

STABILIZING FORCES IN ACOUSTIC CULTURAL EVOLUTION: COMPARING  
HUMANS AND BIRDS

by

Daniel C. Mann

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of  
the requirement for the degree of Doctor of Philosophy, The City University of New York

2019

2019

DANIEL C. MANN

All rights reserved

This manuscript has been read and accepted for the Graduate  
Faculty in Linguistics in satisfaction of the dissertation requirement  
for the degree of Doctor of Philosophy.

JULIETTE BLEVINS

---

Date

---

Chair of the Examining Committee

GITA MARTOHARDJONO

---

Date

---

Executive Officer

MARISA HOESCHELE

DAVID C. LAHTI

MICHAEL I. MANDEL

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

## **Abstract**

# STABILIZING FORCES IN ACOUSTIC CULTURAL EVOLUTION: COMPARING HUMANS AND BIRDS

By

Daniel C. Mann

Advisor: Professor Juliette Blevins

Learned acoustic communication systems, like birdsong and spoken human language, can be described from two seemingly contradictory perspectives. On one hand, learned acoustic communication systems can be remarkably consistent. Substantive and descriptive generalizations can be made which hold for a majority of populations within a species. On the other hand, learned acoustic communication systems are often highly variable. The degree of variation is often so great that few, if any, substantive generalizations hold for all populations in a species.

Within my dissertation, I explore the interplay of variation and uniformity in three vocal learning species: budgerigars (*Melopsittacus undulatus*), house finches (*Haemorrhous mexicanus*), and humans (*Homo sapiens*). Budgerigars are well-known for their versatile mimicry skills, house finch song organization is uniform across populations, and human language has been described as the prime example of variability by some while others see only subtle variations of largely uniform system. For each of these species, I address several questions related to variability and uniformity: What is the typical range of variation? What are the limits of variation? How are those two issues related? And what mechanisms underlie variability and uniformity?

In chapter 3, I investigate a potential domain of uniformity in budgerigar warble: the segment. Segments, units divided by acoustic transitions rather than silence, have been largely ignored in non-human animal communication. I find that budgerigars can achieve a high degree of complexity and variability by combining and arranging these small, more stereotyped units. Furthermore, I find that budgerigar segment organization is not only consistent across independent budgerigar populations but is consistent with patterns found in human language.

In chapter 4, I investigate variability in house finch song. I present data showing that house finches learn sound patterns which are absent in wild house finch populations. These data suggest that cross-population variation in house finch song is narrower than what is permitted by the house finch song learning program.

Finally, in chapter 5, I focus on human language, the most well-described communication system. Here, I research a sound pattern that is absent in the majority of known languages. I find that the rare pattern has independently developed at least six times. In every case, the historical pathway which led to the rare pattern was the same. The historical development in these six linguistic lineages suggests that the overall rarity of the sound pattern is the result of acoustic similarity.

These data illuminate the evolutionary forces that give rise to, and limit, variation. The results of this dissertation have wide-ranging implications, from necessary revisions of linguistic theories, to understanding epigenetic interactions, to the application of evolutionary theory to complex behavior. While these projects within the dissertation are all different, evidence from all three projects support the following claims: (i) cross-population commonality is not evidence for what a species is able to learn; (ii) peripheral mechanisms have a strong influence in limiting cross-population variability; and (iii) high degrees of variation can emerge from uniform traits.

## **Acknowledgments**

First, I would like to thank my supervisor, Professor Juliette Blevins. Her knowledge, guidance, and patience have been invaluable throughout my PhD. I'm immensely grateful for her support when I wanted to take a not-so-typical route through a linguistics PhD.

My committee has also been amazing, and I could not have done this without them. Professor David Lahti has an infectious passion for science which helped convince me to pursue a comparative approach and ultimately won me over to the bird side. Dr. Marisa Hoeschele is a fount of knowledge and has been a source of inspiration. She believed in me when I didn't believe in myself. Dr. Michael I. Mandel was generous with his time and his input was essential in helping me refine my chapters.

I would also like to thank the Ernst Mach Grant committee and the Austrian Science Fund for funding parts of this dissertation, namely Chapter 3. I am also grateful to the Florence Bloch Dissertation Fellowship committee for providing the funds for writing this thesis.

My PhD has been a journey full of self-doubt and self-discovery. This journey has taken this little Arkie from the Graduate Center to the University of Santa Cruz, then to Queens College, and finally to the University of Vienna. Along the way so many colleagues, collaborators, staff, and friends have provided insight, support, and guidance. I want to thank the members of the Lahti Lab, the Pinniped Cognition and Sensory Systems Laboratory, and the Budgerigar Lab. I would also like to thank May Ahmar, Dr. Arafat Angulo-Perkins, Sabrina Bettoni, Nishi Bissoondial, Alexandra Bohmann, Dr. Daniel Bowling, Professor Diane Bradley, Dr. Caroline Casey, Katherine Dawson, Sarah Deventer, Professor Bob Dooling, Adam Fishbein, Professor W. Tecumseh Fitch, Frances Geller, Professor Stefano Ghirlanda, Felix Haiduk, Riccardo Hofer, Professor John Locke, Nadja Kavik, Theresa Matzinger, Phillip Maier, Dr. Ignacio Montoya, the late Professor Paul C. Munding, Dr. Jinook Oh, Petra Pesak, Dr. Stephan Reber, Dr. Colleen Reichmuth, Dr. Andrew Richards, Dr. Cliodhna Quigley, Dr.

Michelle Spierings, Lauren Spradlin, Gabriel Staubmann, Dr. Hsiao-Wei (Vicky) Tu, Laura Waddick, Bernhard Wagner (hi Bernhard!), and Mason Youngblood for thoughtful discussions and input during my PhD.

I'd like to give a special thanks to Shahrzad Afroozeh for her extensive technical help and, more importantly, extended conversations about budgie drama and to Dr. Christen Madsen II for moral support from day one of my PhD and, of course, for the countless hours we spent at chain restaurants in NYC.

Finally, I'd like to thank my family. To Maria and Alfred, Fabi, Andreas and Martina, Oma, and the Loimersdorf Großeltern: for taking care of Cass so that Raffi and I could have a break and for the words of encouragement. Now that the dissertation is complete I can focus more on learning German so that I can understand Andreas's jokes. To Shelli, for the laughs and for being a source of strength. I might have the titles but I'm proud that my little sister is still the smart one. To my dad, Gary, for teaching me to question everything and to keep an open mind. Because of him I learned to love the written word. To my mom, Marcella, words can't express how grateful I am for all of your support. You taught me the value of hard work and perseverance. You sacrificed so much to give us a better life. All my successes I owe to you.

Last, but not least, to Cass and Raffi. My dog, Cass, tried his best to give emotional support when I was swearing at my computer, but, bless his heart, he's a shit emotional support dog. Still, he was always there with loving eyes. He kept me grounded, made me exercise, and forced me to be calm (if only to not wake him up so he wouldn't walk over and "help"). Raffi, you have been an amazing partner. You helped me keep perspective, you made me laugh, you kicked my ass when I needed it. I'm in awe of your brilliance.

## Table of Contents

<b>Chapter 1. Understanding limits on variation in human language and elsewhere.....</b>	<b>1-1</b>
1.1. Introduction.....	1-1
1.1.1. Background and terminology .....	1-4
1.1.2. Vocal learning .....	1-5
1.1.3. Units of analysis .....	1-8
1.1.4. Terminology .....	1-11
1.2. Variation and uniformity in human language .....	1-15
1.2.1. Degree of variation in human language.....	1-15
1.2.2. Limits on variation in human language .....	1-17
1.2.3. Causes of uniformity and variation in human language .....	1-20
1.3. Variation and uniformity in birdsong .....	1-24
1.3.1. Degree of variation in birdsong .....	1-24
1.3.2. Limits on variation in birdsong .....	1-30
1.3.3. Causes of variation and uniformity in birdsong .....	1-31
1.4. Summary .....	1-35
1.5. Structure of the dissertation .....	1-37
1.5.1. Chapter two overview.....	1-37
1.5.2. Chapter three overview.....	1-38
1.5.3. Chapter four overview .....	1-38
1.5.4. Chapter five overview .....	1-39
<b>Chapter 2. Comparing humans and birds: Direct parallels in acoustic communication .....</b>	<b>2-40</b>
2.1. Introduction.....	2-40
2.1.1. Convergent evolution .....	2-40
2.2. Neural similarities .....	2-41
2.3. Genetic similarities .....	2-42
2.4. Similarities in sound perception.....	2-44
2.5. Similarities in sound production .....	2-46
2.6. Summary .....	2-51
<b>Chapter 3. Newfound uniformity: Human-like segments in budgerigar song.....</b>	<b>3-52</b>

3.1. Introduction.....	3-52
3.1.1. Segments in human language .....	3-53
3.1.2. Segments in non-humans?.....	3-54
3.1.3. Comparative approach.....	3-58
3.1.4. Chapter organization .....	3-59
3.2. General methods .....	3-60
3.2.1. Data collection.....	3-60
3.2.2. Utterance/song division .....	3-62
3.3. Experiment 1: Basic units of warble.....	3-63
3.3.1. Experiment 1a: Segmentation.....	3-64
3.3.2. Experiment 1b: Cluster analysis.....	3-67
3.3.3. Experiment 1b: Discussion.....	3-73
3.4. Experiment 2: Population prediction .....	3-74
3.4.1. Experiment 2: Methods .....	3-75
3.4.2. Experiment 2: Results.....	3-76
3.4.3. Experiment 2: Discussion.....	3-76
3.5. Experiment 3: Edge effects .....	3-77
3.5.1. Experiment 3: Methods .....	3-78
3.5.2. Experiment 3: Results.....	3-79
3.5.3. Experiment 3: Discussion.....	3-82
3.6. General discussion .....	3-87
<b>Chapter 4. Newfound variation: House finches learn canary trills .....</b>	<b>4-90</b>
4.1. Introduction.....	4-90
4.2. Methods.....	4-94
4.2.1. Cross-fostering subjects.....	4-94
4.2.2. House finch ontogeny subjects .....	4-98
4.2.3. Wild recorded subjects .....	4-99
4.2.4. Song analysis .....	4-100
4.3. Results.....	4-104
4.3.1. Phrase features.....	4-106
4.3.2. Song features .....	4-108
4.4. Discussion .....	4-111
4.4.1. Why do house finches learn heterospecific song?.....	4-113
4.4.2. Why don't wild house finches trill? .....	4-115

4.4.3. Why is learning incomplete? .....	4-116
<b>Chapter 5. Variation and uniformity in spoken language syllable onsets: the role of sonority in speech.....</b>	<b>5-124</b>
5.1. Introduction.....	5-124
5.2. Methods.....	5-128
5.3. Results.....	5-134
5.3.1. Data.....	5-134
5.3.2. Tautosyllabicity .....	5-150
5.3.3. The evolution of #RO- onsets.....	5-160
5.4. General discussion .....	5-167
5.4.1. Natural sound change? .....	5-167
<b>Chapter 6. Conclusions.....</b>	<b>6-171</b>
6.1. Introduction.....	6-171
6.1.1. Dissertation summary .....	6-171
6.2. Universals and learnability .....	6-173
6.3. Peripheral mechanisms of recurrent patterns .....	6-174
6.4. Variation out of uniformity .....	6-174
<b>Chapter 7. Appendix.....</b>	<b>7-176</b>
7.1. Chapter 3 supplement .....	7-176
7.2. Code .....	7-177
7.3. Sonorant + obstruent onset list.....	7-178
<b>Chapter 8. References.....</b>	<b>8-186</b>

## List of Figures

Figure 1-1. Phylogeny of vocal learners.....	1-6
Figure 1-2. English syllable organization.....	1-9
Figure 1-3. Avian song organization.....	1-9
Figure 1-4. Budgerigar warble organization.....	1-10
Figure 1-5. Human speech articulators.....	1-14
Figure 1-6. Possible design space for learned acoustic communication system.....	1-36
Figure 3-1. Human inter-silence interval.....	3-52
Figure 3-2. Budgerigar warble phrase types.....	3-56
Figure 3-3. Complex phrases.....	3-57
Figure 3-4. Human speech segmentation output.....	3-66
Figure 3-5. Budgerigar warble segmentation output.....	3-67
Figure 3-6. Silhouette values for human segment and phrase cluster sizes.....	3-71
Figure 3-7. Silhouette values for budgerigar segment and phrase cluster sizes.....	3-72
Figure 3-8. Segment and phrase cluster sizes of 2, 3, and 4.....	3-73
Figure 3-9. Duration and F0 of budgerigar medial and final segments.....	3-80
Figure 3-10. Periodicity by in budgerigar segments by phrase position.....	3-81
Figure 3-11. Budgerigar and human phrase initial patterns.....	3-82
Figure 4-1. Canary and house finch song spectrograms.....	4-93
Figure 4-2. Spectrograms of canary-tutored house finches.....	4-104
Figure 4-3. Spectrograms of quiet-reared house finches.....	4-105
Figure 4-4. Length of repeated phrases by the sameness likelihood.....	4-108
Figure 4-5. Similar trill types in a canary and canary-tutored house finch.....	4-111
Figure 5-1. Sonority hierarchy.....	5-125
Figure 5-2. Sonorant + obstruent onset repairs.....	5-151

Figure 5-3. Consonant + vowel gestural synchronicity..... 5-168  
Figure 7-1. Silhouette width comparison for budgerigar warble units..... 7-176

## List of Tables

Table 1-1. Proposed human language universals. ....	1-18
Table 3-1. Acoustic variables measured. ....	3-69
Table 3-2. Budgerigar population classification based on unit type. ....	3-76
Table 3-3. Group means and standard deviations by segment position. ....	3-79
Table 4-1. Individuals & number of quality recordings. ....	4-100
Table 4-2. House finch and canary phrase level measures. ....	4-106
Table 4-3. Random Forest classification of house finch phrases. ....	4-107
Table 4-4. Group medians & p-values for song level. ....	4-109
Table 5-1. Sonorant + obstruent onset languages. ....	5-132
Table 5-2. Klamath consonant inventory. ....	5-134
Table 5-3. Klamath sonorant + obstruent onsets. ....	5-135
Table 5-4. Georgian consonant inventory. ....	5-137
Table 5-5. Georgian sonorant + obstruent onsets. ....	5-138
Table 5-6. Piro consonant inventory. ....	5-140
Table 5-7. Piro sonorant + obstruent onsets. ....	5-141
Table 5-8. Panixtlahuaca Chatino consonant inventory. ....	5-143
Table 5-9. Panixtlahuaca Chatino sonorant + obstruent onsets. ....	5-144
Table 5-10. Russian consonant inventory. ....	5-146
Table 5-11. Russian sonorant + obstruent onsets. ....	5-147
Table 5-12. Tsou consonant inventory. ....	5-148
Table 5-13. Tsou sonorant + obstruent onsets. ....	5-149
Table 5-14. Implicational relationships. ....	5-150

## List of symbols and abbreviations

[]	Acoustic realization of a sound
//	Mental representation of a sound
“”	gloss; definition; new term
C	Consonant
V	Vowel
R	Sonorant
O	Obstruent
N	Nasal
*	Reconstructed
**	Unattested
>	Develops from
>>	Implies
~	Alternates with

## Chapter 1. Understanding limits on variation in human language and elsewhere

### 1.1. Introduction

The most recent shared ancestor of birds and humans lived roughly 310 million years ago (Kumar & Hedges, 1998). This evolutionary distance makes cognitive and behavioral similarities between parrots, songbirds, and humans intriguing, particularly since these similarities are not widespread in avian or mammalian taxa. For one, corvids and many parrot species have cognitive abilities on par with that of primates (Güntürkün & Bugnyar, 2016). Like humans, African grey parrots exhibit transitive inference (Pepperberg, 2010), western scrub jays plan for future needs (Correia, Dickinson, & Clayton, 2007; Raby, Alexis, Dickinson, & Clayton, 2007), and ravens may have a basic theory of mind (Bugnyar, Reber, & Buckner, 2016).

With all these impressive abilities, one would be forgiven for overlooking the small and unassuming budgerigar, a parrot endemic to the arid regions of Australia. Budgerigars have brains that are roughly the size of a jellybean. Yet a quick glance on YouTube turns up videos of budgerigars mimicking the bouncing sound of ping pong balls, the voice of R2-D2, and even the complex dialogue of the Flintstones and Monty Python. In a video of a budgerigar named Disco, the budgie can be heard repeating a famous line from the British sketch comedy: “Nobody expects the Spanish Inquisition”.<sup>1</sup> Even more interesting, Disco repeats part of the phrase while also mixing in previously learned phrases. In one clip, Disco produces “Nobody expects the cheeseburger”, a combination of the aforementioned Monty Python phrase and the meme “I can haz [*sic*] cheeseburger”. Notice that Disco isn’t mindlessly “parroting” but seems to be dividing the utterances into subunits and recombining them in novel and flexible ways.

---

<sup>1</sup> <https://www.youtube.com/watch?v=hNavxvvN9QI>

While budgerigars have the ability to mimic human speech (Banta Lavenex, 2000; Scanlan, 1999), zoos and pet stores are not replete with the sounds of 1970s punchlines, screams of “yabba-daba-doo”, and requests for crackers. In fact, to the human ear, there is little difference between a budgerigar colony at a zoo in Nebraska, United States and another in Vienna, Austria. Published analyses of budgerigar vocalizations confirm that populations share much in common (Farabaugh, Brown, & Dooling, 1992; Farabaugh, Dent, & Dooling, 1998; Tobin, Medina-García, Kohn, & Wright, 2017; Tu, Smith, & Dooling, 2011).

On the one hand, budgerigar song, called “warble”, can vary so much that it can include mimicked human speech. On the other, budgerigar warble in several independent populations share basic properties. How could some individuals or populations deviate so far from the species-typical acoustic communication system? How much further could budgerigars deviate? What is species-typical for them in the first place? And if potential for variability exists, why do most budgerigars operate within a much narrower range of the acoustic communication landscape?

Similar questions can be asked of human language. Language can vary so much that even the basic mode of transmission differs between populations; languages can be spoken or they can be signed (Sandler & Lillo-Martin, 2006). Still, some descriptions hold for the vast majority of languages. All spoken languages, that we know of, use the larynx as the primary sound source. Mark Baker (2001) has referred to interplay of variability and diversity in language as the “code talker’s paradox”. During World War II, the United States marine corps used Navajo speaking Americans to transmit military messages in their native language of Navajo (or Diné). The codes were written in English, translated into Navajo, and then translated back into English with no loss of fidelity. Despite prior success in breaking American codes, cryptographers in the Japanese military were unable to crack the Navajo code. The paradox is that Navajo is similar enough to English that messages of life or death importance can be

translated from one language to the other with precision; at the same time, Navajo is so different from English and Japanese that experts could not figure out how to break the code. For both humans and budgerigars, a complex communication system is simultaneously variable and uniform across individuals and populations.

The long history of research in learned acoustic communication, in human language and birdsong, suggests that systems can be analyzed from either the perspective of variability or the perspective of uniformity. Within and across individuals, populations, and species measurable differences exist. At the same time, all or most individuals in a population or species may share acoustic similarities. The goal of this dissertation, presented in (1-1), is to ask: From the perspective of variation, what is the typical range of the acoustic form?; From the perspective of uniformity, what are the recurrent limits on variation?

***1-1. Question 1***

- a. What is the typical range of variation in (the acoustic form of) human speech and in song of avian species?
- b. What are the recurrent limits on this variation?

A secondary goal of this dissertation, shown in (1-2), is to try to explain what underlies the recurrent limits on variation within systems.

***1-2. Question 2***

- a. What mechanisms underlie the recurrent limits on variation in these systems?
- b. Can study of variation and its limits in one species inform that of others?

Descriptions and analyses of the range and limits of variation naturally feed into questions of “why”, particularly when cross-population comparisons uncover asymmetries in sounds and sound patterns. Multiple independent developments of some aspect of an acoustic communication system is unexpected if no limiting factors exist. But if recurrent patterns do exist, comparative work can suggest if that pattern is the result of some functional pressure, articulatory limitation, tendency for misperception, cognitive bias, etc. With enough data

points, it is possible to formulate precise hypotheses to explain why certain sounds or sound patterns are more common than others and why some are unattested.

In this dissertation, I present data on the learned acoustic communication system of one parrot species – budgerigars, one songbird – the house finch, and one primate species – humans. The budgerigar is well-known for variability in its song (Banta Lavenex, 2000; Gramza, 1970; Scanlan, 1999), in chapter 3, I investigate a potential domain of uniformity: the segment. Segments, units divided by acoustic transitions rather than silence, have been largely ignored in non-human animal communication, in spite of research showing strong cross-population tendencies in human segment inventories and segment organization (Hyman, 2008; Vaissière, 1983). In chapter 4, I turn to the house finch, a songbird species that has shown cross-population uniformity in its song (Bitterbaum & Baptista, 1979; Ju, 2015; Mundinger, 1975; Pytte, 1997; Tracy & Baker, 1999). In this chapter, I present data showing that they learn sound patterns which are not found in wild populations. Finally, in chapter 5, I focus on human language, the most well-described communication system. Here, I research a sound pattern that is absent in the majority of known languages and I investigate the historical development of the pattern in the small number of languages in which it is present. While specific questions differ in all of the projects, all focus on variation and uniformity. All discuss the typical range in the species, the recurrent limits on variation, and the possible explanations underlying variability and uniformity.

### **1.1.1. Background and terminology**

This dissertation bridges research between the study of human language and the study of animal communication systems. It is certainly not the first to do so, but care must always be taken with interdisciplinary research as assumptions and terminology vary across fields. Throughout the dissertation, I do my best to make concepts, terms, and assumptions as clear as

possible for a broad audience. However, it is valuable to lay out a few core ideas from the onset.

### **1.1.2. Vocal learning**

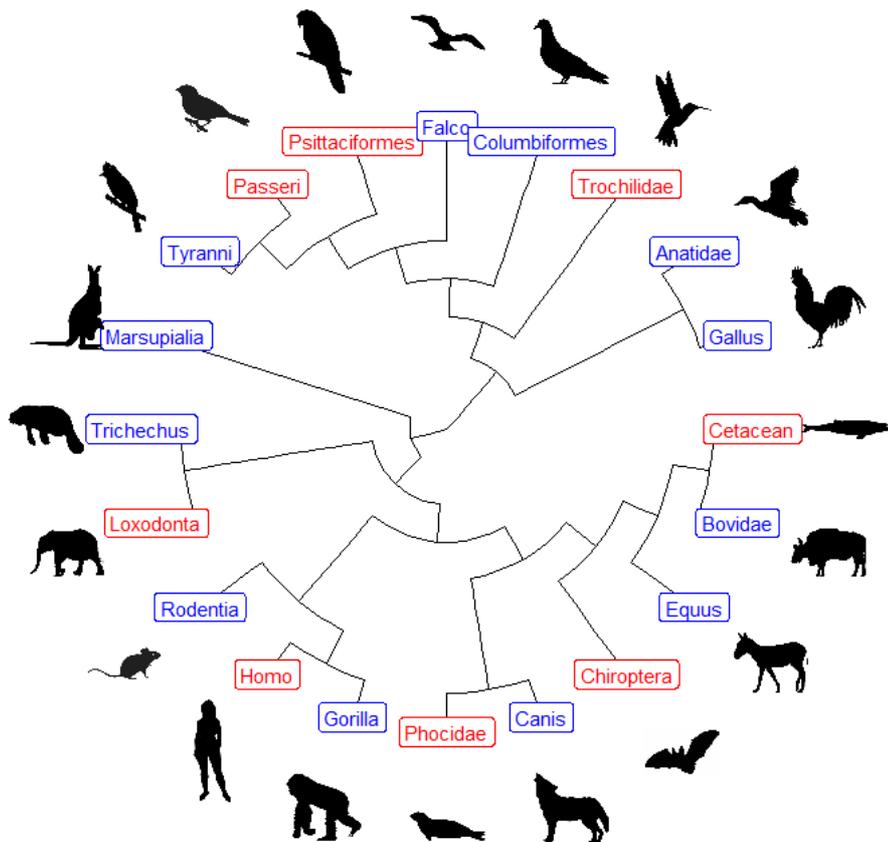
Learning of acoustic communication systems can take several forms, but research typically focuses on acoustic communication signals that are learned through the mimicry of others. This ability is called “vocal production learning” or simply “vocal learning”. Vocal learning is defined by the production of novel signals or the modification of preexisting signals due to interaction with other individuals (Janik & Slater, 1997).

Vocal learning is an important component of human language. Through vocal learning a child learns the speech sounds and sound patterns of their social group. More specifically, humans learn the sounds of those they interact with, independent of genetic relationships. Vocal learning likely plays a role in the ability generate a practically infinite set of possible sound combinations. These sound combinations can be applied to objects, ideas, actions, etc. which allows for the vast number of symbols used in human language (Studdert-Kennedy, 1998).

Vocal production learning is distinguished from other types of acoustic communication learning such as comprehension learning and usage learning. Comprehension learning is the association of a preexisting signal to a novel context while usage learning is learning to produce a preexisting signal in a novel context. Comprehension and usage learning are shared amongst a wide range of animal species (Janik & Slater, 1997). In this dissertation I use the term “vocal learning” only for vocal production learning.

Unlike usage and comprehension learning, vocal learning is relatively rare in the animal kingdom. Of the species that have been investigated thus far, only three avian taxa and five mammalian taxa show evidence of vocal learning. Vocal learning birds include parrots, songbirds, and hummingbirds (Araya-Salas & Wright, 2013; Bradbury & Balsby, 2016;

Marler, 1970b). Vocal learning mammals include cetaceans, bats, pinnipeds, elephants and humans (Janik, 2014; Reichmuth & Casey, 2014; Stoeger & Manger, 2014; Vernes, 2017). Because most of these vocal learning taxa have sister taxa which are not vocal learners, see the phylogenetic tree in Figure 1-1, vocal learning is likely to have evolved independently multiple times (Jarvis, 2006).



**Figure 1-1. Phylogeny of vocal learners.** Evidence for vocal production learning has been found in at least eight taxa (shown in red). The taxa in blue are those that are closely related to vocal learners, yet lack evidence for vocal learning. All vocal production learning taxa are represented in the figure, while only a small subset of non-learners is presented. Because vocal production learning is relatively rare and broadly distributed across the classes Mammalia and Aves, vocal production learning seems to have evolved independently multiple times. Phylogenetic data extracted from Michonneau, Brown, & Winter (2016). Silhouette images taken from Chamberlain (2018).

The evidence for vocal learning in these taxa is not all equal. For pinnipeds, bats, and hummingbirds, it is unclear if vocal learning is widespread within the groups or if vocal learning is limited to a few species (Araya-Salas & Wright, 2013; Reichmuth & Casey, 2014; Vernes, 2017). In pinnipeds and elephants, unambiguous data are limited to a few individuals (Ralls, Fiorelli, & Gish, 1985; Stoeger et al., 2012). For example, Hoover, a harbor seal raised by a fisherman off the coast of New England, would yell at passersby in a New England accent (Ralls et al., 1985). Koshik, an Asian elephant, would use his trunk to modify his vocal tract in order to mimic his Korean-speaking trainer (Stoeger et al., 2012). Suggestive evidence in these species, however, is more widespread. Many pinniped species have dialects in their acoustic systems and some have been trained to create novel sounds (Reichmuth & Casey, 2014). Dialects are also present in elephants and the neuroanatomy of elephants suggests they have the requisite control needed for vocal production learning (Stoeger & Manger, 2014).

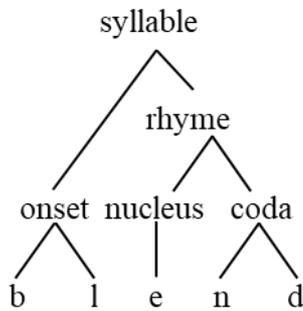
In humans, songbirds, and parrots, however, the evidence for vocal learning is robust (Doupe & Kuhl, 1999; Farabaugh, Linzenbold, & Dooling, 1994; Jarvis, 2007; Nottebohm, 1972; Pepperberg, 2010; Williams, 2004). Evidence typically comes in two forms. The first is when species-typical acoustic input is deprived or manipulated (Marler & Sherman, 1983; Senghas & Coppola, 2001; Thorpe, 1958). For vocal learners, if input is absent or atypical, vocal behavior develops abnormally. In non-vocal learners, vocal behavior is largely unaffected under those same conditions. For instance, cross-fostered primates use the call of their genetic parents, not fostered parents (Owren, Dieter, Seyfarth, & Dorothy, 1992). Chickens raised in isolation produce species-typical vocalizations (Konishi, 1963). The second line of evidence for vocal learning comes from the mimicry of other species. Learning is likely the only explanation for why one species could reproduce the signal of another (Dobkin, 1979). Other lines of evidence include variation between populations and changing of vocalizations when exposed to a novel environment. However, the variation between populations, or

“dialects”, can arise in non-learning species (Seddon, 2005). Modifications of a vocalization in a novel environment could be an instance of usage learning where a signal was already present but only used when exposed to the novel group or situation (Fischer, Wheeler, & Higham, 2015).

### **1.1.3. Units of analysis**

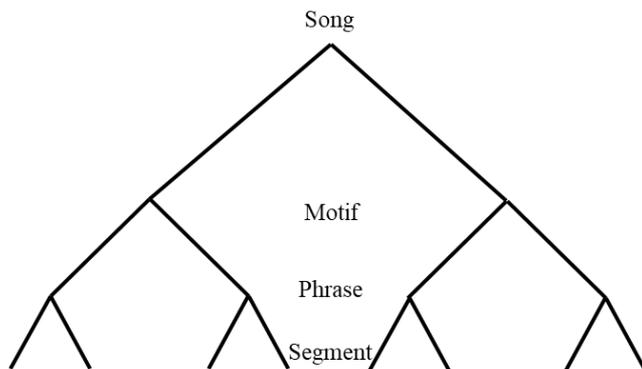
When making comparisons across species one of the most difficult problems is knowing which vocal units are comparable. Terminology for unit types in learned acoustic communication reflects this difficulty. For instance, a birdsong “syllable” has been defined as a unit which is divided by a silence (Isaac & Marler, 1963), a unit of a set duration (Thompson, LeDoux, & Moody, 1994), a consistent sequence of subunits (Lemon, 1975), or a breath group (Franz & Goller, 2002). While these definitions are not always mutually exclusive, they are very different from definitions of syllables in spoken human languages. In human language, syllables are often defined by a cyclical rise and fall in intensity. However, there is no clear, precise acoustic definition for the syllable, at least not one that can account for syllables across all languages (Ohala & Kawasaki-Fukumori, 1997). However, evidence from speech errors, language games, and language-specific grammatical rules suggest that for many languages the syllable is an important perceptual unit (Blevins, 1995). Human syllables are typically built around a nucleus, often the most acoustically intense segment in the syllable (usually a vowel). Syllables can also take an onset and a coda. The nucleus combines with the coda to form a rhyme. Cross-linguistically, syllables with an onset and nucleus are more common than those with only a nucleus or those with a nucleus and coda (Blevins, 2006b). Languages vary in the segments that are permitted in syllable positions. Vowels, which are loud, periodic segments, are the most common syllable nuclei. Short, aperiodic bursts are common in the onset. The structure in Figure 1-1 demonstrates the internal structure of an English syllable, *blend*. The

segment *e* [ɛ] is the nucleus. The consonant cluster *bl* form the onset and the consonant cluster *nd* form the coda.



**Figure 1-2. English syllable organization.**

Because the syllable is a perceptual unit, there is no clear analog of the syllable in avian communication. Syllables organize segments in human language, but in birdsong very little research has been done at the level of the segment. However, like human language, birdsong can be hierarchically organized, so subsequent research may uncover more analogous structures. In Figure 1-3 one possible structure is presented.

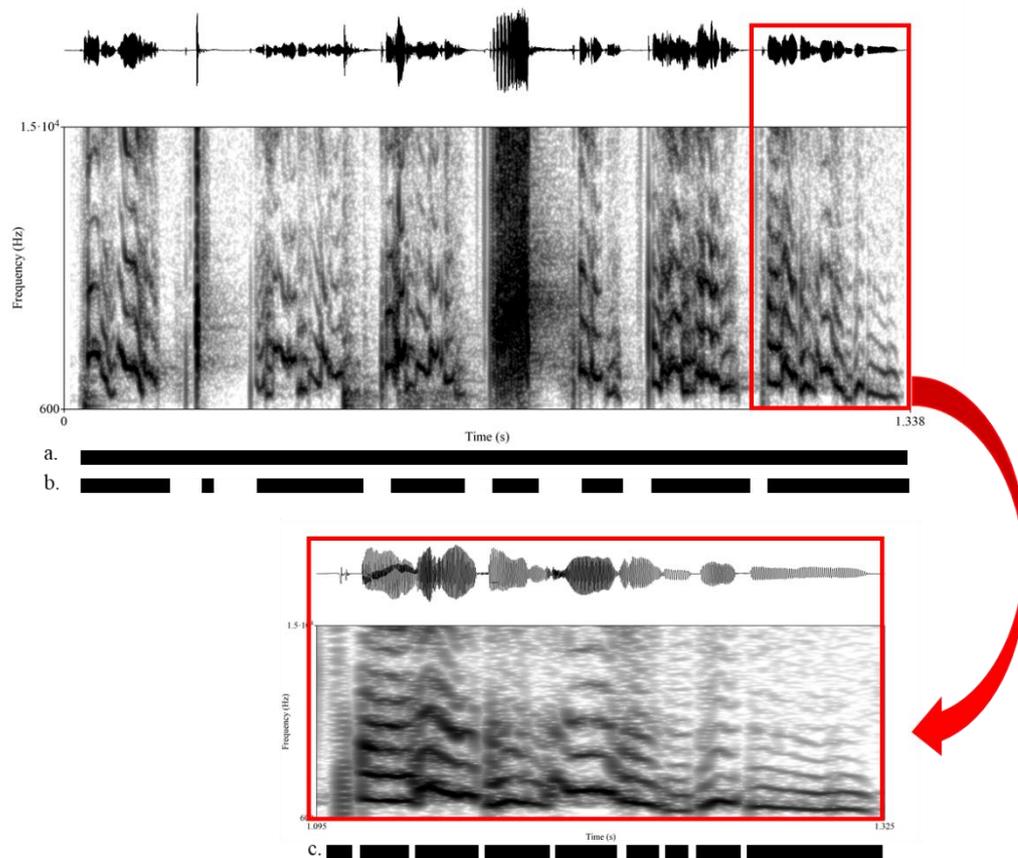


**Figure 1-3. Avian song organization.** Song can be organized hierarchically in birds. Segments are the smallest unit. Full song is the highest. Between phrases and segments organizational units like (human) syllables could exist.

Precise definitions for bird vocal units are usually species specific, but general statements hold true for most species. Song is generally the largest unit. It is typically longer than a second and buffered by ~ 100 milliseconds of silence. Song is divided into phrases (also called syllables). Phrases are 50 – 500 milliseconds long and usually surrounded by 10 – 100

milliseconds of silence. Phrases are divided down into segments, also called notes. Segments are the smallest unit and are usually divided by short periods of silence or rapid acoustic transitions. Some species have motifs which are consistent sequences of phrases. Many species also have trills which are fast repetitions of a phrase or segment (Catchpole & Slater, 2008; Thompson et al., 1994).

In Figure 1-4, three levels of a specific system, budgerigar warble, are shown.



**Figure 1-4. Budgerigar warble organization.** Sequence of a budgerigar song. The phrase in red is zoomed in to better see the segmental structure.

A) Song: Usually longer than a second, composed of smaller units, divided by longer pauses.

B) Phrase: Usually 50-500 msec. May have subunits. Divided by 10-100 msec. Also called syllable.

C) Segment: Smallest unit. Divided by very small periods of silence or acoustic transitions. Also called note.

Budgerigar song is complex and variable. Songs are composed of phrases, though there seems to be little stereotypy in phrase organization. Tu (2009) suggests that structure and

regular patterns do exist. I discuss the level of the segment in Chapter 3 of this dissertation. Intermediate units between phrases and segments may also exist.

#### 1.1.4. Terminology

In this section, I list a few other key terms that will be useful throughout this dissertation. General terms are in 1.1.4.1 and sound class terminology can be found in 1.1.4.1.1 and 1.1.4.1.2. More complicated terminology related to the units of analysis (e.g., such as syllable, song, phrase) are defined in the previous section, 1.1.3. Some terminology requires a more detailed discussion and may not be pertinent to every chapter. These terms are defined within the relevant chapter. For instance, Chapter 3 contains a list of acoustic variables in Table 3-1 and in Chapter 5, I defined the Sonority Hierarchy and the Sonority Sequencing Generalization, Figure 5-1 and (5-1), respectively.

##### 1.1.4.1. General terminology

**Allophone:** Acoustic variants of a mental phonemic category. They do not contrast meaning. In English, the phoneme /n/ can be produced with an alveolar articulation [n], as in *new* [nu], or with a dental articulation [n̪] when the sound precedes *th* [θ], as in *tenth* [tɛn̪θ].

**Aperiodic:** Sounds that lack a clear fundamental frequency; do not have a repeating pattern.

**Aspiration:** “A period of voicelessness after the release of an articulation, as in English in *pie* [p<sup>h</sup>aɪ]” (Ladefoged & Johnson, 2011, p. 305).

**Coda:** Within a (human) syllable, the unit or units that follow the nucleus

**Conspecific:** Same species.

**Features:** Subsegmental units. Descriptive properties of sounds. E.g., [+voiced] for sounds with vocal fold vibration.

**Formant:** “A resonance of the vocal tract; formants act as frequency ‘windows’ (band-pass filters) that allow certain frequencies to pass through the vocal tract relatively unhindered. Formants are the central acoustic cue separating different vowel sounds” (Fitch, 2010, p. 514).

**Fundamental frequency (F0):** “The lowest frequency in a periodic waveform; central determinant of voice pitch” (Fitch, 2010, p. 514).

**Heteromorphic:** Belonging to the different morphemes.

**Heterospecific:** Different species.

**Honest signal/cue:** “Signals that provide accurate information to perceivers, either about the quality or properties of the signaler itself (e.g. advertisement calls), or about something in the environment (e.g. alarm calls)” (Fitch & Hauser, 2003, p. 65).

**Index:** “In animal communication, signals which contain information because of a physical law (e.g. formants as an indexical cue to body size)” (Fitch, 2010, p. 515).

**Innate:** “Reliably developing or ‘canalized.’ Innate traits are those whose development shows a high robustness in the face of environmental variation; this does not imply

that they are necessarily inflexible or that environmental information does not play a role in their development” (Fitch, 2010, p. 515).

**Manner of articulation:** How articulation is made in the vocal tract.

**Monomorphemic:** Belonging to the same morpheme.

**Morpheme:** The smallest unit of meaning in language.

**Natural class:** “Any complete set of sounds in a given language that share the same value for a feature or set of features” (Hayes, 2009, p. 43).

**Nucleus:** Within a (human) syllable, the sonority peak or the most prominent part of the syllable. Obligatory part of the syllable.

**Onset:** Within a (human) syllable, the unit or units that come before the nucleus.

**Periodic:** Sounds that have a clear fundamental frequency; repetition at regular intervals.

**Phonation:** “Vibration of the vocal folds” (Ladefoged & Johnson, 2011, p. 309).

**Phone:** A sound.

**Phoneme:** “The smallest distinctive unit in the structure of a given language” (Ladefoged & Johnson, 2011, p. 309). Contrastive sound. Mental sound category.

**Phonetics:** “The detailed nature of speech production and perception” (Fitch, 2010, p. 516). Involves the production, perception, and acoustics of speech sounds.

**Phonology:** “Unconscious rules for sound patterning that are found in the mind/brain of a person who speaks a particular language” (Hayes, 2009, p. 19).

**Phonotactics:** “The principles of phonological well-formedness in a particular language” (Hayes, 2009, p. 64). How sounds are organized in a language and what sound patterns are permitted.

**Place of articulation:** Where an articulation is produced in the vocal tract.

**Rime:** Within a (human) syllable, the nucleus and the coda.

**Source:** The conversion of airflow into sound; these sounds are “filtered by the vocal tract. Voicing, friction, and burst noises are the main sound sources in human speech” (Ladefoged & Johnson, 2011, p. 310).

**Source-filter theory:** “A theory that assumes the time-varying glottal airflow to be the primary sound source and the vocal tract to be an acoustic filter of the glottal source” (Fitch, 2010, p. 517).

**Spectrum:** “Short for ‘spectrum of frequencies’; a display of relative magnitudes or phases of the component frequencies of a waveform (Fitch, 2010, p. 517).

**Syncope:** The loss of a sound within a word.

**Taxon:** “A natural grouping of animals, linked by descent from a common ancestor. Species are [taxa], and so are families (like the cat family) or classes (like mammals)” (Fitch, 2010, p. 513).

**Voiced:** “Having vibrations of the vocal folds during an articulation, as in English [m] in *me*” (Ladefoged & Johnson, 2011, p. 311). Typically used in reference to human linguistic sounds.

**Voiceless:** “Pronounced without vibrations of the vocal folds, as in English [s] in *see*” (Ladefoged & Johnson, 2011, p. 311). Typically used in reference to human linguistic sounds.

#### 1.1.4.1.1. Sound classes (manner)

Generally speaking, human linguistic sounds can be divided into two broad categories: vowels and consonants. Vowels are louder, periodic sounds made with a relatively open vocal

tract. Consonants are produced with some constriction in the vocal tract. Because of the constriction, consonants typically have lower amplitudes and energy that is more widely dispersed across the sound spectrum. Both sound classes can be subdivided, but I focus more on divisions within consonants in this dissertation.

**Sonorant:** Subset of consonants; “sounds produced with a vocal tract cavity configuration in which spontaneous voicing is possible” (Chomsky & Halle, 1968, p. 302). Includes nasals, liquids, laterals, rhotics, glides.<sup>2</sup>

**Types of sonorants:**

**Glide:** Sonorant consonants “which are vowel-like, but which, unlike vowels, do not occupy the head position in the nucleus of a syllable” (Carr, 2008, pg. 63). As in English [w] in *win* and [j] in *yes*.

**Lateral:** Sonorant consonants where “the airstream flows over the sides of the tongue, as in English [l] in *leaf*” (Ladefoged & Johnson, 2011, p. 308).

**Liquid:** Sonorant consonants that “have the characteristic acoustic quality of *l*-like and *r*-like sounds” (Hayes, 2009, p. 7).

**Nasal:** Sonorant consonants where “the velum is lowered, allowing air to escape through the nose. Most nasal consonants have a complete blockage within the mouth at the same time.” (Hayes, 2009, p.7). As in English [m] in *mine* and [n] in *nine*

**Rhotic:** Sonorant consonants; *r*-like sounds. The most common rhotics are “trills made with the tip or blade of the tongue” (Ladefoged & Maddieson, 1996, p. 215). However, sounds within the rhotic class show substantial variation in their manner and place of articulation. A lowered 3<sup>rd</sup> formant may be the characteristic which is shared by all members of the class.

**Tap/flap:** Sonorant rhotic consonants in which “an articulator makes a rapid brush against some articulatory surface” (Hayes, 2009, p. 7). As in American English [ɾ] in *butter*.

**Trill:** Sonorant consonants where “an articulator is made to vibrate by placing it near an articulatory surface and letting air flow through the gap” (Hayes, 2009, p. 7). As in Spanish [r] in *perro*.

**Obstruent:** Subset of consonants; sounds that “are produced with a cavity configuration that makes spontaneous voicing impossible” (Chomsky & Halle, 1968, p. 302). Includes stops, affricates, and fricatives.

**Types of obstruents:**

**Affricate:** Obstruent consonant; “a stop followed by a fricative, made at the same location in the mouth in rapid succession so that the result has the typical duration of a single speech sound.” (Hayes, 2009, p.7). As in English [tʃ] in *cheap* and [dʒ] in *jeep*.

**Click:** Obstruent consonant “made with an ingressive velaric airstream” as Zulu [!] (Ladefoged & Johnson, 2011, p. 306).

---

<sup>2</sup> The definition of “sonorant” often includes vowels. In this dissertation, I use sonorant only to refer to sonorant consonants.

**Ejective:** Obstruent consonant “made with an egressive glottalic airstream, such as Hausa [k’]” (Ladefoged & Johnson, 2011, p. 306).

**Fricative:** Obstruent consonants where “a tight constriction is made, so that air passing through the constriction flows turbulently, making a hissing noise” (Hayes, 2009, p.7). As in English [θ] in *thin* and [s] in *sin*.

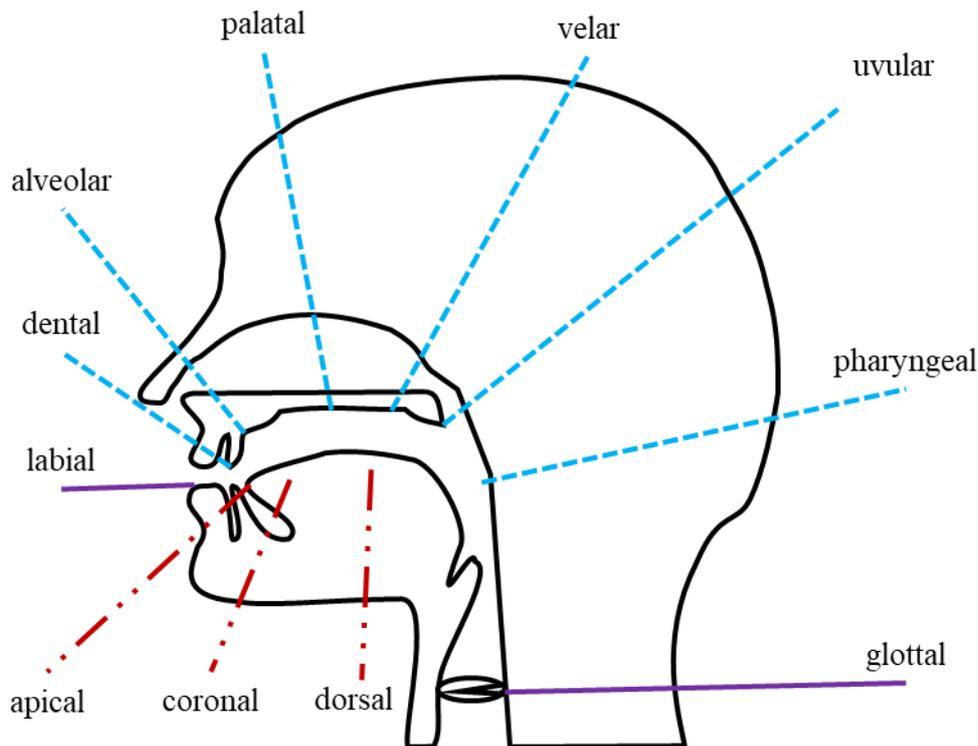
**Implosive:** Obstruent consonant “made with an ingressive glottalic airstream, such as Sindhi [ɓ]” (Ladefoged & Johnson, 2011, p. 307).

**Sibilant:** “A speech sound in which there is high-amplitude, turbulent noise, as in English [s] and [ʃ] in *sip* and *ship*” (Ladefoged & Johnson, 2011, p. 310). A type of fricative.

**Stop/Plosive:** Obstruent consonants created when “airflow through the mouth is momentarily closed off” (Hayes, 2009, p.7). As in English [p] in *pay* and [d] in *day*.

#### 1.1.4.1.2. Sound classes (place of articulation)

Figure 1-5 shows the major places in the vocal tract where a constriction can be made.



**Figure 1-5. Human speech articulators.** Common active (red alternating lines) and passive (blue dashed lines) articulators used during speech production. Purple solid lines do not fall cleanly into either category.

The blue consistent dashes show the points of articulation that do not move during speech, “passive articulators”. “Dental” sounds are made with articulation at the back of the teeth, “alveolar” are made at the alveolar ridge, “palatal” at the palate, “velar” at the soft palate,

“uvular” at the uvula, and “pharyngeal” are made in the back of the throat. The red alternating lines in Figure 1-5 show the “active articulators”, those that move to form a constriction. “Dorsal” sounds are made with the back of the tongue, usually at the uvula or the soft palate. “Coronal” sounds are those made with the tip of the tongue or the tongue blade, typically at the alveolar ridge or the teeth. “Apical” is a subset of coronal; apical sounds are made with the tongue tip. The purple solid line in Figure 1-5 show those articulators that can be active or passive. “Glottal” sounds are made at the larynx. “Labial” sounds are produced with the lips.

In the following sections, I review instances of variability and uniformity in human language and avian song. I also discuss theoretical frameworks which attempt to explain variability and uniformity in learned acoustic communication systems.

## **1.2. Variation and uniformity in human language**

### **1.2.1. Degree of variation in human language**

Language can be transmitted via the oral-auditory channel (spoken languages) or the manual-visual channel (signed languages) and can still convey the same (linguistic) information. That the most basic property of a communicative system can differ between communities suggests that language is highly variable. While it remains understudied, linguistic features have also been found in a haptic form of American Sign Language used in blind-deaf communities (“tactile American Sign Language”; Edwards, 2012), suggesting even further variability. While they are parasitic on spoken language and not learned as first languages, linguistic communication can be whistled (Meyer, 2008), drummed, hummed (Stern, 1957), and written.

Spoken languages seem to make use of a wide range of what is physically possible (Catford, 1977). Sounds which contrast words (phonemes) can be made by pushing air out of the lungs (pulmonic) or sucking air into the oral cavity (ingressive), with vocal fold vibration (voiced) or without (voiceless), with a fully open vocal tract (vowels) or by complexly blocking

air flow (plosives), with a closure at lowest point of the vocal tract (glottis: glottal) or the furthest (lips: bilabial), with one point of articulation (e.g., lip closure: [p]) or multiple (e.g., lip constriction and tongue dorsum at velar ridge: [w]; Ladefoged & Maddieson, 1996).

The International Phonetic Alphabet (IPA) was created with the intent to have a set of symbols which could be used to transcribe contrastive sounds in all of the world's languages. The initial version had 47 symbols for sounds and another 7 diacritics.<sup>3</sup> As more languages have been rigorously described and analyzed, the number of symbols has continued to grow with each successive version. In its most recent edition, that number has expanded to 112 symbols and 31 diacritics. Another 23 symbols are used to describe features which can extend beyond a single segment, such as tone or stress (International Phonetic Association, 2018). Even the current IPA chart does not have a symbol for a velar click because the sound was thought to be impossible. Clicks are produced with an anterior constriction made by the lips, tongue tip, or tongue blade and a posterior constriction made by the back of the tongue at the velum or uvula. Once the anterior and posterior constrictions are in place, the tongue body moves down. The movement of the tongue rarefies the air between the two constrictions. Once the anterior closure is released, pressure is equalized creating a loud, transient burst of energy (Ladefoged & Maddieson, 1996). Velar clicks were thought to be impossible because the tongue cannot form an anterior constriction at the velum coupled with a more posterior constriction while also leaving enough space to form an air cavity. However, recent work has found that speakers of Laal can create a velar click by releasing the velar constriction before the anterior closure (Ball, Howard, & Miller, 2018; Lionnet, To appear).

---

<sup>3</sup> The International Phonetic Alphabet is maintained by the International Phonetic Association. Originally named Dhi Fonètik Títcerz' Asóciécon, the group developed from 'L'Association phonétique des professeurs d'Anglais, a teachers organization. It gradually shifted to an organization for phoneticians (Macmahon, 1986).

In terms of variability in segment inventories, some sounds, like /m/, /k/, and /i/, are found in more than 84% of all spoken languages, while others like clicks, interdental, and labio-velar plosives are only used by a handful of languages (Maddieson, 2013a, 2013b; Maddieson & Precoda, 1989). The inventory of contrastive consonant sounds in languages ranges from roughly six in Rotokos to 95 in !Xu (Maddieson, 1984; Maddieson & Precoda, 1989). For vowels, many languages have only three, while !Xu has 24 (Maddieson, 1984; Maddieson & Precoda, 1989).

Sound sequencing, or “phonotactics”, is highly variable as well. Hawaiian doesn’t allow any sequences of consonants (Schutz, 1981). Georgian, on the other hand, can have sequences of eight consonants before a vowel is produced (Butskhrikidze, 2002). For most languages, a syllable nucleus can only be a vowel (Blevins, 1995). Other languages, like English, will allow sonorous consonants like /r/, /l/, /m/, and /n/ to fill that role. Imdlawn Tashlhiyt Berber is far less restrictive. In addition to vowels and sonorous consonants, Imdlawn Tashlhiyt Berber also permits obstruents to serve as syllable nuclei (Dell & Elmedlaoui, 1985).

### **1.2.2. Limits on variation in human language**

In spite of the diversity of human languages, some linguists have claimed that certain grammatical properties hold for all languages. Often these claims are abstract and can be difficult to falsify. For example, Greenberg (1962) claims that all languages have syllables. Gokana, a Niger-Congo language spoken in Nigeria, is an exception to this proposed universal (Hyman, 2011). As Hyman (2011) notes, though, it is possible to analyze Gokana as having syllable structure. No language-internal evidence motivates the use of the syllable as a meaningful unit in Gokana, but it is theoretically possible to force a syllable analysis on the language.

Over the last few decades, cross-linguistic research has expanded to include more languages from across the globe. While this research has found strong tendencies across

languages, researchers are also finding that substantive inviolable universals likely do not exist. In Table 1-1, are a few linguistic phenomena which have been proposed as universal. In each of these cases subsequent research has found counterevidence for universality.

***Table 1-1. Proposed human language universals.***

<b>Domain</b>	<b>Proposed claim</b>	<b>Proposed counterevidence</b>
Phonetics	Labial-alveolar plosives are not possible (Maddieson, 1983)	Yélfí Dnye (Ladefoged & Maddieson, 1996)
Phonology	Coronal phonemes are present in all sound inventories (Hyman, 2008)	Northwest Mekeo (Blevins, 2009)
Phonotactics	CV is the only universal syllable type (Jakobson & Halle, 1956, p. 37)	Arernte (Breen & Pensalfini, 1999)
Semantic	All languages have numerals (Greenberg, 1963)	Pirahã (Gordon, 2004)
Syntax	All languages have recursion (Hauser, Chomsky, & Fitch, 2002)	Pirahã (Everett, 2005)

Humans make use of a broad range of what is physically possible, but spoken languages are still constrained by human physiology. Some sounds which were believed to be physically impossible have since been discovered. Maddieson (1983) postulated that doubly-articulated labial-alveolar plosives, sounds made with a closure at the lips and synchronous closure at the alveolar ridge, were likely impossible. Simultaneous closure prevents an appropriate pressure buildup behind the lips which would make the labial release imperceptible. Maddieson argued that examples of labial-alveolar sounds were actually sequences of two sounds rather than a single unit. However, along with Peter Ladefoged, he later found that in Yélfí Dnye, a language on Rossel Island in Papua New Guinea, labial-alveolar sounds were produced with simultaneous closure (Ladefoged & Maddieson, 1996), invalidating his previous claims of impossibility.

In contrast to labial-alveolar sounds, coronal sounds are widespread in languages. All of the spoken languages in the UPSID database have at least one coronal in their inventory (Maddieson & Precoda, 1989). Paradis & Prunet (1991, p. 2) state that coronals have a special

status and that coronals may be a “default” place of articulation. Hyman (2008) includes the universality of coronals in his list of proposed phonological universals, stating that all spoken languages contrast at least one coronal segment with at least one other place of articulation. However, Northwest Mekeo, a language of Papua New Guinea, lacks coronal sounds (Blevins, 2009). The Northwest Mekeo consonant system has bilabial, palatal, and velar sounds but no coronal phonemes or allophones.

At the level of segment organization, Jakobson & Halle (1956) claim that CV, the syllable structure with a consonant in the onset and vowel in the nucleus, is the only universal syllable structure. Subsequent researchers have gone further, claiming that CV is the default syllable type and that consonant onsets are required (Itô, 1989; Prince & Smolensky, 2002). Yet, Arrernte, a Central Australian language, is a counter example to this universal. Syllables in Arrernte are preferentially VC (Breen & Pensalfini, 1999). While some syllables do have consonant onsets, stress assignment and morphological processes suggest speakers may mentally represent an initial vowel in these cases as well (Breen & Pensalfini, 1999).

Outside of sound patterns, absolute universals have been proposed, as well (Pinker & Bloom, 1990). However, many of these are more difficult to assess. Often these claims are highly abstract and heavily intertwined with theory. For example, Hauser, Chomsky, & Fitch (2002) claim that recursion is universal across all languages and may be the only property of human language that is not shared with other species. Recursion is the embedding of a phrase or clause within another phrase or clause, though specific definitions vary (Christiansen & Chater, 2015; Hauser et al., 2002; Koopman, 2014). Hauser et al. (2002) argue that recursion is the root of the infinite generative abilities of language. Daniel Everett (2005), however, claims that recursion’s centrality to language is overstated. He claims that Pirahã, an Amazonian language he studied for over twenty years, lacks recursion altogether. In a recursive language, a clause like “they left” can be embedded within a larger sentence, [I said that [they

left]]. Through testing Pirahã speakers, though, Everett found that Pirahã speakers use juxtaposition of clauses: e.g., “my saying, they leave”. Everett states that these juxtaposed clauses typically have slight pauses, further suggesting a lack of embedding.

Pirahã is a fascinating example of linguistic variability. Pirahã lacks color terms, barely uses pronouns, and violates another proposed universal: that all languages have numeral words (Everett, 2005; Gordon, 2004). At first glance, Pirahã seems to have a word for “one”, “two”, and “many”, *hói*, *hoí*, *aibaagi*, respectively. However, *hói* and *hoí* are used in a wider range of contexts. When all of these contexts are taken together, *hói* seems to mean “small size or amount” or “roughly one”, *hoí* means “larger size or amount” (Everett, 2005; Gordon, 2004).

At least one language violates each proposed universal in Table 1-1. However, while these universals are not absolute, they are evidence of strong tendencies across languages. In the next section I explore proposed explanations for why strong tendencies exist across languages.

### **1.2.3. Causes of uniformity and variation in human language**

Within the past several decades, a resurgence of work in linguistics has come closer to biological models in assessing potential explanations for recurrent properties. One model in particular, that of Evolutionary Phonology (Blevins, 2004, 2006a, 2008, 2009, 2015) has been successful in both demonstrating the degree of variation in sound patterns, and explaining recurrent limits on this variation. Where variation is limited and strong tendencies or recurrent patterns exists, there are multiple potential explanations (Blevins, 2004, 2006a, 2015):

- 1) Inheritance from a shared mother tongue
- 2) Parallel evolution in the form of parallel phonetically motivated sound change (where “phonetic” means “based on aspects of speech perception & production”)
- 3) Non-phonetic constraints on form & function, including: general cognitive mechanisms, potential phonological (non-phonetic) universals, emergent effects of feedback loops in the course of language acquisition/maintenance
- 4) “Non-natural” or external factors including: language contact, prescriptive norms, literacy, and second-language learning
- 5) Chance

These five categories can account for most recurrent sound pattern in human language. Putative universals and universal tendencies receive considerable attention, but any similarity between two populations could potentially be of interest.

Inheritance is the most straight-forward source of shared sound patterns. Two languages may share patterns simply because the pattern was passed down from a shared ancestor. Inherited systems, whether through cultural or genetic mechanisms, can naturally diverge given enough time and separation (Cavalli-Sforza, Feldman, Chen, & Dornbusch, 1982). Small, gradual changes will lead to major differences over time, but similarities may still exist. Sanskrit, Latin, and Greek are vastly different languages, yet similarities in sound-meaning correspondences are great enough to strongly suggest a common ancestor, Proto-Indo-European, spoken 5,000 years ago (Fortson, 2004; Schleicher, 1877).

The second factor in Blevins' model, phonetic factors, are based in articulation, acoustics, aero-dynamics, or the interactions between them. Final obstruent devoicing is an example. In multiple languages like Afar, Basque, Chadic Arabic, Fyem, Russian, Malay, and Turkish, restrictions have independently developed on producing sounds like /b d g/ at the end of a word (Blevins, 2006a). For instance, Russian speakers produce the voiceless counterpart [p] when /b/ is in word-final position: клуб /klub/ [klup] "club". But when not at the end of a word, the sound is [b]: клуба /kluba/ [kluba] "club's" (Wetzels & Mascaro, 2001). Integrating, and expanding upon, the aerodynamic work of Ohala (1997), auditory work of Steriade (1999), and the articulatory work of Vaux & Samuels (2005) and Ohala (1983), Blevins (2006) argues that final devoicing is the result of many phonetic factors. First, laryngeal spreading or closing gestures, which take the vocal folds out of a voicing configuration, may be associated with the end of prosodic domains. The lack of final release also reduces the perceptibility of voicing cues which could lead to the perception of a voiceless segment. Final lengthening often causes

devoicing of a segment, as well. Finally, the vocal tract configuration of obstruents is poorly suited to support voicing.

Non-phonetic constraints are the third potential source for recurrent sound patterns in Blevins' model. Non-phonetic explanations are those which are internal to humans but are not a result of the peripheral mechanisms on producing or hearing the signal. These may be more rooted in cognitive factors, like categorization, memorization, attentional biases, or even domain-specific abilities.

Innate cognitive properties that are specific to human language could also fall within this category. For some linguistic frameworks, innate cognitive mechanisms specific to the linguistic system are the main source of cross-linguistic similarities (Chomsky, 1965, 1966, 2007). Under this view, all humans are born with a language faculty which constrains what is potentially learnable from environmental input. Cross-linguistic variation is tightly linked to the absolute limits of what is possible (Baker, 2001; Chomsky, 2007; Pinker & Jackendoff, 2009). Within phonology, Optimality Theory (Prince & Smolensky, 2002) argues that an innate language faculty constrains certain sounds and sound combinations. Structures that are not subject to the constraints will be more common cross-linguistically. Constraints, however, can be overruled so the lack of absolute universals across languages is due to constraint violations. Evolutionary Phonology does not rule out innate, linguistic-specific mechanisms (Blevins, 2006a, 2007) as a potential source of recurrent sound patterns. However, linguistic specific mechanisms are not the null hypothesis. Unless compelling data exists, positing human and linguistic specific cognitive mechanisms is unwarranted. Non-phonetic explanations may be specific to humans or the linguistic faculty, but they need not be human- or domain-specific. In fact, many of the most basic properties of phonological systems, like the existence of mental categories, category classes, and unit sequencing, are based in general cognitive properties such

as pattern recognition and categorization (Bybee, 2010; Feigenson & Halberda, 2004; Kuhl & Miller, 1975).

External factors are the fourth potential source of recurrent sound patterns. External factors are those in which recurrent sound patterns arise through interaction with the social or physical environment. The clearest case would be similarities due to cultural contact, like the English initial consonant cluster of /vl/. This consonant cluster is found primarily in words of Russian origin, as in the name *Vladimir*. In the non-loan words English does not permit /vl/ in the onset.

External factors also include interactions with the physical environment. Humans, as well as a wide range of species, increase the amplitude of their voice when noise levels rise (Brumm & Zollinger, 2011). However, very little research has gone into exploring ecological effects on linguistic structure. So far, no convincing evidence exists that the environment plays a significant role in shaping languages. Caleb Everett and colleagues (2017; Everett, Blasí, & Roberts, 2016) have provided data suggesting drier climates have fewer vowels and utilize tone contrasts less frequently. However, multiple language researchers have critiqued the work for presenting a simplistic description of the linguistic data (Ladd, 2016; Progovac & Ratliff, 2016) and not properly controlling for historical population movement (Collins, 2016; Donohue, 2016; Moran, 2016).

The final source of recurrent sound patterns in Evolutionary Phonology is chance. Chance leaves open the possibility that sounds, or sound patterns could be similar due to randomness. The odds that any two languages would share a sound-meaning correspondence is quite high. In a sample of 68 independent languages, Ringe (1999) calculates the likelihood that any two of the languages would share a word is slightly higher than 90%.<sup>4</sup> For a real world example,

---

<sup>4</sup> Ringe calculates this for “CVC-roots”, that is consonant-vowel-consonant sequences, like English *dog* or *pat*. He calculates the probabilities for many other sound sequences to illustrate the high probability any two languages will share sound-meaning correspondences is well within the realm of possibility.

English and Mbabaram, an Aboriginal Australian language, both use the sound sequence [dɔŋ] to refer to members of the species *Canis lupus familiaris*. In both languages, this development was independent (Blevins, 2004, p. 49; Hale, 1964). However, the more populations that share a pattern, the less likely chance is a valid option.

### **1.3. Variation and uniformity in birdsong**

#### **1.3.1. Degree of variation in birdsong**

There are over 10,000 extant species of birds, roughly double the number of extant mammalian species (O’Leary et al., 2013; Prum et al., 2015). Since the extinction of the non-avian dinosaurs, birds have radiated across the globe occupying a variety of ecological niches and evolving a wide diversity of morphological and behavioral traits. One of the most obvious domains of behavioral diversity in birds is vocal behavior.

Many birds are vocal, and their vocalizations can vary drastically across species. Cassowaries can produce low booms which have a fundamental frequency as low as 23 Hz (Mack & Jones, 2003), almost below the range of human hearing (<20 Hz). Blue-throated hummingbirds can produce ultrasonic vocalizations, as high as 30 kHz, even though their auditory range does not extend higher than 7 Hz (Pytte, Ficken, & Moiseff, 2004). Songbirds typically vocalize with an open beak, but closed beak vocalizations have evolved in 16 distinct avian lineages, including in pigeons, ostriches, pheasants, and kakapos (Riede, Eliason, Miller, Goller, & Clarke, 2016).

Most birds are not vocal learners and only three taxa of birds show evidence for vocal learning: Parrots, songbirds, and hummingbirds (Araya-Salas & Wright, 2013; Bradbury & Balsby, 2016; Marler, 1991). Vocal learning seems to have evolved independently in all three branches. The closest related birds to hummingbirds are swifts and nightjars (Prum et al., 2015; Zhang et al., 2014), both non-vocal learning taxa. Parrots and songbirds are closely related, but songbirds are more closely related to the non-vocal learning suboscines (Prum et al., 2015;

Zhang et al., 2014). There is little evidence that suboscines are vocal learners (though see Kroodsma et al., 2013 for possible evidence). Parrots and songbirds either independently evolved vocal learning or the ancestor of parrots, songbirds, and suboscines was a vocal learner and suboscines subsequently lost the trait.

Vocal learning research has historically focused mostly on songbirds. Songbirds make up nearly half of all avian species and are found in most terrestrial habitats (Barker, Cibois, Schikler, Feinstein, & Cracraft, 2004). As early as the 18<sup>th</sup> century, Daines Barrington (1773) performed song learning experiments on common linnet males and stated that “[n]otes in birds are no more innate than language is in man, and depend entirely upon the matter under which they are bred [and] as far as their organs will enable them to imitate the sounds which they have frequent opportunities of hearing.”<sup>5</sup> Vocal learning research in birds rapidly developed in the 20<sup>th</sup> century. Konishi (1964), Marler (1970), and Thorpe (1958) found that acoustic deprivation severely impacted the production of species-typical vocalizations in juncos, white-crowned sparrows, and chaffinches, respectively. Non-vocal learning chickens, on the other hand, produce normal, species-typical vocalizations after being deafened (Konishi, 1963).

While vocal learning songbirds failed to produce normal song, song was often present, just in an aberrant form. Konishi (1965b) proposed the “auditory template model” of song learning. The basic version of the model states that songbirds are born with a crude template of their species-typical song. The template is refined through exposure and practice. The song learning process occurs in two discrete stages. In the first, the perceptual stage, the young bird listens to songs in its environment. If the bird hears a song that matches their crude template, they use details from the input to refine their internal template. The second stage, the sensorimotor stage, occurs at the start of the bird’s first breeding season, usually in late winter or early spring. In

---

<sup>5</sup> Commas in original removed for clarity and to match more closely with modern usage.

the sensorimotor stage, the juvenile practices the songs it memorized during the perceptual stage. The young bird practices more songs than will eventually be in its repertoire. Toward the end of the sensorimotor stage, the bird drops songs from its repertoire. The remaining songs will end up being the only songs it sings during the rest of its life; it's "crystallized" song. During the perceptual stage the bird does not produce song and during the sensorimotor stage the bird does not memorize new song. Under this model, variation across species in song output is largely the result of differences in the inherited template (Marler & Sherman, 1983). Within-species variation can be the result of variations in song exposure or genetic differences between individuals (Nelson, Whaling, & Marler, 1996). Once the bird has fully developed, within-individual variation is limited to the few songs the individual selected during its first spring.

In white-crowned sparrows and swamp sparrows the perceptual stage and the sensorimotor stage are separated by several months (Marler, 1970a, 1970b; Mooney, 1999). However, this is not true for all vocal learning birds. Zebra finches, for instance, have overlapping perceptual and sensorimotor stages, both of which occur in their first 100 days of life (Clayton, 1987; Mooney, 1999). Even within a species, the sensitive period for memorization can vary. Most zebra finches memorize song from stimuli presented during days 35-70 post-hatch (Clayton 1987). However, social factors influenced the learning window. Clayton (1987) found that individuals who had their tutors switched on day 35 were more likely to reproduce song learned before day 35. Similarly, Böhner (1990) found that zebra finches whose fathers were removed before day 35 learned song just as well as those who had their father for the full duration of their song learning period. Kroodsma & Pickert (1980) found that the memorization period in marsh wrens was related to the amount of daylight the hatchlings were exposed to. Marsh wrens born in August receive less acoustic input than those born earlier in the summer, but the perceptual stage stays open for August hatchlings until the next spring.

For species like zebra finches, swamp sparrows, and white-crowned sparrows, memorization is limited to first few months of life, after which no new songs can be memorized. These species are “closed-ended learners”. The temporal window for other species is more permissive. “Open-ended learners”, like mockingbirds and starlings, can memorize and practice novel songs throughout their life (Gammon & Altizer, 2011; West, Stroud, & King, 1983). Brenowitz & Beecher (2005) state that the learning window for vocal learning birds forms a continuum, with closed-ended learners at one end and open-ended learners at the other end. In between the extremes are birds like canaries and brown-headed cowbirds. Canaries learn novel units throughout their life, but do so only during breeding seasons (Nottebohm, Nottebohm, & Crane, 1986). Song in the male brown-headed cowbird seems to be less dependent on memorization of adult song than on the interaction between males and females and active selection of vocal units. Brown-headed cowbirds are divided into two subspecies, each of which have distinct songs. King & West (1983) housed male brown-headed cowbirds with females of the other subspecies. The males had never been exposed to song and female brown-headed cowbirds do not sing. However, the males produced song that matched the song of the female subspecies. Subsequent research found that females reinforced certain songs and song units (West & King, 1988). Female brown-headed cowbirds produce a visual display during some vocalizations which prompts males to reproduce those vocalizations more often.

Outside of the songbird clade, many parrot species fall on the less restrictive, open-ended, edge of the continuum. Adult female and adult male budgerigars, for instance, learn novel social calls when exposed to a new group of budgerigars (Farabaugh et al., 1994; Hile & Striedter, 2000). Hummingbird vocal learning research is less developed but there is suggestive evidence that at least some species may be open-ended learners (Araya-Salas & Wright, 2013).

The differences in learning windows, acoustic learning templates, and amount of learning permissivity results in a variety of different vocal outputs. House finch song may last a couple

of seconds (Mundinger, 1975) while canary song can last for over a minute (Williams, 2004). Budgerigar warble can last for 10 – 30 minutes (Farabaugh et al., 1992). The phrase repertoire size can range from roughly 2 – 4 phrases in the *Locustella* warblers (Szekely, Catchpole, Devoogd, Marchl, & Devoogd, 1996) to over 100 in mockingbirds (Howard, 1974). In budgerigars, the units that comprise song may never repeat so that both the number of songs and the number of phrases may both be unlimited (Farabaugh et al., 1992). The combination of song units can be strict such that a bird may have only one unique song type, like the white-crowned sparrow (King, 1972), or it could be flexible to allow infinite combinations, like the brown thrasher (Kroodsma & Parker, 1977).

To make more analogous comparisons to human language, variability of birdsong should be assessed across populations within a species. Swamp sparrow dialects all seem to operate within a relatively narrow range; all have the same basic units below the level of the song and they all string those units together into trills (Balaban, 1988; Marler, 1997). In the most well-studied songbird, the zebra finch, song can vary between individuals in terms of units (repertoire size, unit types, acoustic features, introductory sounds, etc.), audience (only female directed vs female and male directed), song-to-song stereotypy, and duration (Williams, 2004). One of the earliest indications that songbirds learned their song was the presence of dialects, that is, cross-population variation within species – typically geographically based. In most species, song varies gradually which results in a dialect continuum (Farabaugh, Brown, & Veltman, 1988). However, often there are radical changes between areas, like that of white-crowned sparrow dialects in California. Members of a sedentary subspecies of white-crowned sparrows will sing one song type in one area and another type in an area a few kilometers away, with an abrupt shift occurring at the “border” (MacDougall-Shackleton & MacDougall-Shackleton, 2001). House finches have different dialect types depending on the region of the United States. On the West Coast, syllables are shared but full songs are not and variation is

gradient rather than discrete (Bitterbaum & Baptista, 1979). In the Eastern populations, however, house finches have clear dialect boundaries. These populations share full songs with only slight variations (Mundinger, 1975).

Heterospecific mimicry is perhaps the most obvious evidence of within species variation. For some species, the copying of sounds from other species is rare. In Dobkin's (1979) survey of heterospecific mimicry, evidence for many taxa came from a single individual. Baptista & Morton (1981) found a single Lincoln sparrow who shared sounds and sound patterns of white-crowned sparrows. Eberhardt & Baptista (1977) and Kroodsma (1972) state that heterospecific mimicry in their study species is rare, song sparrows and vesper sparrows, respectively. Both hypothesize that the heterospecific mimicking individuals lacked conspecifics in their area, so they learned whatever was available. The rareness in these species is of interest because it suggests a much greater level of plasticity that could only be seen given the right set of circumstances. It also hints that a greater number of species may be able to mimic heterospecifics but may never encounter the appropriate conditions.

For other species, heterospecific mimicry is more common. The European starling, for instance, differs radically from one population to the next. The species is endemic to Central Eurasia yet, due to European colonization, has been introduced across the globe. Since it can mimic sounds from other species, starling song often reflects the sounds of the local bird populations (Hausberger, Jenkins, & Keene, 1991). Human speech mimicry has been documented for budgerigars (Banta Lavenex, 1999; Scanlan, 1999), African grey parrots (Patterson & Pepperberg, 1994; Pepperberg, 1990), European starlings (West et al., 1983), and Indian Hill mynahs (Klatt, 1974). An African grey parrot, Alex, was able to approximate many human speech sounds, including bilabial stops, despite not having lips (Patterson & Pepperberg, 1998). Even with a physical limitation Alex was able to find a strategy to match the acoustic input.

### **1.3.2. Limits on variation in birdsong**

When compared to human language, more evidence for strict limits in birdsong exists. Much of these data come from heterospecific song exposure in experimental settings. For example, swamp sparrows cannot learn to reproduce song sparrow song units (Marler & Peters, 1977), despite the two species being so genetically similar that viable hybrids survive in the wild (Marler, 1990). Chaffinches were not able to reproduce song of other species unless that species song had “the right tonal quality” (Thorpe 1958). Of course, even in these cases, the failure to reproduce some acoustic signals could have been affected by the experimental protocol (Marler, 1997).

In swamp sparrows, Podos (1996) and Lahti, Moseley, & Podos (2011) found that there is an optimal trill rate range. If the repetition of sounds is too fast or too slow swamp sparrows do not faithfully imitate trills. Motor limitations seem to constrain the ability to trill at high rates (Podos, 1996). Swamp sparrows reduced the rate of trilling or inserted pauses throughout song. For slower rates, swamp sparrows increased the trill rate producing faster trills than the model (Lahti et al., 2011). Trill rate is used to attract mates and higher trill rates are perceived as “sexier”. The authors suggest that rates were sped up so that the singer could indicate that he is a high-quality male. These data suggest that the trill rate ceiling is limited by the physical apparatus while the floor is limited by the need to produce high performance song.

For species that are more prolific mimics, limits on variability seem less rigid, but still exist. Zollinger & Suthers (2004) found that mockingbirds were able to mimic high repetition rate canary trills, but they were unable to mimic the trill duration. While canaries can maintain a 22 notes per second trill for several seconds (4.1-7.7 seconds), a canary-tutored mockingbird is only able to maintain this rate for half a second. Marsh warblers mimic at least 100 different species, yet there is little evidence of mimicry of low frequency signals (Dowsett-Lemaire, 1979).

### **1.3.3. Causes of variation and uniformity in birdsong**

At a basic level, variations in the vocal behavior of individuals can be the result of genetic factors or environmental factors. However, genes and environment are not dichotomous. Most behaviors are the result of genes and the environment working together. Research into avian vocal behavior, especially that of songbirds, has been crucial in developing our understanding how nature and nurture interact to build behavior. In this section, I will briefly explore factors that affect variability and uniformity in avian vocal behavior. I review factors that relate to genetic inheritance, the environment, and the interaction between the two.

#### **1.3.3.1. Genetic factors**

Variance that results from genetic factors is more obvious in species that do not learn their vocal repertoire. In non-vocal learners, acoustic input is not necessary for vocal development. For instance, McCracken & Sheldon (1997) found that the differences in acoustic structure of vocal units was closely related to phylogeny in non-vocal learning herons. Genetically inherited physiological traits are responsible for uniformity and variance within and across species. The acoustic signal is often directly tied to the size and shape of the vocal production organs (Fant, 1960; Fitch, 1999). Vocal morphology can be under intense selective pressure because vocalizations can serve as a proxy for fitness. Vocal tract resonances, or “formants”, can also provide receivers with fitness information. The spacing of formants are directly related to the size of the vocal tract (Bowling et al., 2017; Fant, 1960; Fitch & Kelley, 2000). The size of the vocal tract relates to overall body size; as such, potential mates, rivals, or predators can use formant spacing as an acoustic cue to judge body size (Reby et al., 2005). In fact, body size exaggeration has been proposed as a hypothesis for the descent of the larynx in humans (Fitch, 2000). In birds, Fitch & Kelley (2000) found that whooping cranes can distinguish formant variations in synthetic whooping crane calls, suggesting they could use the information as an index to body size. Elongation of the trachea has evolved independently at least six times

during avian evolution (Fitch, 1999). In birds with tracheal elongation, the trachea coils around the clavicle, thorax, or sternum often resulting in a trachea that is longer than the length of the bird. Fitch (1999) argues that lengthening the trachea will reduce formant spacing and, thus, can give an acoustic impression of a larger individual. Inherited physiological traits affect vocal learners as well. Eberhardt (1994) found that singing in male Carolina wrens was more metabolically costly than any activity other than flying. Because singing is energy expensive, potential mates or rivals can use song to judge the fitness of the singer. The inability to match trill rates, mentioned for swamp sparrows and mockingbirds in 1.3.2, is largely the result of inherited physical constraints (Podos, 1996; Zollinger & Suthers, 2004).

In vocal learning species, Konishi (1965b), Marler & Sherman (1983), and Marler (1997) have proposed that innate auditory templates may explain within-species uniformity. As reviewed by Marler (1997), an extreme version of a template model states that the template is highly detailed and pre-encodes species-universal song units and patterns. Richly-specified templates could explain why some species have units which appear in all populations and why some species can learn songs from only a handful of repetitions. Swamp sparrows, for instance, vary in the organization of song units but the basic units are consistent across populations (Balaban, 1988). Nightingales only need to hear a song fifteen times in order to learn it (Hultsch & Todt, 1989). Richly-specified templates are similar to the models in human language which argue that grammatical structures are innate and only a few exposures are needed to trigger learning (Baker, 2001; Chomsky, 1981; Prince & Smolensky, 2002).

However, for many songbird species, detailed templates seem inadequate to explain within-species variability. Under the right circumstances some songbirds, even species whose learning program seem highly restrictive, can learn song from other species (Marler, 1997; Marler & Peters, 1977). A template that specifies another species' song repertoire seems unlikely. As such, templates likely encode just enough information so that the young bird

recognizes its species-typical song (Marler, 1997). For example, Soha & Marler (2000) found that the species-typical introductory whistle in white-crowned sparrow song plays a role in song learning. White-crowned sparrows will not learn conspecific song that does not have the introductory whistle. They will, however, learn vocalizations from birds and mammals as long as the whistle precedes the vocalizations.

### **1.3.3.2. Environmental factors**

The most obvious source of environmentally-induced variation in vocal behavior is variation in acoustic input. The learning by instruction model assumes that birds have little to no innate information of their species-typical song. A strong form of the model assumes that species should be able copy any sound that is within their physiological range (Marler, 1997). Under this model, for any two individuals within a vocal learning species, the degree of similarity will be related to the similarity in acoustic input. For instance, house finches in Colorado share fewer song units in common with populations that are more distant than with those that are geographically closer (Tracy & Baker, 1999). The marsh warbler can mimic over 100 species and only seems to be limited by its physical ability to produce certain sounds (Dowsett-Lemaire, 1979).

### **1.3.3.3. Genes and the environment**

Behaviors can be classified as being the byproduct of genes or environment, but more often behavior is the result of genes as mediated by the environment. Innate vs learned is not sufficient to account for much of the developmental behavior seen in vocal learning birds. Even traits that are thought of as classically heritable, like body size, are subject to environmental effects during the developmental process (West-Eberhard, 2003).

Isolation experiments show that acoustic input is required to produce typical song for most vocal learners. Many birds also seem to have preferences for conspecific song, though, suggesting some preexisting knowledge (Marler, 1997). Marler (1970a) found that white-

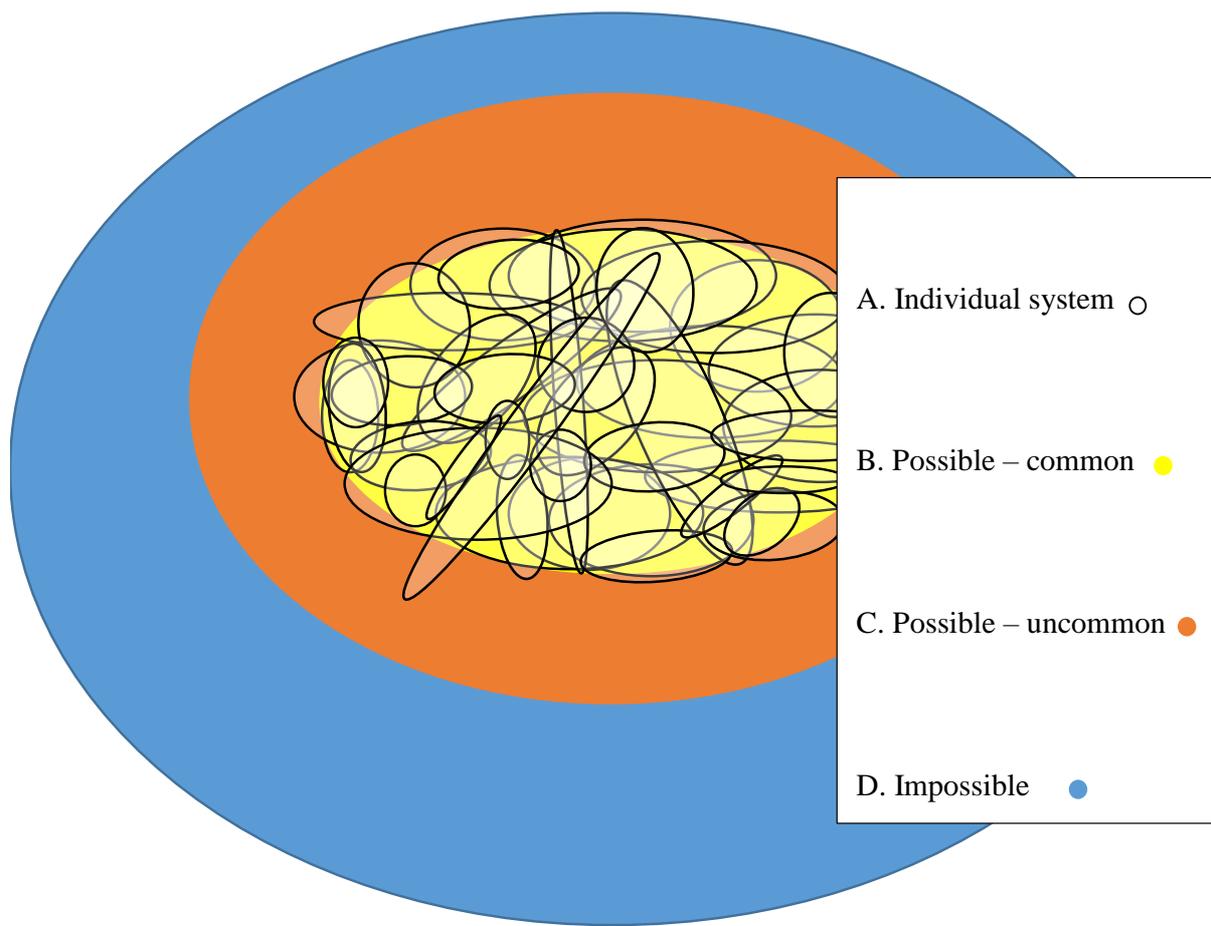
crowned sparrows preferentially learned conspecific song when exposed to tape recordings of several species. However, social interaction can affect the preference for conspecific song. White-crowned sparrows learned song from heterospecific tutors even if they had exposure to conspecific song, provided that the heterospecific tutor was the only bird the juvenile had social interaction with (Baptista & Petrinovich, 1984).

The interaction of genes, environment, and learning converge in the developmental stress hypothesis (Nowicki, Peters, & Podos, 1998). The developmental stress hypothesis suggests that song provides information about the singer's fitness because song is a reflection of the individual's post-hatch development, at least for close-ended temperate zone songbirds. The memorization stage of song learning occurs when the bird is reliant on its parents and when the individual is most at risk for nutritional deficiencies. The most fit individuals or the individuals with the best parents will be able to cope with developmental stress while simultaneously learning complex song. Female songbirds can use song to assess whether a male has good genes and/or good parenting skills.

Experimental evidence has confirmed that stress during development negatively affects adult song. Zebra finches who experienced restrictions in the amount of food during development produced less complex song (Zann & Cash, 2008). The affected zebra finches produced song units at a slower rate and produced lower peak frequencies. Brumm, Zollinger, & Slater (2009) found less accurate copying in food restricted zebra finches. Similar deficiencies have been found in canaries, starlings, blue tits, and swamp sparrows (MacDougall-Shackleton & Spencer, 2012). Variation in the post-natal stressors and variation in an individual's ability to cope with stressors lead to variation in song within a species. These variations are used by female songbirds to inform mate choice.

#### **1.4. Summary**

Any acoustic communication system can be conceptualized as fitting into a theoretical “design space” (Pinker & Jackendoff, 2009). The full design space includes all the theoretical communicative possibilities (see Figure 1-6). Of this infinite set, some systems will be impossible for a species, some will be possible but uncommon, and some will be possible and will occur regularly. As I reviewed in this chapter, for some species, what is common and what is possible overlap considerably; for others, there is likely a large domain of unattested possibilities that are learnable.



**Figure 1-6. Possible design space for learned acoustic communication system.**

- A) Individual system: An acoustic communication system (or part of an acoustic communication system) from an individual population.*
- B) Possible – common: The patterns, elements, features, etc. that most or many populations share.*
- C) Possible – uncommon: A system that is rare or could theoretically exist but doesn't.*
- D) Impossible: A system that could not possibly exist.*

Represented in Figure 1-6, an acoustic system or component of a system can fit into the domain of what is species-typical. Common systems include spoken languages that use both vowels and consonants (potentially universal), introductory whistles in white-crowned sparrow song (near universal; Soha & Marler, 2000), and frequency contour contrasts in the song of multiple songbird species (Lachlan, Verhagen, Peters, & ten Cate, 2010).

The orange domain in Figure 1-6 (the second ring from the outside) represents systems or traits that are possible but rare or unattested. These would include the preference for vowel initial syllables in Arrernte (rare; Breen & Pensalfini, 1999); mimicry of humans in parrots and corvids (occurs only under rare conditions), and a human language that uses an oral-auditory modality but has no periodic signals (unattested).

The final domain, the blue area in Figure 1-6 (outside circle), are systems or traits that are impossible. Theoretically they could exist, but physiology, cognitive limitations, or general physical properties of reality prevent their development. These could include a human system that uses only 100 kHz vocalizations or 10 Hz vocalizations (impossible to produce or hear), true labial sounds in birds (lack of lips), or precise mimicry of human speech in Macaques (lack vocal learning/lack of “speech-ready brain”; Fitch, Boer, Mathur, & Ghazanfar, 2016).

## **1.5. Structure of the dissertation**

### **1.5.1. Chapter two overview**

Language is often held to be one behavior, if not *the* behavior, that separates humans and non-human animals (Berwick & Chomsky, 2017). However, decades of research has shown that while animals may not show evidence of language in its totality, many, and maybe all, of the core properties of the human linguistic system are shared by other species (Fitch, 2010, 2018). Even if human language is unique, it is composed out of non-unique parts (Fitch, 2010). In this chapter, I discuss mechanisms of vocal behavior that are shared between birds and humans. I review neurological, genetic, perceptual, and vocal production mechanisms which converge in the avian and human lineages. The evidence presented in this chapter will show that cross-species comparisons are essential in understanding a behavioral trait. Cross-species comparisons are particularly relevant for the study of human language because researchers often make claims about the uniqueness of language without testing non-humans (Fitch, 2010).

### **1.5.2. Chapter three overview**

While budgerigar communication has been well studied, most of the focus has been on budgerigar contact calls. Very little work has been done on budgerigar song, “warble” (Farabaugh et al., 1992; Tobin et al., 2017; Tu et al., 2011). Unlike humans (and unlike house finches: Chapter 4) we do not yet have a broad basis for what might be typical in budgerigar warble. This chapter adds four independent budgerigar groups to the growing literature on budgerigar warble: two independent populations housed at the University of Vienna, a pair of pet budgerigars, and a population housed at the University of Maryland. A novel aspect of this research is the segment analysis. Budgerigars show evidence of segmental patterns that are more broadly shared, namely fundamental frequency falls, final lengthening, and initial plosive-periodic combinations. These patterns are shared among the four groups. There is also evidence that they are shared with other species, suggesting a non-species typical mechanism.

### **1.5.3. Chapter four overview**

The range of variation in acoustic communication within and across avian species is quite extensive, yet we are only beginning to understand the full extent of this variability. The house finch is an interesting model species to fit within the broader picture of songbird acoustic variability. The house finch has a moderately complex song, has expanded its range across all of temperate North America over the last 70 years, and there is a solid basis of wild recordings which permit us to address questions of uniformity and variation. While house finches do show cross-population differences in song organization and composition, all house finch songs seem to follow similar patterns (frequency modulated units, buzzes, 2-3 seconds in duration, etc.). Other wild recordings suggest house finches are even more flexible in what they can produce (Baptista, 1972). In this chapter, I analyze data from a cross-fostering experiment in which house finches were fostered by canary parents. Some individuals received canary song input, while others received no input. These data are then compared to each other and to species-

typical house finch and canary song. I find that canary-tutored house finches learn to trill as part of their song. Since trills are not present in typical house finch song studies of cross-population song patterns clearly underestimate what a species is capable of learning and producing.

#### **1.5.4. Chapter five overview**

Recurrent sound patterns are often argued to be the result a species and trait specific language faculty (Berent, 2013; Prince & Smolensky, 2002). For instance, linguistic-specific mechanisms have been invoked for the rarity of onset clusters of a sonorant and obstruent, like *rta* or *lba* (Berent, Steriade, Lennertz, & Vaknin, 2007). However, complex and abstract cognitive mechanisms are often unnecessary to account for cross-linguistic generalizations. In many cases, recurrent patterns develop as a result of common historical sound changes. Rare sound patterns, on the other hand, are often the result of uncommon sound changes (Blevins, 2004, 2006a, 2015, 2017). In this chapter, I review a rare sound pattern, sonorant + obstruent onset clusters, and the languages that permit these rare patterns. I investigate the historical development of sonorant + obstruent clusters, like *lba* and *rta*, to assess whether their rarity can be explained by historical factors. I find that the cross-linguistic and historical data suggest that the development of these rare onset clusters is the result of phonetic pressures and that linguistic specific cognitive mechanisms are not necessary to account for the typological asymmetry.

## **Chapter 2. Comparing humans and birds: Direct parallels in acoustic communication**

### **2.1. Introduction**

While no clear analog to human language exists in the animal kingdom, language is built off subsystems and behaviors which are widely shared across species (Fitch, 2017). As such, research into non-human communication has yielded important discoveries which have provided deep insights into the mechanisms of human language. Vocal learning birds have been at the center of this research (Bolhuis & Everaert, 2013; Doupe & Kuhl, 1999). Despite significant differences between humans and vocal learning birds, language and avian song shared deep neurological, genetic, perceptual, and anatomical parallels.

#### **2.1.1. Convergent evolution**

In human language, cross-linguistic observations and generalizations have been used to show historical relatedness among languages (Hock, 1991). Schleicher (1877) found similar sound-meaning correspondences between Sanskrit, Latin, and Greek, such as Vedic Sanskrit *tráyas* ‘three’, Latin *trēs* ‘three’, and Greek *treīs* ‘three’. The consistent similarities suggested the three languages shared a common ancestor, despite modern speakers being separated by thousands of kilometers. Similarly, cross-species comparisons can inform on relatedness among species and trait evolution. Homologous traits, those shared by hereditary descent, can inform on the trait’s ancestral form. A defining characteristic of birds is the presence of feathers. Since this is a trait shared by all members of the taxa, feathers were likely to have been present in the first birds. Paleontological research has confirmed that feathers pre-dated birds and were present in the last common ancestor of avian species (Brush, 2000).

Conversely, shared traits that do not result from a common ancestor can reflect similar historical pressures. In biology, these “analogous traits” are those that are similar between two species but were not present in the last common ancestor. For example, birds and bats have wings. Because wings are rare and the closest relatives of bats and the closest living relatives

of birds do not have wings (primates and crocadians, respectively; O’Leary et al., 2013; Kumar & Hedges, 1998), the bird-bat ancestor is unlikely to have had wings. Analogous traits are of particular interest because they can reveal similar evolutionary mechanisms. Independent evolution of a trait, or “convergent evolution”, suggests that the species either now, or at some point in their evolution, have similar functional needs, face similar environmental pressures, are subject to similar development constraints, or possess similar underlying architecture (Harvey & Pagel, 1991; Shubin, Tabin, & Carroll, 2009). In the case of wings, birds and bats both seem to have developed wings to achieve powered flight.

## **2.2. Neural similarities**

As reviewed by Emery (2006), early comparative cognition researchers were somewhat skeptical that avian brains could provide insights into “higher-level” cognitive abilities. Bird brains are smaller and lack the folding associated with the human neocortex. The outer layers of avian neural tissue were believed to be derived from the basal ganglia, a region considered to be primitive (reviewed in Emery, 2006). Recent research suggests that early skepticism of bird intelligence was unwarranted. Work from Jarvis et al. (2005) discovered that much of the avian forebrain emerges from the pallium, matter which is also responsible for the mammalian neocortex. Also, neurons are more densely packed in avian brains, so the smaller size is not reflective of lesser intelligence. In fact, the larger corvids and parrots have similar number of forebrain neurons as much larger primates (Olkowicz et al., 2016).

In terms of the neural anatomy responsible for vocal behavior, evidence for similarities between humans and vocal learning birds in neural structure comes from lesion research or brain activation techniques like PET and MRI imaging (Jarvis, 2013). Lesions affecting the left face motor cortex in humans and the nidopallial and arcopallial vocal nuclei in songbirds both disrupted learned vocalizations while leaving innate calls, cries, and screams relatively unaffected (reviewed in Doupe & Kuhl, 1999; Jarvis, 2013). In adult zebra finches, lesions to

the HVC region caused subjects to regress to producing juvenile-like vocalizations (Aronov, Andalman, & Fee, 2008). Lesions to the nidopallial vocal nucleus disrupted the production of mimicked vocalizations in human speech-mimicking budgerigars (Banta Lavenex, 2000). In both humans and vocal learning birds, vocal behavior is strongly lateralized. In zebra finches the right hemisphere is responsible for learned vocalizations, while in canaries it is the left side, as with humans (Nottebohm, 1977; Williams, Crane, Hale, Esposito, & Nottebohm, 1992). In humans and birds, if a lesion unilaterally affects the dominant side early in life, the opposite hemisphere can compensate for the injury and the species can still develop species-typical vocalizations (Hertz-Pannier et al., 2002; Nottebohm, 1977).

Crucially, the neural regions and pathways associated with vocalization are, in many ways, more similar between humans and vocal learning birds than between humans and primates or vocal learning birds and non-vocal learning birds. The three vocal learning avian taxa and humans all control their vocalizations via brain regions in the cerebrum and through pathways connecting three anterior and four posterior regions (Jarvis, 2007; Jarvis et al., 2000). These regions and pathways have not been found in non-vocal learning species, such as apes or chickens, whose vocalizations are only controlled in the midbrain and/or medulla (Kuypers, 1958; Wild, 1997).

### **2.3. Genetic similarities**

The search for the genetic underpinnings of language and birdsong have revealed similarities across vocal learning species, the most studied of which is *FoxP2* gene. *FoxP2* was discovered in a family that had high rates of languages disorders (Vargha-Khadem, Gadian, Copp, & Mishkin, 2005). The gene became famous because it seemingly caused linguistic impairment without affecting motor control or general cognition (Gopnik, 1990; Gopnik & Crago, 1991; Pinker, 1994, p. 323). Subsequent research found that disruption of the gene was not, in fact, local to language but caused deficits in general intelligence, articulation, and the

learning of non-linguistic orofacial movements (Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995). While initial interest may have overstated its importance for language, the gene has been found to be important for vocal learning (Vargha-Khadem et al., 2005). Mutations in *FoxP2* have been implicated in disruption of learned vocalizations in both humans (Vargha-Khadem et al., 1995) and avian species (Haesler et al., 2007). Haesler et al. (2004) found higher *FoxP2* expression in zebra finches during their song learning period. Haesler et al. (2004) also found rise and fall in *FoxP2* expression in canaries that correlates with their yearly open period of vocal learning.

Despite the importance of *FoxP2* in speech and birdsong (and possibly echolocation in bats, see Li, Wang, Rossiter, Jones, & Zhang, 2007) the gene is highly conserved across all tetrapod species. In fact, a comparison of 1,880 genes in humans and mice found that *FoxP2* was in the top five percent of most conserved genes (Enard et al., 2002). The human *FoxP2* gene and mouse *FoxP2* gene differ in only three amino acids. In humans, two of the amino acid changes occurred after the chimpanzee-human split (the orangutan is the only other primate to have undergone an amino acid change; Enard et al., 2002). Expanding to avian species there is a 98% convergence of the *FoxP2* proteins in humans and zebra finches (Haesler et al., 2004).

The findings of a role for *FoxP2* in song learning in zebra finches and canaries is the first evidence for “deep homology” in complex communication (Fitch & Mietchen, 2013). Deep homology is the sharing of a trait between two or more species that was not present in the last common ancestor but which has similar underlying genetic regulatory and developmental mechanisms (Shubin et al., 2009). *Hox* genes, for instance, play an essential role in the development of limbs in all arthropods (e.g., spiders, crustaceans, insects, etc.) and vertebrates (e.g., amphibians, fishes, mammals, birds, etc.), despite the fact that the last common ancestor

of these two groups was limbless (Shubin, Tabin, & Carroll, 1997).<sup>6</sup> In related species which do not share the trait, the underlying genetic structure or developmental pathways may only be slightly different. The cross-species similarities in *FoxP2* reveal common mechanisms which form the basis for complex communication. Only a few species have fully developed vocal-learning abilities, but the underlying framework seems to be present in many other species. Recent work in non-vocal learning rats suggest that, given the right conditions, rats show evidence for limited vocal learning (Arriaga, Zhou, & Jarvis, 2012).

#### **2.4. Similarities in sound perception**

Some of the strongest claims of human linguistic uniqueness have come from the domain of speech perception. The linguistic signal is referential and generative. In isolation, these characteristics are not unique, however the combination seems to only exist in humans. Many species have (functionally) referential signals, communicative units which apply to real world entities, but the inventories are typically limited in size and individuals do not create novel labels (Seyfarth & Cheney, 2003). Many whales, parrots, and songbirds have productive capabilities in that they can create novel songs by rearranging smaller units, but these signals do not seem to convey any semantic meaning (Marler, 2000). In human language, the listener must extract an acoustic signal that matches with a lexical meaning in a complex and rapidly fading signal. Because the number of potential labels is infinite, many units are likely to share acoustic similarities. Additionally, speech is not discrete and invariant, multiple units can exist within a breath group and units affect, and are affected by, the production of other units. All of these factors suggest that humans would need a highly specialized system for extracting meaningful units from the acoustic signal.

---

<sup>6</sup> The ancestor of vertebrates and arthropods would have been at the protostome-deuterostome split over 555 million year ago. According to the fossil record, the ancestors of both groups wouldn't have developed limbs until the Cambrian period at least 15 million year later (Erwin & Davidson, 2002).

Lieberman, Harris, Hoffman, & Griffith (1957) proposed that the perceptual mechanism of “categorical perception” evolved to aid in speech processing. Categorical perception refers to the tendency for listeners to respond to gradient acoustic stimuli as if the stimuli were categorical. Two tokens of a sound category could vary considerably in their acoustic signal but listeners have trouble perceiving them as different so long as they are both on one side of a perceptual boundary (Lieberman et al., 1957). For instance, voiced and voiceless plosives differ in the timing of vocal fold vibration relative to the stop burst. In English, the voice onset time (VOT) of the voiced /d/ is roughly equal to the timing of the burst, while the VOT for the voiceless /t/ is as much as 70 milliseconds after (Lisker & Abramson, 1964). In perceptual experiments where the voice onset time is manipulated in incremental steps, category perception does not change incrementally. Rather, listeners switch from one category to the other at ~20-30 milliseconds (Pisoni & Lazarus, 1974). The relevance of categorical perception for phonology seemed clear: sound categories distinguish meaning, “phonemes”, so the perception of the correct category is important, but speech is imprecise and variant. A perceptual mechanism which causes the listener to abstract away from fine-grained acoustic variations could help the listener more quickly ascertain linguistic meaning in the signal.

However, further research discovered that categorical perception was not unique to human language or speech. Cutting & Rosner (1974) found that humans had categorical responses to non-speech acoustic stimuli (see also Cutting, 1982).<sup>7</sup> Non-human mammals can also perceive continuous voice onset time differences as categorical (Kuhl & Miller, 1975). In birds, Dooling, Soli, et al. (1987) trained two budgerigars to respond to synthetic stimuli. One was trained to respond to stimuli with a voice onset time of 0 milliseconds and the other was trained

---

<sup>7</sup> Rosen & Howell (1981) found methodological issues with Cutting & Rosner's (1974) experiment which found that humans responded categorically to synthetic stimuli which morphed from a violin pluck sound to a violin bow sound. In response, Cutting (1982) adapted the methodology and still found categorical perception of violin plucks and bows.

to peck for stimuli with a VOT of 70 milliseconds. In the experiment, the budgerigars were given intermediate stimuli. Both individuals performed human-like in that their response rate did not shift with the VOT continuum, rather, they maintained a high response rate until completely switching to no response. Nelson & Marler (1989) found a similar effect with more naturalistic stimuli. Swamp sparrows have roughly six units, two of which are rapid frequency down-sweeps. These two down-sweeps differ in duration, one of which has a mean duration of 6.6 milliseconds and the other has a mean of 25.1 milliseconds. Synthetic units of intermediate length were not perceived as intermediate. Rather, swamp sparrows perceived the intermediate units as belonging to either the short or long category.

Other perceptual abilities have been postulated as specialized perceptual phenomena, such as “perceptual compensation for coarticulation”, the “perceptual magnet effect”, or “phonetic trading relations”, yet comparative research has yet to find any abilities which are human specific (Carbonell & Lotto, 2014; Lotto, Kluender, & Holt, 1997). In fact, many of these abilities are shared with avian species that are not vocal learners. Japanese quails, a non-vocal learning species, can learn to categorize human stop consonants even when there is no consistent acoustic cue shared in the stimuli, an ability that was thought to be a human adaptation for categorizing acoustically variable speech input (Kluender, Diehl, & Killeen, 1987). These comparative data suggest that speech processing is built off more general abilities.

## **2.5. Similarities in sound production**

Out of all of the traits that could be similar between humans and bird acoustic communication, the mechanisms of sound production would seem to be the least promising. Birds have beaks instead of lips, they lack a protruding nose or outer ear, and their most obvious sounds are either high pitched, melodious songs or harsh shrieks. Furthermore, humans, like other mammals, use the larynx to convert air into sound. Birds use the syrinx. The larynx is

located at the top of the trachea and first evolved in air-breathing fish to keep water from entering the lungs. This respiratory protection function has been maintained in most tetrapods, including humans and birds (Berke & Long, 2010). The use of the larynx for vocalization evolved later, but the structure has not changed greatly across species. Any adaptation in the larynx could comprise the larynxes ability to protect the respiratory system; as such, modifications have been modest, at least when compared to the syrinx (Hoh, 2010).

The syrinx is a novel structure which is only found in the avian lineage. This organ is found at the bottom of the trachea where the tube splits into the two bronchi. Unlike the larynx, the syrinx is almost exclusively used for sound production. As such, syrinxes are more varied across avian species than the larynx is across mammalian species (Garcia et al., 2017). For instance, while most birds have one set of vibrating tissues within the syrinx, many songbirds have two sets which allows them to produce two discrete sounds from either side of the syrinx (Nowicki & Capranica, 1986; Zollinger, Riede, & Suthers, 2008). The rufous ovenbird, a subsongbird, has three sets, one set at each of tracheobronchial junctions and another set which lies higher in the syrinx (Garcia et al., 2017).

While birds and humans use different vocal organs, the basic bio-mechanics of the syrinx and the larynx are quite similar. Humans and birds use a set of vibrating tissues to convert the expiratory air into sound. In humans, the vibrating tissues are the vocal folds (or vocal cords). For parrots and pigeons, the vibrating tissues are the lateral tympaniform membranes and in songbirds, the medial labia (Larsen & Goller, 1999, 2002; Suthers, 2001). In general terms, muscles within the syrinx or larynx bring the tissues together, air passes across the tissues causing them to vibrate, creating sound. The conversion of air into sound is called the “source”. The rate of tissue vibration is called the fundamental frequency ( $F_0$ ), which is perceived as pitch. More specifically, intrinsic syringeal or laryngeal muscles lightly adduct the tissues which causes pressure to increase below the closure. The increase in pressure forces the tissues

apart creating a “tissue wave” which travels along the caudo-cranial axis of the tissues, that is, from the bottom to the top. As the wave travels along the caudo-cranial axis of the tissues, elastic recoil brings the two sides back together. The superior (top) and inferior (bottom) of the tissues are out-of-phase with each other, creating a self-sustaining oscillation (Elemans et al., 2015). The self-sustained oscillations created by pressure changes and elastic recoil has been understood in humans and mammals for decades (Titze, 1994), but it has only recently been confirmed to hold for avian species (Elemans et al., 2015). In both mammals and birds, the mass of the tissues affects their rate of vibration. Tissues with higher mass create a source with a lower fundamental frequency.

The acoustic result of tissue vibration is not the only component of avian and human vocalizations. Both adhere to the “source-filter theory of vocal production” (Fant, 1960). The source-filter theory states that vocalizations are a composite of the output of two independent structures: the source and filter. The source signal can be modified by changing air pressure below the tissues or by adjusting the tension or placement of the vibrating tissues. The source signal is filtered by the vocal tract. Differences in the size and shape of the vocal tract influence how the source is filtered. Importantly, these two components can be modified independently and, for the most part, modifying one does not affect the other. For illustration, a human can change from the sound [a] to [i] by modifying their vocal tract, the filter, while maintaining the same F<sub>0</sub>, the source. Conversely, a human can also maintain [a], the filter, and can shift the larynx up to create a “high pitched voice” or can change the placement of the vocal folds to produce an irregular vibration pattern, the source.

The filter can modify the source input by creating constrictions in the vocal tract. The degree and location of the constriction can determine the overall amplitude envelope, the amount of energy at certain frequencies, and how widely distributed energy is across the spectrum (Fant, 1960). The vocal tract filter is one of the principle ways that humans create

sound contrasts in spoken language. Vowel sounds are created with little constriction in the vocal tract while plosives completely block airflow. Lowering the velum permits airflow through the nasal cavity, which is used to create sounds like [m] and [n] as well as the nasalized vowels found in languages like French and Portuguese. The placement of the constriction in the vocal tract contrasts sounds, as well. [b], [p], and [m] are made with a constriction of the lips while [d], [t], [s], [l], and [n] are made with the tongue blade touching the alveolar ridge.

The filter creates formant frequencies which are used to contrast vowel sounds in spoken languages. Formants are frequency peaks that result from the size and shape of the vocal tract. In any column of air – like a tube or the vocal tract – there are “natural frequencies” at which the air prefers to vibrate (Fant, 1960; Raphael, Borden, & Harris, 2011). These natural frequencies are determined by the size and shape of the air column. Sound passing through the air column will excite those natural frequencies while dampening other frequencies. Formants are a natural byproduct of sound in a column of air, as such, they should exist in non-human vocalizations (Fitch, 1997). In birds, evidence of formants has been found for numerous species, including song sparrows, black-capped chickadees (Nowicki, 1987), corn crakes (Budka & Osiejuk, 2013), and whooping cranes (Fitch & Kelley, 2000). Whooping cranes and corn crakes can perceive formants in their own calls, suggesting that formants may be biologically meaningful for at least some species of birds (Budka & Osiejuk, 2013; Fitch & Kelley, 2000).

Because pure tones are common in birdsong, many researchers were skeptical that the source and the filter operated interdependently for the majority of avian vocalizations (Casey & Gaunt, 1985). Nottebohm (1976) and Casey & Gaunt (1985) proposed a model in which avian pure tones are the result of a whistle-like mechanism. In whistles, the source is the result of the air passing through a narrow constriction and the fundamental frequency is tied to the volume of the resonant space behind the constriction (Casey & Gaunt, 1985). This mechanism was

ruled out by Nowicki (1987) who recorded several songbird species in a helium-oxygen chamber. Because the speed of sound is greater in a helium than in oxygen, the frequencies the result from resonances in a column of air should have been shifted up in frequency. Frequencies that are the result of contact between tissues, as in vocal fold vibration, will be unaffected.<sup>8</sup> In Nowicki's (1987) songbirds, the pure tone did not shift to a higher frequency, as would be expected with a whistle; rather, the fundamental stayed the same and harmonics appeared in the song. The presence of harmonics suggests that in order to produce pure tones, many songbirds configure their vocal tract to create a formant window which allows the energy at the fundamental to pass through while filtering out the harmonics.

Whether birds can modify their vocal tract to the same degree as humans is less clear. Beak gape has an acoustic effect on the acoustic output (Goller, Mallinckrodt, & Torti, 2004), but evidence for other articulators is less robust. At least some parrots and songbirds can modify the vocal source by lingual articulation. Experimental manipulation of a euthanized monk parakeet's tongue – using an artificial source signal – produced formant and amplitude changes (Beckers, Nelson, & Suthers, 2004). X-rays of live monk parakeets during vocalization showed that the tongue placement changes throughout the production of calls. Tongue position differs for different call types (Ohms, Beckers, ten Cate, & Suthers, 2012). In songbirds, Suthers, Rothgerber, & Jensen (2016) found support for lingual manipulation in northern cardinals. For sounds above 2 kHz, northern cardinals track the fundamental frequency of their song by altering the volume of the oropharyngeal-esophageal cavity (Riede, Suthers, Fletcher, & Blevins, 2006). However, from 2 kHz to 0.8 kHz the oropharyngeal-esophageal cavity does not change. Through x-ray cineradiography, the authors found that northern cardinals tongue

---

<sup>8</sup> This can be heard in humans when someone inhales helium and produces a “Donald Duck” voice; the fundamental frequency is unaffected but formant frequencies are shifted up.

elevation correlated with lower frequencies, suggesting that the tongue plays an active role in filtering the acoustic signal.

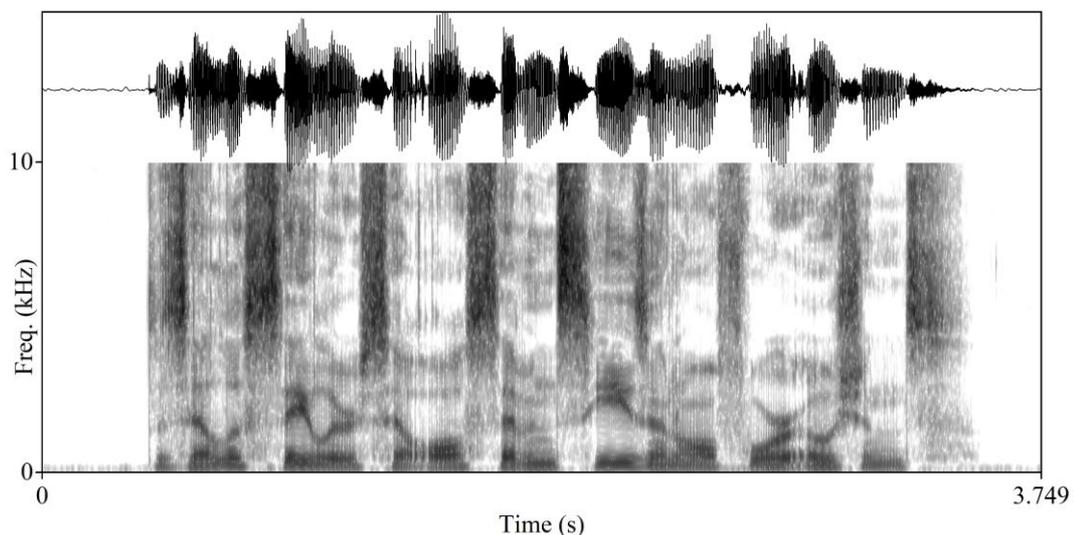
## **2.6. Summary**

Humans and birds share neurological, genetic, perceptual, and production mechanisms for vocal communication. The examples presented in this chapter are only a small subset of the research that has been conducted. Despite the extensive research there is still work to be done. Much of the work has focused only a handful of species, like the zebra finch and canary (Williams, 2004). Furthermore, structural comparisons have often been concerned with analogs for human syntax (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Bolhuis & Everaert, 2013), while similarities in phonetics, phonology, prosody, and phonotactics have received less attention (though this is changing, see Hoeschele & Fitch, 2016; Mol, Chen, Kager, & ter Haar, 2017; Yip, 2006). For instance, most avian vocal communication work has focused on elements divided by periods of silence (Kershenbaum et al., 2014), despite silence being only one of many cues which divide units in human language and speech. In the next chapter I start to close this gap by exploring units in budgerigar warble which seem to be analogous to the segment in human language.

## Chapter 3. Newfound uniformity: Human-like segments in budgerigar song

### 3.1. Introduction

One of the most essential, and difficult, tasks in understanding animal vocal communication is defining the units of production. Typically, units are defined relative to silences within the signal, with the most basic unit being uninterrupted sound (Kershenbaum et al., 2014). However, while the use of silence is useful, this approach may lose important information, particularly when comparing a system to human language. In humans, an English speaker can utter a long, complex phrase with no intervening silence, as in the phrase “the zealous sailors sail all seven seas and all four oceans”, seen in Figure 3-1. During speech, humans rapidly and actively modify their articulators in both the vocal tract (e.g., tongue) and larynx (e.g., cricothyroid cartilages). The result is that there are rapid shifts in one or more acoustic parameters; shifts that often mark perceptually discrete unit boundaries. These units are known as segments (Raphael et al., 2011).



**Figure 3-1. Human inter-silence interval.** In the speech stream, words and sounds are rarely discrete, stand-alone units. Rather, they are usually highly interconnected. The phrase “the zealous sailors sail all seven seas and all four oceans” contains many words and sounds and yet lacks silence intervals.

### 3.1.1. Segments in human language

In humans, research into segments and their organization (phonetics, phonology, phonotactics) has been one of the central focuses of the study of language and is arguably the area of linguistics where our knowledge of universal patterns, and their underlying mechanisms, is most advanced. For instance, extensive cross-linguistic typological research has revealed that, in spite of the great diversity of languages, all spoken languages have two broad classes of segments: plosives, transient bursts of energy (e.g., *p*, *d*, *k*), and vowels, periodic signals with clear harmonic structure that are typically made with little to no vocal tract obstruction (e.g., *i*, *u*, *a*; Hyman, 2008; Ladefoged & Maddieson, 1996; Lindblom & Maddieson, 1988; Maddieson, 1984). Lindblom & Maddieson (1988) suggest the universal use of these two segment classes is based in articulation, perception, and communication. Among the different segment classes, plosives and vowels are the most acoustically and articulatorily distinct from each other. Therefore, if speakers need to maximize perceptual contrasts alternations of plosives and vowels would be the most optimal pattern.

Furthermore, organization of these segment classes is asymmetrical, with plosive-vowel patterns, like *ka*, being near universal while the reverse pattern, *ak*, is much less common (related to the consonant-vowel preference and the Margin Hierarchy; Blevins, 1995; Breen & Pensalfini, 1999; Clements, 1990; Lowenstamm, 1996; Prince & Smolensky, 2004). Research into the plosive-vowel pattern has informed questions related to innate grammatical constraints (Chomsky, 1965; Prince & Smolensky, 2002); the importance of exceptions to putative universals (Breen & Pensalfini, 1999); the role of misperception in shaping common sound patterns (Blevins, 2004; Ohala & Kawasaki, 1984); the timing, production, and coordination of articulatory gestures (Nam, Goldstein, & Saltzman, 2009); and whether or not segments are distinct early on or emerge during development (Davis & MacNeilage, 1995; Giulivi, Whalen, Goldstein, Nam, & Levitt, 2011; MacNeilage, 1998).

Importantly, the two universal segment classes and near universal plosive-vowel pattern would be largely missed if researchers were to segment speech using only periods of silence, as is often done in animal acoustic communication research. Numerous species, particularly songbird species, have been described as having repertoire sizes that are infinite or impossible to determine (Macdougall-Shackleton, 1997; Read & Weary, 1992). These infinite repertoires are likely to be composed of a smaller set of more basic units, similar to how humans can build infinite sets of words from a finite set of segments.

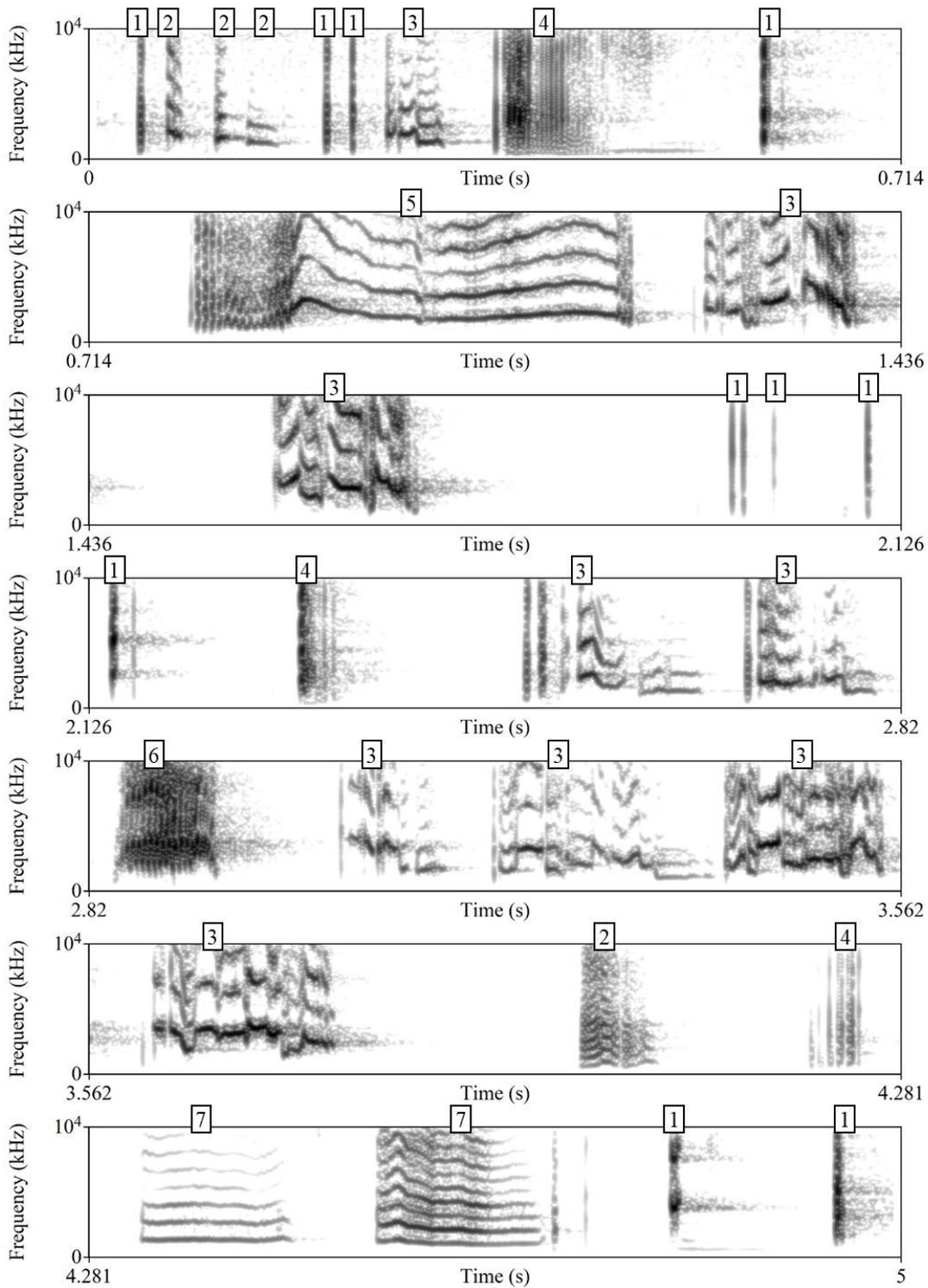
### **3.1.2. Segments in non-humans?**

Despite the importance of the segment in language science research, the segment has been largely unexplored in non-human species. Small units divided by acoustic transitions have been acknowledged as a potential level of analysis (Berwick et al., 2011; Du & Troyer, 2006; Kershenbaum et al., 2014), though most work has either focused on relatively simple concatenations (as in the banded mongoose: Jansen, Cant, & Manser, 2013) or extraordinary phenomena like double voicing (Zollinger et al., 2008). However, this level is essential in understanding complex acoustic communication in that it allows us to better understand how some species can have a seemingly infinite number of units in their vocal repertoire or how two individuals can possess vastly different acoustic signals yet still be clearly recognized as members of the same species (Fitch, 2006). Furthermore, at least in humans, because the production of segments is rapid and inter-connected, segments affect, and are affected by, surrounding segments. Recurrent contextual sound patterns may provide insight into articulatory or perceptual mechanisms which may promote those patterns (Blevins, 2004).

One species that is of particular interest for segmental analysis is the budgerigar, a small parrot native to the arid regions of Australia. The species has been used extensively as a model for vocal learning (Farabaugh et al., 1994; Heaton & Brauth, 2000; Hile & Striedter, 2000) and speech perception (Dooling & Brown, 1990; Dooling, Okanoya, & Brown, 1989; Dooling,

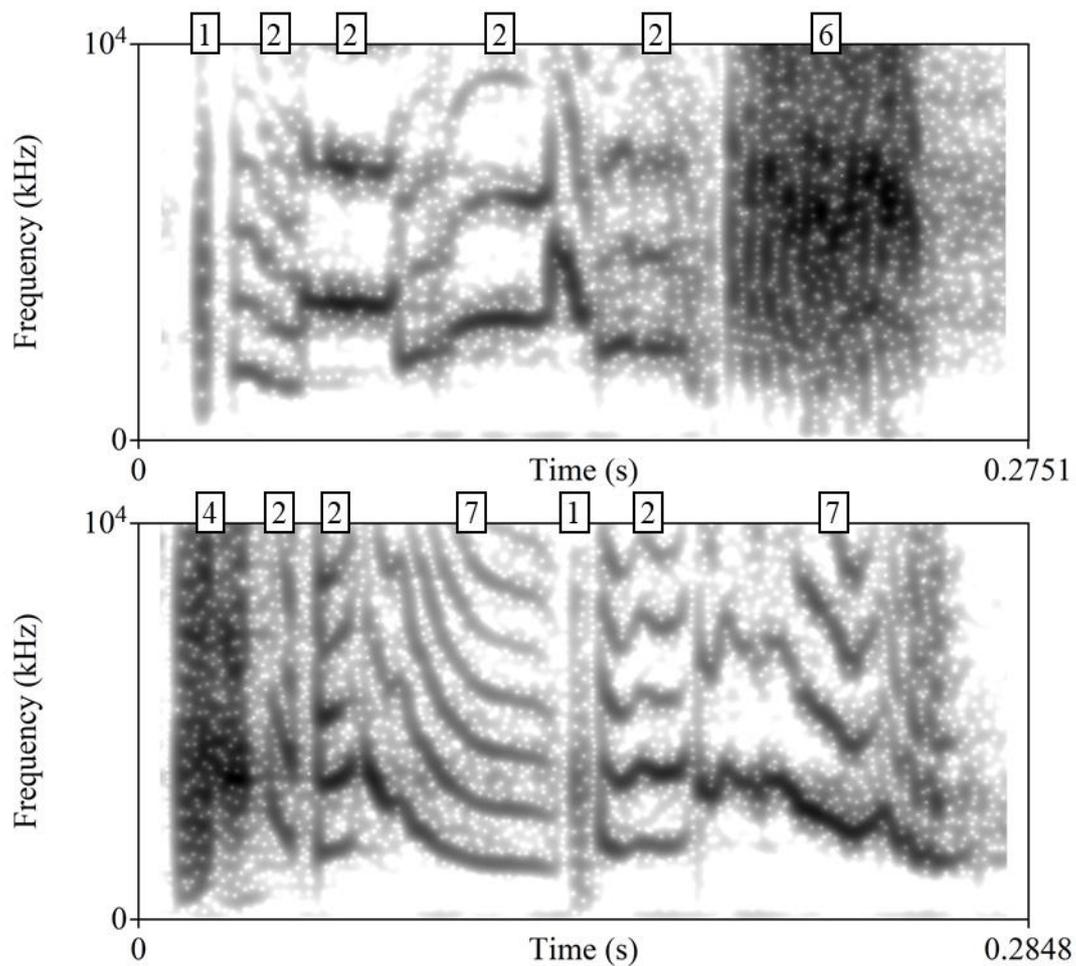
Park, Brown, Okanoya, & Soli, 1987; Hoeschele & Fitch, 2016), but its learned song, called “warble”, has received little attention.

Due to the complexity of warble, only a few published studies have done detailed analyses (Farabaugh et al., 1992; Tu et al., 2011). Tu, Smith, & Dooling (2011) found that phrases (“elements” in their terminology) could be classified into eight broad types: clicks, noisy, alarm call-like, short harmonic, long harmonic, pure tone, contact call-like, and compound (see Figure 3-2). The first six are relatively simple in that within-type tokens are more stereotyped and variation is limited to a few acoustic parameters such as duration. The final two types, however, are far more variable. Compound phrases are those which combine multiple phrase types, typically contact call-like phrases and one of the broadband (alarm call-like, noisy, or click) classes. The contact call-like phrases are the most frequent phrase type in warble. Typically, they are 100 – 200 milliseconds long, mostly periodic, and have high degrees of frequency modulation (Farabaugh et al., 1992; Tu et al., 2011). According to Farabaugh et al. (1992), “exact repeated renditions” of these complex phrases are incredibly rare.



**Figure 3-2. Budgerigar warble phrase types.** Five seconds of budgerigar warble with the phrase types described by Tu et al. (2011). (1) click, (2) short harmonic, (4) noisy, (6) alarm call-like, and (7) long harmonic phrase types have lower degrees of within-type variation. (3) contact call-like and (5) compound are less stereotyped. Phrase type (8), pure tone, does not seem to be widespread. It is the least frequent phrase type in Tu et al. (2011) and does not appear in the warble analysis of Farabaugh et al. (1992) or groups A, B, and C of this study.

A visual inspection of the two complex phrase types suggests that contact call-like phrases may also be a form of compound phrases. From the contact call-like phrase and compound in Figure 3-3, both types can seemingly be subdivided into units which broadly fit into the other phrase types.



**Figure 3-3. Complex phrases.** Two of the phrase types described by Tu et al. (2011) seem to be composed of units similar to the other phrase types. In the top spectrogram, the compound phrase type contains a click (1), several short harmonic sounds (2), and an alarm call-like element (6). The bottom spectrogram, a contact call-like phrase, contains a noisy element (4), a click, and long (7) and short (2) harmonic sounds. The final sound in the bottom spectrogram is labeled as a long harmonic, but it seems more complex and may be better described as multiple short harmonic elements with amplitude modulation, which is characteristic of the alarm call-like units, at the end.

Importantly, the contact call-like and compound units seem to be poor candidates for basic warble units. They lack stereotypy, are complex, and seem to be built from more simple elements. These factors make them prime candidates for a segmental analysis, though.

### **3.1.3. Comparative approach**

Humans are generally better than computers at recognizing patterns (Duda, Hart, & Stork, 2012), a fact that has been of great use in division and classification of animal acoustic signals, particularly with the aid of spectrograms (Catchpole & Slater, 2008; Janik, 1999). However, manual segmentation can be subject to human biases, can be functionally impossible with a large amount of data, and may attend to properties that are not biologically meaningful for the research species. These issues can be further compounded if the signal is being segmented by something less obvious than an interval of silence (Kershenbaum et al., 2014). Furthermore, in budgerigar warble some acoustic transitions are less clear-cut than others, as such, without prior knowledge of what is meaningful for the budgerigar, the magnitude of acoustic changes which mark segment boundaries is often unclear.

In this chapter, I designed a custom algorithm to segment phrases to make analysis more objective and to deal with a large dataset. To try to overcome the problem of biological relevance, I looked to human speech. Human speech is the most well-studied acoustic communication system and we have a much better idea of the acoustic changes that are biologically relevant. Despite their immense differences, human speech and other acoustic communication systems often make use of similar acoustic principles. Therefore, I created a lax segmentation algorithm which would take advantage of these similarities.

In most tetrapods, acoustic output is the result of three systems: respiratory, phonatory, and filter (Fitch & Hauser, 2003; Janik & Slater, 2000). Air is expelled from the lungs which is converted to sound by quasi-periodic tissue vibration (e.g., vocal folds in humans) and is then filtered by a vocal tract. The algorithm measures changes in three acoustic parameters

which could result from modification in these systems: amplitude, fundamental frequency, and Wiener entropy. Changes in amplitude can result from modification of pulmonic energy or vocal tract aperture. Changes in fundamental frequency, the physical correlate of pitch, are the result of adjustments of the vibrating tissues or the pulmonic energy input. Finally, changes in Wiener entropy, a measure of energy dispersion across the power spectrum, can result from modifications of the vocal tract, vibrating tissues, or pulmonic energy. Based on the similarities in physiological features, I believe an algorithm that successfully divides the signal of one species can successfully divide the signal of another species.

Our algorithm defines segment boundaries as large, rapid changes in fundamental frequency, amplitude, or Wiener entropy or smaller, correlated shifts between these acoustic parameters. I defined the specific magnitude of those changes by fitting the values to the manual segmentation of human phrases. In order to scale the algorithm to the faster and higher frequency song of budgerigars, I made the analysis window sizes dependent on fundamental frequency.

#### **3.1.4. Chapter organization**

The principle goal of this chapter is to assess whether I can find the segments of the unstereotyped budgerigar warble phrases. I built an algorithm to divide budgerigar phrase into segments based on a model trained on human phrases. If the algorithm is successful in segmenting phrases, I expect to find certain patterns. First of all, segments should be more stereotyped and form more distinct categories than phrases (Abler, 1989; Studdert-Kennedy, 1998). Furthermore, phrases should carry more cues to identity and segments should be more homogenous across populations (Marler, 2000). A secondary goal of this chapter is to assess whether segments vary by their position in the phrase. In humans, the production of segments is often affected by position, in part because of the rapid and interconnected nature of their

production (M. E. Beckman, De Jong, Jun, & Lee, 1992). I expect that budgerigar segments should similarly be affected by their prosodic context.

This chapter is comprised of three experiments. In the first, I create a segmentation algorithm to divide human speech into segments. Using the human data, I apply to algorithm to budgerigar warble. I then assess the status of segments as a basic unit by clustering segment and phrase data. In the second experiment, I test the reliability of segments and phrases at predicting population and individual identity. In the third experiment, I test the effect of position on the acoustic realization of segments.

## **3.2. General methods**

### **3.2.1. Data collection**

#### **3.2.1.1. Budgerigar warble data collection**

We segmented and analyzed the vocalizations of a total of fourteen budgerigars, thirteen males and one female, from four independent populations (groups A, B, C, & D). Individuals from groups A and B were recorded in their aviaries (Group A: 2.5 x 2 x 2 m; Group B: 2 x 1 x 2 m) in the Department of Cognitive Biology at the University of Vienna. The aviaries are located in separate, non-adjacent rooms at the university. Group B shares a room with another aviary with which it has acoustic, but not visual or physical, contact (no individuals were recorded from this other group). The rooms of both A and B are lined with acoustic foam padding (Basotect®30 Plain) to reduce echo and outside noise. The colony from which group A was recorded has a total of 12 budgerigars, six of which are male. The colony of group B has six individuals with three males. I was able to record seven individuals from group A and one from group B.

Group C is comprised of two pet budgerigars who were recorded at a home in Arkansas, USA. They were recorded in a metal wire cage (70 x 60 x 50 cm) lined with the same acoustic foam as with groups A and B.

We habituated groups A, B, and C to the presence of a human with recording equipment in their social environment and then opportunistically recorded individuals so that I could record song that is as close to their naturalistic performance as possible. These groups were recorded with an H4N Zoom recorder and a Sennheiser directional shotgun microphone at a sampling rate of 44.1 kHz. I mounted a GoPro Hero 4 to the top of the shotgun microphone. I recorded video (30 frames/sec) during the recording sessions in order to precisely identify the vocalizing individual.

The final four individuals (Group D) were recorded at the Laboratory of Comparative Psychoacoustics at the University of Maryland. The recordings of three of the individuals in group D were from archival recordings presented in Tu (2009), Tu, Smith, & Dooling (2011), and Tu & Dooling (2012). The final individual from Group D was recorded ten years later. All individuals were recorded under the same conditions; recording details can be found in Tu (2009), Tu et al. (2011), and Tu & Dooling (2012).

### **3.2.1.2. Human language data collection**

Because no data currently exist for budgerigar segments I used human speech to guide research into budgerigar warble. I used vocalizations from five historically unrelated languages: Chickasaw, Georgian, English, Vietnamese, and !Xóǀ. I chose these languages based on a combination of language relatedness, access to audio files with phrases, good signal-to-noise ratio, and speaker sex (Female: Chickasaw, Georgian & English; Male: !Xóǀ, English, & Vietnamese). All these factors serve to increase the diversity of acoustic signals. For example, males have, on average, a lower fundamental frequency than females, !Xóǀ has one of the largest segment inventories across all languages, Vietnamese and Chickasaw use F0 to differentiate words, and Georgian uses complex sound sequences. These factors also help to prevent overfitting a segmentation model to a specific language, language family, sex, or individual. With the exception of one English speaker, the files were collected from the UCLA

Phonetics Lab Archive (“The UCLA Phonetics Lab Archive,” 2007). In the archive, each sound file is accompanied by recording details and transcripts. Most of the sound files have speakers uttering short words, so I specifically looked through the transcripts for those that had longer utterances.

We used three English speakers. The first two were taken from the Vietnamese and Georgian sound files. In these files, English speakers often prompted phrases or described what was occurring in the recordings. The English speaker in the Georgian file was a native English-speaking female. In the Vietnamese file, the speaker was male and a non-native speaker, likely of Vietnamese, though the recording notes did not make the native speaker’s linguistic background explicit. The final English speaker was the author, a native speaker of American English. Those utterances were recorded in a semi-anechoic room at the University of Vienna using an H4N Zoom recorder at a sampling rate of 44.1 kHz.

### **3.2.2. Utterance/song division**

From the recordings of the budgerigars I extracted bouts of warble using a custom Praat script. The script used Praat’s *Annotate: To Textgrid (silences)* function to label sections of the recording as potential warble bouts. I used -45 dB for the amplitude threshold and one second duration for the threshold for silence, meaning that if amplitude was less than 45 dB down from the peak amplitude for longer than one second, the section was labeled as silence. The rest was labeled as a vocalization. Vocalizations that were longer than 2.5 seconds were labeled as warble. I manually coded those sections by individual and quality. To code individual, I cross-checked with the recording notes and video files. For quality, I excluded bouts where two or more individuals were vocalizing simultaneously, and I could not determine which vocalizations belonged to which individual. In some cases, I was able to extract sections of one individual vocalizing from these longer, multi-vocalization bouts.

We ran each warble through another custom Praat script that divided the warble into phrases. The script used a pass Hann band filter (*Minimum frequency: 1 kHz, Maximum frequency: 15 kHz, Smoothing: 100 Hz*) to exclude any noise outside of the typical budgerigar song range. It then created an intensity envelope by calculating the root-mean-square (RMS) of the sound pressure (window duration: 25 msec; time step: 5 msec). The algorithm identified phrases by checking for intervals where sound pressure RMS dipped below 1/6th of the song sound pressure RMS for longer than 10 milliseconds.

Our algorithm labeled phrases based on the Tu, Smith, & Dooling (2011) classification. I collapsed the compound and contact call-like phrases into a single “complex phrase” category. These complex phrases were extracted for segmentation.

### **3.3. Experiment 1: Basic units of warble**

Abler (1989) states that systems with infinite forms must be built from a finite set of discrete and invariant units. This “particulate principle” holds for many natural systems; for example, chemical elements are built from atoms, proteins are built from amino acids which are built from a small set of nucleotides. In systems with multiple levels, like human language, basic level units may be used to create an infinite set which is, in turn, used to create another higher level. Even if units in an intermediate level are not finite, these units are still more stereotyped and discrete than the units in the level above them. For instance, a language may have a set of 30 phonemes which speakers organize into an infinite set of words. Those words are used to create an infinite, and even more variable, array of utterances/sentences (Abler, 1989; Studdert-Kennedy, 1998).

The ability to generate novel units from more low level units has been cited as a trait that separates humans from other animals: human language is generative, animal repertoires are finite and static (Hockett, 1960). Studdert-Kennedy (1998) suggested – while offering the appropriate caveats given the state of parrot research at the time – that “parrots evidently imitate

speech in a continuous, non-particulate fashion, much as they imitate fire sirens and lawnmowers” (p. 209). Learned acoustic communication systems, like the warble of budgerigars and the songs of whales and songbirds, seem to adhere to the particulate principle at higher levels, as with the organization of phrase types within songs. Whether this principle holds for the creation of phrases, though, is unclear. For many species, the inter-silence interval seems to be the most basic unit (Kershenbaum et al., 2014). If the segment is an appropriate level of analysis for budgerigar warble, I expect that I should be able to segment budgerigar phrases and the resulting segments should be more stereotyped than phrases. If phrases are the most basic unit, I would expect to find inconsistency with segments and higher degrees of stereotypy at the phrasal level.

This experiment is divided into subsections. The first describes the segmentation algorithm that was used to divide budgerigar and human phrases. The second validates the segmentation output by clustering the segments and comparing them to phrase clusters.

### **3.3.1. Experiment 1a: Segmentation**

#### **3.3.1.1. Experiment 1a: Methods**

##### **3.3.1.1.1. Experiment 1a: Segmentation algorithm**

For automatic phrase segmentation, the algorithm took multiple measurements at regular intervals throughout the phrase for amplitude, fundamental frequency, and Wiener entropy. I made the algorithm easily scalable between humans and budgerigars by making measurement windows and sampling intervals dependent on a species minimum fundamental frequency. Minimum human F0 was set at 50 Hz and budgerigar F0 was set at 400 Hz, roughly the bottom F0 range for each species. For each acoustic measurement, the algorithm calculated the percent change between the acoustic measurement and the subsequent measurement. A percent change greater than a predetermined value marked a segment boundary. (The specific magnitudes are discussed below.) A second pass searched for smaller magnitude changes that are correlated

between intensity and either F0 or Wiener entropy. That is, a smaller scale change in amplitude may mark a segment boundary if within the same time window a small change in F0 or Wiener entropy is also present. The size of the window is determined by dividing 0.5 seconds by the minimum species F0. I didn't include correlations between F0 and Wiener entropy because a change in F0 is necessarily correlated with a change in Wiener entropy. Finally, to prevent the insertion of multiple boundaries associated with the same change, I added a buffer. The buffer is the same duration as the window size for correlations between acoustic changes, e.g., in a human phrase, a boundary cannot be inserted within 10 milliseconds of another boundary (0.5 seconds/50Hz).

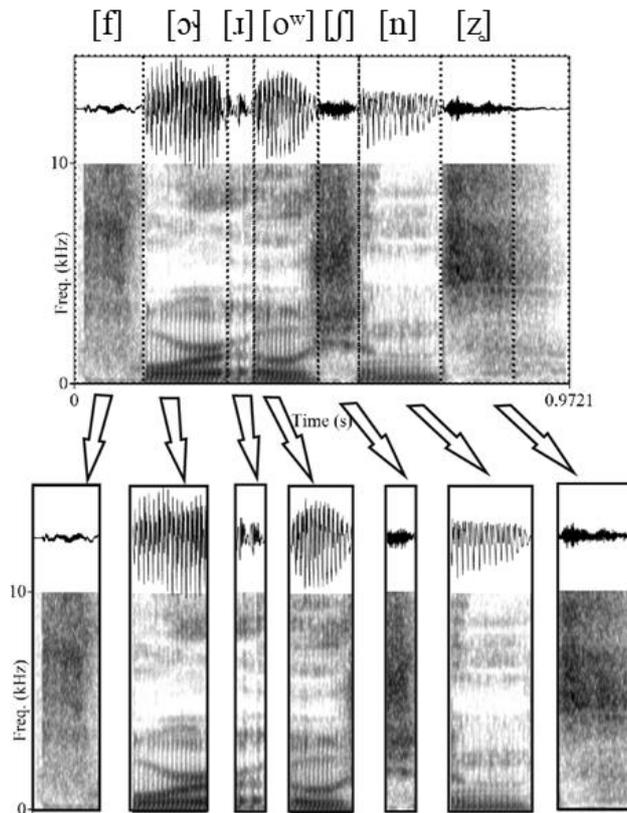
To find the specific thresholds which mark segment boundaries, I applied the algorithm to a manual segmentation of human language. I manually segmented and transcribed each human language phrase based on the conventions laid out in Keating, MacEachern, Shryock, & Dominguez (1994). I then adjusted the sensitivity of the algorithm, so the output would agree with the manual segmentation. Using smaller percent change values makes the algorithm more sensitive, potentially leading to false boundary insertions, while larger values runs the risk of missing segment boundaries. I adjusted the input values until the algorithm produced a mean accuracy rate of 65% for the human phrases, a rate comparable to published segment boundary detection algorithms (Dusan & Rabiner, 2006; Kuo & Wang, 2006), which is particularly high when considering that I am limiting ourselves to three acoustic parameters. I defined accuracy rate as the number of correct segments divided by the sum of correct segments, false insertions, and false deletions.

The most accurate model was one which had large transition values of 6 dB per frame for amplitude, 190 Hz per frame for F0, and 150 per frame for Wiener entropy. For smaller, correlated transitions the most optimal values were 2 dB per frame for amplitude, 7 Hz per

frame for F0, and 90 per frame for Wiener entropy. With those values I applied the algorithm to budgerigar phrases.

### 3.3.1.2. Experiment 1a: Results

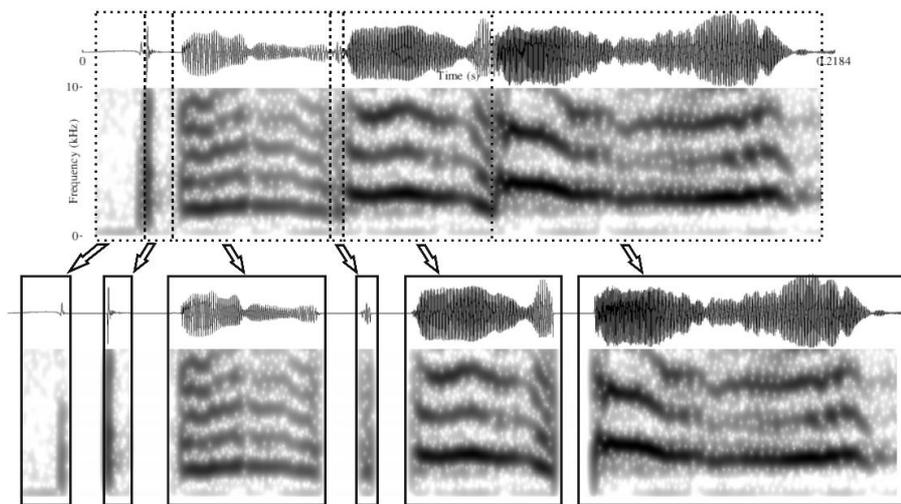
An example of the algorithm's segmentation of human speech is shown in Figure 3-4.



**Figure 3-4. Human speech segmentation output.** The phrase “four oceans” as divided by the segmentation algorithm. Top image shows the inter-connected speech phrase and bottom shows the segmentation.

As can be seen from Figure 3-4, the segmentation algorithm divided the interconnected speech into the major speech segments. The algorithm missed some glide transitions but divided non-glide consonants from vowels.

An example of the segmentation algorithm as applied to budgerigar warble is shown in Figure 3-5.



**Figure 3-5. Budgerigar warble segmentation output.** Spectrogram showing a budgerigar phrase (top) and the division into segments (bottom).

As shown in Figure 3-5, the algorithm successfully located moments of rapid acoustic transitions and divided the phrase based on those transitions.

### 3.3.2. Experiment 1b: Cluster analysis

#### 3.3.2.1. Experiment 1b: Methods

##### 3.3.2.1.1. Experiment 1b: Preliminary human language analysis

I performed a cluster analysis on human language data so that I would have a clearer idea of what to expect from the results of the warble segmentation. I used the recordings described in 3.2.1.2 as input for the clustering analysis.

I took fifteen random phrases from each of the seven speakers. As in the previous case, I manually segmented the phrases based on the guidelines in Keating et al. (1994). I then broadly labeled each segment for manner (vowel, glide, approximate, nasal, fricative, affricate, stop, and click), place of articulation (labial, coronal, palatal, velar, glottal, high/mid/low, front/central/ back), and whether the segment was voiced or voiceless, nasalized, glottalized, a rhotic, a lateral, a tap, or a trill.

I ran the phrases and segments through a Praat script which extracted acoustic parameters. I chose only a few parameters that could be relevant at both the segmental and phrasal level: duration, intensity, mean fundamental frequency, standard deviation of fundamental frequency, F1, F2, center of gravity, and spectral standard deviation. I took a subset of the segments, sampling equally from individual and segment manner, so that the number of segments would equal the number of phrases, 105. I scaled the acoustic parameters and clustered the units using the function *eclust*(*hc\_method* = “ward.D2”, *hc\_metric* = “spearman”) in the R package *factoextra* (Kassambara & Mundt, 2017). Using the *factoextra* function, *fviz\_silhouette()*, I ran silhouette analyses and calculated the average silhouette score for cluster sizes from two to eight. In a silhouette analysis, each unit is placed in a cluster and gets a silhouette value based on the distance to other units within its cluster and to other units in the next nearest cluster. Silhouette values range from -1 to 1. Negative values mean the unit was likely misclassified, numbers closer to 1 suggest that the unit is in a tight and non-overlapping cluster, and a value close to 0 suggests the unit lies between two clusters (Rousseeuw, 1987; Wadewitz et al., 2015). For each cluster size, I obtained an average silhouette value. More basic units would be expected to repeat and therefore cluster together more clearly which should lead to higher silhouette values. Units with subunits should show more acoustic overlap as they may have some subunits in common leading to lower silhouette values. This is expected to be true even if the number of clusters in the data is not necessarily optimal.

#### **3.3.2.1.2. Experiment 1b: Warble clustering**

For each budgerigar segment and phrase, I measured and extracted 21 acoustic variables (listed in Table 3-1).

**Table 3-1. Acoustic variables measured.**

<b>Mean Fundamental frequency (F0):</b> Frequency. Measured in Hertz (Hz). Result of (quasi) periodic vibration of labia (lateral tympaniform membranes; Larsen & Goller, 2002) in the syrinx (vocal source in avians). Measured by creating a pitch contour from Praat's <i>To Pitch (ac)</i> ... function (along with all other F0 calculations) then calculating the mean frequency for the sound by using <i>Get mean</i> .... The advanced settings were the same as in the division algorithm, except that the minimum frequency (which determines the analysis window) was set to a standard species minimum of 500 Hz: <i>To Pitch (ac): 0, 500, 15, "no", 0.03, 0.45, 0.05, 0.15, 0.04, 10000</i> .
<b>Minimum F0:</b> Frequency. Hz. Lowest F0 value found in frequency contour. Praat <i>Pitch</i> object function: <i>Get minimum</i> ....
<b>Maximum F0:</b> Frequency. Hz. Highest F0 value found in frequency contour. Praat <i>Pitch</i> object function: <i>Get maximum</i> ....
<b>F0 range:</b> Frequency. Hz. Difference between maximum F0 and minimum F0.
<b>Start F0:</b> Frequency. Hz. F0 at unit beginning. Measured by taking the Praat <i>Pitch</i> object, starting from the onset of the sound and searching until Praat's F0 calculation was able to find evidence of voicing.
<b>Mid F0:</b> Frequency. Hz. F0 at the halfway point of the unit. Measured by finding the midpoint of the unit and then taking the F0 measurement from the Praat <i>Pitch</i> object.
<b>End F0:</b> Frequency. Hz. F0 at unit end. Measured by taking the Praat <i>Pitch</i> object, starting from the offset of the sound and searching backwards until Praat's F0 calculation was able to find evidence of voicing.
<b>F0 slope:</b> Frequency. Hz. The difference between the start and end F0. Negative numbers mean F0 falls over the course of the vocalization.
<b>F0 slope/time:</b> Frequency. Hz/msec. Slope corrected for duration. It is calculated as the difference between the fundamental frequency at the unit start and unit end divided by total duration.
<b>Time-frequency excursion:</b> Frequency/temporal. Hz. The sum of frequency modulations divided by the unit duration. Calculated by taking the Praat <i>Pitch</i> object and summing the absolute difference between each frequency measurement. The sum is then divided by the unit duration.
<b>Jitter:</b> Frequency. Percent. Perturbations/deviations in the fundamental frequency. Calculated with Praat's <i>PointProcess (periodic, cc)</i> ... and <i>Get jitter (local)</i> ... functions. I set <i>Period floor</i> at 0.0001, the <i>Period ceiling</i> at 0.00167, and <i>Maximum period factor</i> at 1.3.
<b>Shimmer:</b> Amplitude. Perturbations/deviations in amplitude. Praat <i>PointProcess</i> object function <i>Get shimmer (local)</i> .... Same input parameters as jitter, plus <i>Maximum amplitude factor</i> at 1.6.
<b>Duration:</b> Temporal. Seconds. Amount of time the unit lasts. Praat's <i>Get total duration</i> ... gives a measurement in seconds so I converted to milliseconds by multiplying by 1000.
<b>Time to maximum amplitude:</b> Temporal. Percentage. Calculated by taking time point of the maximum amplitude ( <i>Get time of maximum</i> ...) in the unit and dividing by the total unit duration.
<b>Periodicity:</b> Quality. Present/Absent. If a (quasi)-periodic signal is present in the unit. Uses the output of Praat's <i>Pitch</i> object.

**Wiener entropy:** Spectral. Unitless. A measure of how much energy is spread across the sound spectra. I calculated Wiener entropy by dividing the sound into 10 msec windows, with a 9 msec overlap, and calculating the amount of energy at each frequency bin (100 Hz). I then calculated the geometric and arithmetic mean energy across the bins for each 10 msec window slice. I took the logarithmic score of the geometric mean divided by the arithmetic mean so that white noise (energy at all frequencies) is 0 and a pure tone is negative infinity. The final measurement is the mean Wiener entropy score for each window in the sound. Since Praat does not have a built in Wiener entropy function, I built my own (with aid from Gabriel J. L. Beckers Wiener entropy script: [http://www.gbeckers.nl/pages/praat\\_scripts/wiener\\_entropy.praat\\_script](http://www.gbeckers.nl/pages/praat_scripts/wiener_entropy.praat_script)).

---

**Center of gravity:** Spectral. Hz. The average frequency over the whole spectrum of a sound weighted by the spectrum. Center of gravity is calculated such that for a sine wave the center of gravity is the same as the frequency of the sine wave, while the center of gravity for white noise is half of the Nyquist frequency. I used Praat's *Spectrum* object function *Get centre of gravity...* with a *Power* setting of 2.

---

**Standard deviation:** Spectral. Hz. The standard deviation in the center of gravity. *Spectrum* object function *Get standard deviation...* with a *Power* setting of 2.

---

**Skewness:** Spectral. A measure of the symmetry in the spectral distribution, that is how different is the energy distribution above and below the center of gravity. *Spectrum* object function *Get skewness...* with a *Power* setting of 2.

---

**Kurtosis:** Spectral. A measure for how different the energy distribution across frequency bins (centered on the center of gravity) is from a Gaussian distribution. *Spectrum* object function *Get kurtosis...* with a *Power* setting of 2.

---

**Intensity:** Amplitude. dB. The acoustic correlate of loudness. Measured with *To Intensity...*

---

For budgerigar warble, I randomly selected a subset of segments ( $n = 840$ ) and a subset of phrases ( $n = 840$ ), sampling equally from each individual (60 was the minimum number of complex phrases for the budgerigars). For each unit, I used the measurements listed in Table 3-1.

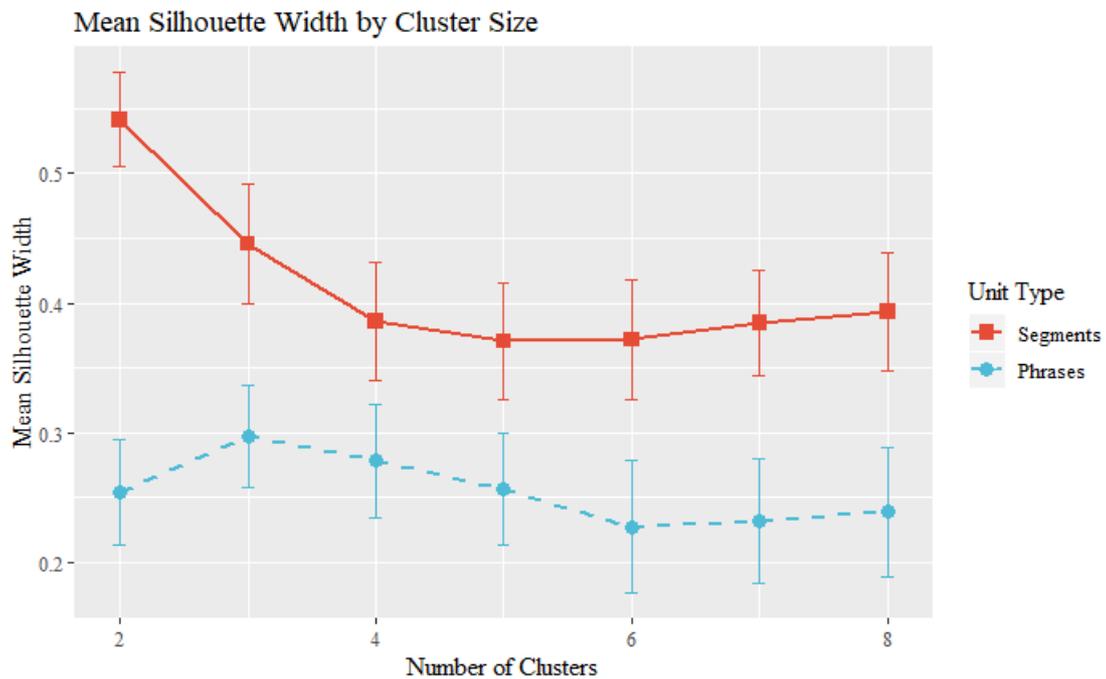
As with the human data, I scaled the data for each acoustic parameter and performed a hierarchical clustering using the function `eclust(hc_method = "ward.D2", hc_metric = "spearman")` in the R package *factoextra* (Kassambara & Mundt, 2017). Using the *factoextra* function, `fviz_silhouette()`, I ran silhouette analyses for cluster sizes from two to thirty.

### 3.3.2.2. Experiment 1b: Results

#### 3.3.2.2.1. Experiment 1b: Human language

For the human language data, shown in Figure 3-6, silhouette scores were higher for segments for every cluster size. Even the best performing cluster size for phrases, three clusters,

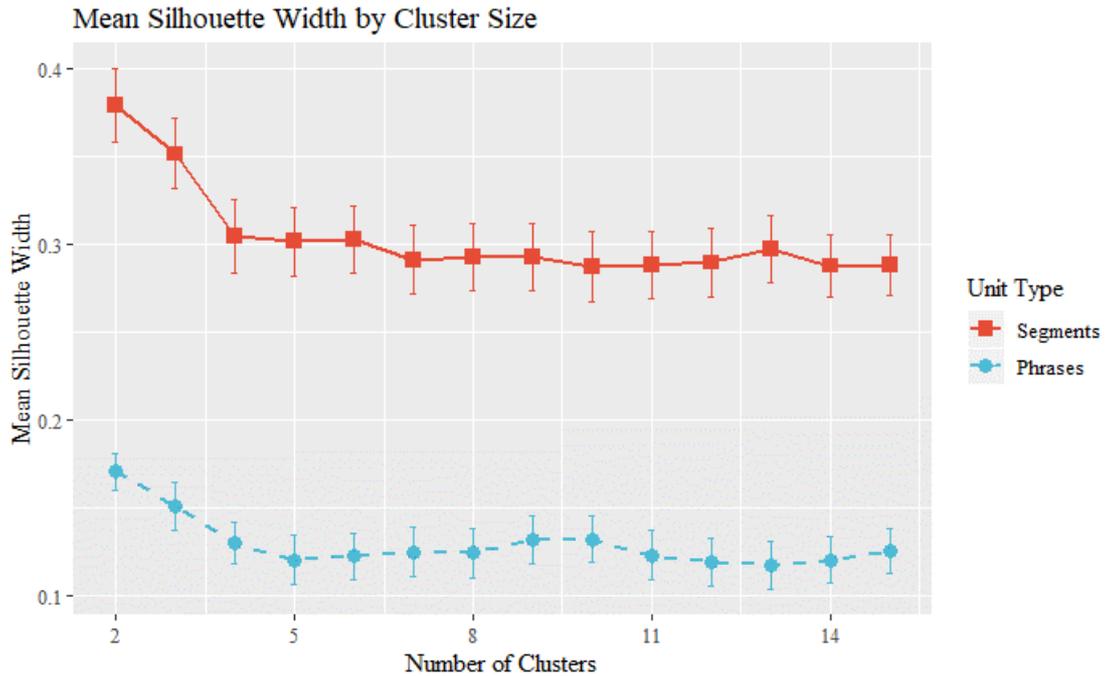
had a lower mean silhouette width than the worst performing cluster size of segments, 0.371 for five clusters.



**Figure 3-6. Silhouette values for human segment and phrase cluster sizes.** Each unit in the segment and phrase datasets receive a silhouette value based on within and across cluster distances. A score of 1 means the unit is a member of a well-defined and compact cluster, a score of 0 means the unit lies between two clusters, and a negative value means the unit is likely placed in the wrong cluster. The average of silhouette values provides a metric for how well the data clusters. I tested cluster sizes from 2 to 8. For humans, segments have higher mean silhouette width values for all clusters. Error bars are the standard error of the mean.

### 3.3.2.2.2. Experiment 1b: Budgerigar warble

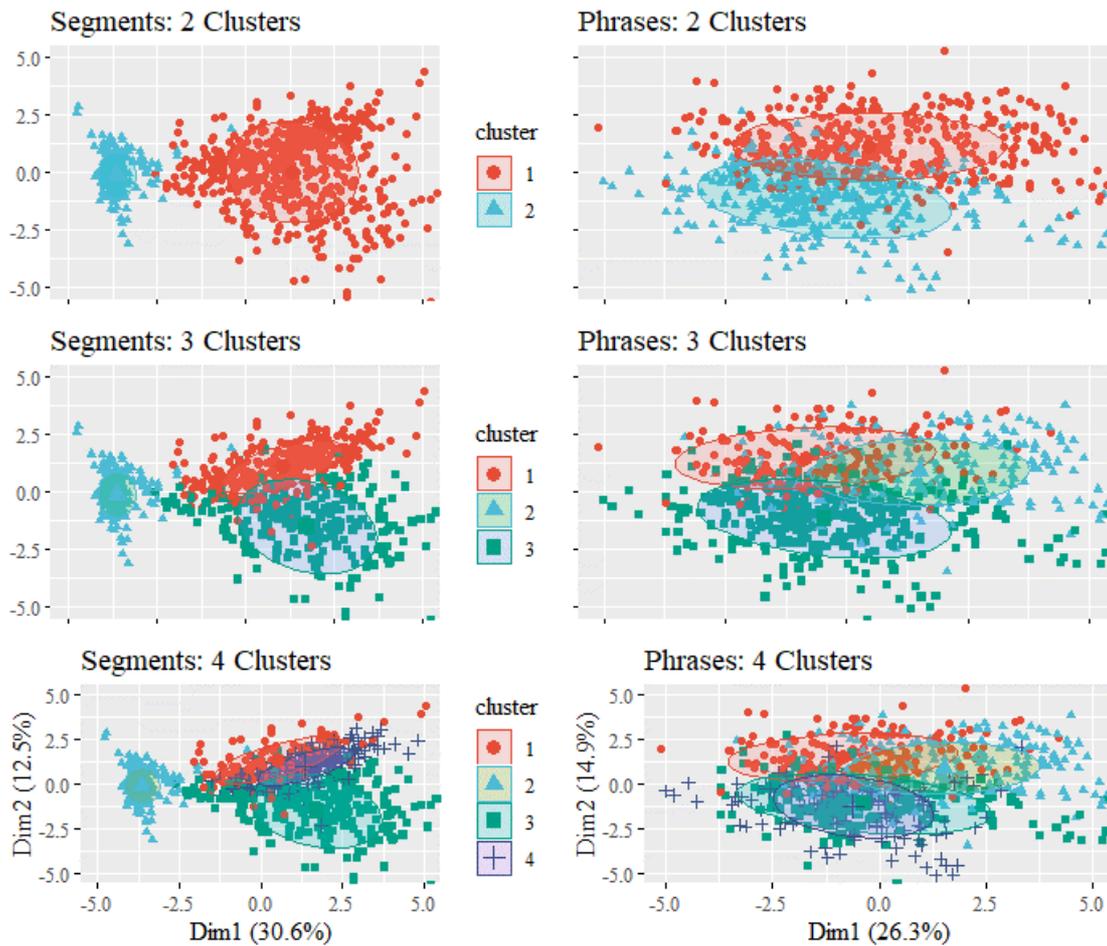
Mean silhouette widths for segments and phrases at different clusters sizes are shown in in Figure 3-7.



**Figure 3-7. Silhouette values for budgerigar segment and phrase cluster sizes.** Each unit in the segment and phrase datasets receive a silhouette value based on within and across cluster distances. A score of 1 means the unit is a member of a well-defined and compact cluster, a score of 0 means the unit lies between two clusters, and a negative value means the unit is likely placed in the wrong cluster. The average of silhouette values provides a metric for how well the data clusters. I tested cluster sizes from 2 to 15. For budgerigars, segments cluster better than phrases. Error bars are the standard error of the mean. Further silhouette comparisons can be found in 7.1.

At every cluster size, budgerigar warble segments produced more clearly defined clusters. Mean silhouette scores for segments are higher than 0.25 for all cluster sizes and reach a maximum 0.38. The highest silhouette score for phrase clusters is 0.17.

Figure 3-8 shows how individual warble units were categorized for clusters sizes of 2, 3, and 4.



**Figure 3-8.** Segment and phrase cluster sizes of 2, 3, and 4. The function `fviz_cluster()` performs a principal components analysis and plots the data points with the cluster information. Ellipses represent a multivariate normal distribution with a .5 confidence interval.

Categorization of warble segments into two clusters produces clusters with very little overlap, as seen in the top left corner of Figure 3-8. These two clusters were periodic and aperiodic segments. The other cluster sizes produced segment clusters with overlap, though overlap is less than with phrases.

### 3.3.3. Experiment 1b: Discussion

The segments that resulted from the segmentation algorithm produced more discrete clusters than the complex warble phrases. For every cluster size, mean silhouette width was higher for segments which suggests higher degrees of stereotypy for segments. Phrases

clusters, on the other hand, were largely overlapping. This is expected if segments are more basic units that are combined to create complex phrases.

To the best of my knowledge, these are the first data to suggest that non-humans have productive capabilities at the level of the segment. Whales, songbirds, and parrots are able to create novel songs by reorganizing preexisting phrases (Marler, 2000), an ability once considered uniquely human (Hockett, 1960). These data suggest that non-humans can also combine units which are not bound by intervals of silence. Budgerigar vocal complexity makes use of similar generative combinatorial abilities found in human phonology. Furthermore, these data suggest that budgerigars are able to create novel structures at, at least, two levels: combining and rearranging phrases within warble (Tu, 2009; Tu & Dooling, 2012) and combining and rearranging segments within phrases.

### **3.4. Experiment 2: Population prediction**

An interesting aspect of the particulate principle is that it helps explain how a vast amount of diversity can emerge from the combination of a small set of discrete elements. If elements were blended together the result would be an average of the input and the full range of variation can be no greater than that of the initial set (Abler, 1989). Combining discrete elements, on the other hand, can create an almost infinite set of novel arrangements, even with a small set of basic units.

Because the number of combinations can be vast, substances or systems that are clearly different at a higher level of organization are less clearly so when analyzing their component parts. For example, hydrogen and oxygen could combine into water, H<sub>2</sub>O, but the same two elements could combine into hydrogen peroxide, H<sub>2</sub>O<sub>2</sub>, which is highly toxic if ingested. Similarly, attempting to guess a language based only on the presence of a few segments would be difficult – or impossible depending on the segments. The voiceless velar stop [k] is present

in 89.36% of all languages, the bilabial nasal [m] is in 94.24%, and the high front unrounded vowel [i] is in 87.14% (Maddieson, 1984; Maddieson & Precoda, 1989).

Low level units, like segments, in spoken human language are tightly linked with vocal anatomy and auditory capabilities (Browman & Goldstein, 1992; Ohala, 1983). Since these traits are largely uniform across all human populations, all human languages should share basic patterns (Ohala, 1983). I expect that the same should hold for budgerigar warble and other learned acoustic communication systems. The more basic a level is, the more likely its units are tied to more widespread factors shared by most members of the species. As such, higher level warble phrases should be more reliable at predicting individual or population identity. Since higher organizational levels are less stereotyped, the likelihood that two individuals will share a unit by random chance is relatively low. Lower level segments, on the other hand, should be more species-wide and, therefore, worse at population or individual prediction.

### **3.4.1. Experiment 2: Methods**

We used a random forest algorithm (Breiman, 2001; *R* package *randomForest*, Liaw & Wiener, 2002) to assess the possibility that the acoustic cues in segments and phrases could be used by budgerigars to determine group or individual identity. I then trained four random forest models; the models varied in the type of acoustic input data (phrases vs segments) and in the classification output (group vs individual identity). Because some groups and individuals had more samples than others, I took a random subset of the data for each model. For the model classifying individuals using segment data, I used 500 segments from each individual. For the model classifying groups using segment data, I used 500 segments from each group. Because I had a much smaller number of phrases in the dataset, I took a sample size based on the group or individual with the smallest number of samples. For the model classifying individuals using phrase data, I took a random sample of 61 phrases from each individual. For the final model, group classification from phrase data, I took a random sample of 384 phrases from each group.

To assess whether the algorithms performed above chance at classification, I used an exact binomial test, *binom.test()* in *R*, for each of the models. Finally, I ran a paired t-test, *t.test(paired = T)* in *R*, to evaluate whether segment or phrase input performed better at classification.

For both sets of tests, I corrected for multiple testing by using *p.adjust(method = "Holm")* in *R*.

### 3.4.2. Experiment 2: Results

Overall, all of the models performed significantly better than chance at group/individual classification ( $p < 0.001$  for all). Using segment data as input, success rates were an average of 13% lower than with phrase data at classifying individuals ( $t = 4.91$ ,  $df = 13$ ,  $p < 0.001$ ) and 18% lower classifying groups ( $t = 5.17$ ,  $df = 3$ ,  $p = 0.014$ ).

**Table 3-2. Budgerigar population classification based on unit type.**

Unit/ Population	Success Rate	Chance Level	Binomial Test	Phrase vs Segment
Phrase – Group	71%	25%	>0.001	$t = 5.1$ , $df = 3$ , $p = 0.015$
Segment – Group	48%	25%	>0.001	
Phrase – Individual	40%	7%	>0.001	$t = 5.9$ , $df = 13$ , $p < 0.001$
Segment – Individual	24%	7%	>0.001	

Data for individual predictions were non-normal so I performed a paired Mann-Whitney U signed rank test using *wilcox.test(paired = T)*:  $V = 105$ ,  $p > 0.001$ . I then ran a t-test to check if the results were different. Both tests found significance, so I presented the results of the t-test to be more parallel to the group data.

### 3.4.3. Experiment 2: Discussion

Budgerigar phrases are more reliable than segments at predicting group and individual identity. As with the previous experiment, these data suggest that the warble segmentation

algorithm is finding units within inter-silence intervals. Segments are more stereotyped and consistent across populations and budgerigars are able to create group- or individual-specific phrases via segment combination and organization. These data are consistent with what is known about segments in human language. The human sound inventory can be, generally, described with a single page of symbols (the International Phonetic Alphabet; Ball et al., 2018) and certain sounds (such as [i], [m], [k]) are shared among the vast majority of languages (Maddieson & Precoda, 1989). A single language, however, needs hundreds of pages to approximate the number of words or phrases in its inventory. Furthermore, similarities in segment inventory are not evidence for historical relatedness or cultural contact. Aleut, Arrernte, and Tsimshianic are unrelated languages and yet all have a three vowel systems comprised of the same vowels: /a/, /i/, and /u/ (Hyman, 2008; Maddieson, 1984; Maddieson & Precoda, 1989).

### **3.5. Experiment 3: Edge effects**

Because segments are produced rapidly and in tight coordination over a short time period, their production is affected by acoustic context. Some segments may obscure the acoustic cues of other segments, some articulatory transitions are more complex than others, and aerodynamic pressures change over the course of a vocalization. Similar positional biases across multiple populations can reveal that the pattern is the result of a species-wide biomechanical, cognitive, and/or functional pressure. Similarly, shared patterns across species can reveal more general principles on acoustic communication. Humans are the only species for which segmental data exists and there is clear evidence that structural context affects the production of human segments, most obviously at phrase boundaries (Fougeron & Keating, 1997). Human segments at the end of a phrase tend to be longer in duration and lower in fundamental frequency and intensity (Fuchs, Petrone, Rochet-Capellan, Reichel, & Koenig, 2015; Ladd, 1984). Initial segments are less periodic than medial and final segments (Fougeron

& Keating, 1997). In this section, I evaluate these four acoustic parameters in four independent budgerigar populations to assess whether budgerigars share these positional biases.

### 3.5.1. Experiment 3: Methods

To evaluate the effect of segment position on the acoustic signal, I divided the segments into three categories based on their relative position in the phrase: initial, medial, and final. I defined initial segments as the first segment of a phrase, final as the last, and the medial group included everything in between. I evaluated four acoustic measurements: mean fundamental frequency, duration, and intensity, and periodicity.

We used mixed effect models to assess whether segment position has an effect on the acoustic output. Mixed effect models, particularly generalized linear mixed models, allow for more flexibility and accuracy when analyzing non-normal data like ours which vary in samples per position, samples per individual, and number of individuals per group (Bolker et al., 2009). I used the *lme4* (Bates, Machler, Bolker, & Walker, 2015) package in *R* with segment position and population as fixed effects. I included population as a covariate to better assess if the segment patterns occur independent of group. I included individual identity as a random effect. I used the *anova()* function in *R*'s base *stats* package to compare the model with a null model that excludes the fixed effect of segment position. I also compared the full model with a model in which the covariate group was removed. I used the *vif()* function in *car* (Fox et al., 2011) to check for collinearity.

Visual assessment of the residuals for periodicity, mean fundamental frequency, and duration were all non-normal, so I used *lme4*'s *glmer()* function to fit a generalized linear mixed model (F0: Gaussian distribution with “log” link; Duration: an inverse Gaussian distribution with “identity” link; Periodicity: binomial distribution with “logit” link.) For intensity, the residuals were normal and homoscedastic, so I fitted a linear mixed model.

### 3.5.2. Experiment 3: Results

*Table 3-3. Group means and standard deviations by segment position.*

Group	Segment position	Intensity dB	Duration MSec	Periodicity %	F0 Hz
A N = 7	Initial	46.3 (±2.72)	4.99 (±0.75)	28.22 (±4.25)	2633 (±181)
	Medial	56.75 (±3.19)	6.13 (±0.58)	71.19 (±3.58)	2611 (±152)
	Final	52.51 (±3.37)	10.77 (±2.7)	49.79 (±6.89)	2081 (±110)
B N = 1	Initial	54.62	6.01	34.83	2442
	Medial	62.35	7.86	66.66	2479
	Final	58.05	10.53	44.77	1828
C N = 2	Initial	48.72 (±1.89)	7.16 (±1.03)	39.57 (±5.74)	1705 (±208)
	Medial	59.27 (±1.59)	6.43 (±0.73)	79.24 (±1.57)	2231 (±105)
	Final	55.47 (±3.24)	6.82 (±0.97)	50.44 (±0.36)	2118 (±206)
D N = 4	Initial	44.57 (±3.7)	5.82 (±1.93)	42.07 (±15.76)	2578 (±50)
	Medial	52.94 (±3.95)	6.82 (±1.07)	78.11 (±1.14)	2524 (±132)
	Final	50.51 (±6.23)	16.97 (±5.63)	68.49 (±4.88)	2085 (±240)

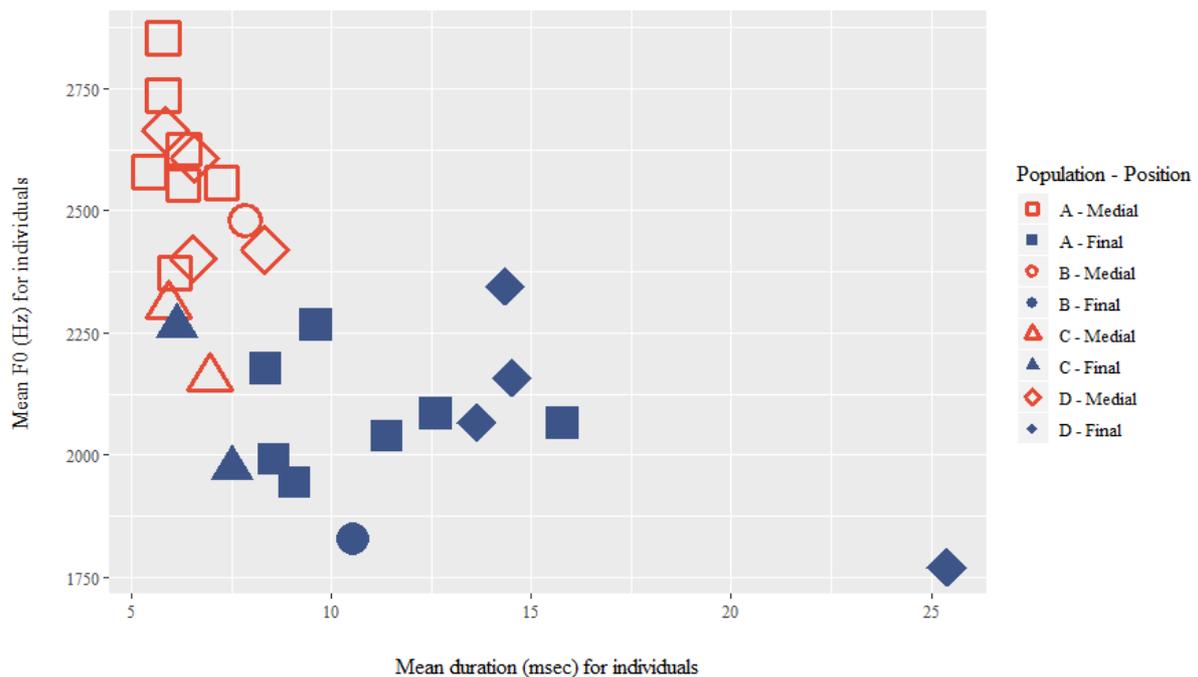
#### 3.5.2.1. Experiment 3: F0

The full model, with group and segment position as fixed effects, performed better than the null model, which did not include position as a fixed effect:  $X^2 = 2353.4$ ,  $df = 2$ ,  $p < 0.001$ .

The model without group performed worse than the full model:  $X^2 = 11.752$ ,  $df = 3$ ,  $p = 0.008$ .

For all four groups and all 14 individuals, mean F0 measurements were lower for segments in phrase-final position when compared to medial segments (mean of individual means:  $n = 14$ , Medial:  $\bar{x} = 2522$  Hz,  $\sigma = \pm 181$  Hz ~ Final:  $2069 \pm 165$ ).

Duration and F0 for Medial and Final Segments



**Figure 3-9. Duration and F0 of budgerigar medial and final segments.** Medial (unfilled red points) and final (filled blue points) segments. Data points are mean values for each individual budgerigar. Shapes represent group.

### 3.5.2.2. Experiment 3: Duration

For duration, the full model outperformed the null model:  $X^2 = 3209.6$ ,  $df = 2$ ,  $p < 0.001$ .

There was no difference between the full model and the model with group removed:  $X^2 = 5.672$ ,  $df = 3$ ,  $p = 0.129$ .

Final segments for all individuals are longer in duration, on average (mean of individual means:  $n = 14$ , Medial:  $\bar{x} = 6.5$  msec,  $\sigma = 0.84$  msec ~ Final:  $12, \pm 4.84$ ).

### 3.5.2.3. Experiment 3: Intensity

The full model outperformed the null model for intensity:  $X^2 = 22186$ ,  $df = 2$ ,  $p < 0.001$ .

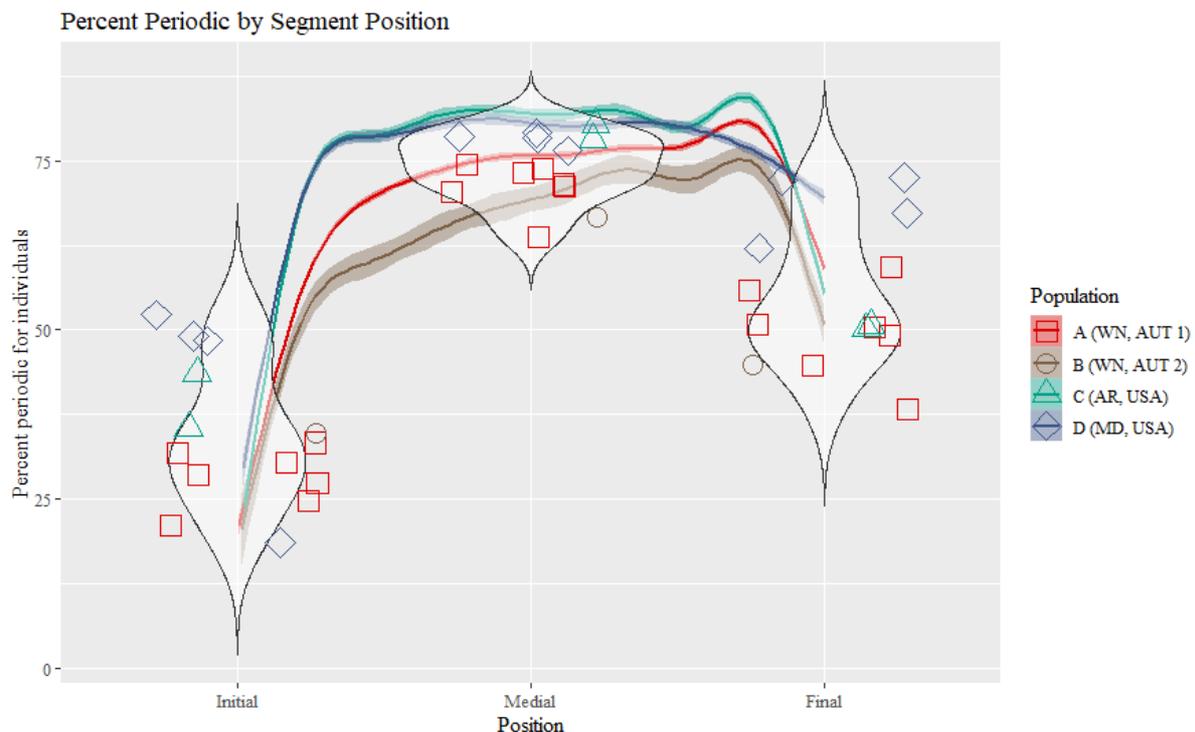
The full model also outperformed the model with only position:  $X^2 = 15.491$ ,  $df = 3$ ,  $p = 0.001$ .

For all individuals, intensity is lowest in initial position (Initial:  $\bar{x} = 46.7$  dB  $\sigma = 3.71$  dB ~ Medial:  $56.4 \pm 4.03$  ~ Final:  $52.8 \pm 4.46$ ). For all but one individual, a male from group D, intensity was lower in final position than in medial.

### 3.5.2.4. Experiment 3: Periodicity

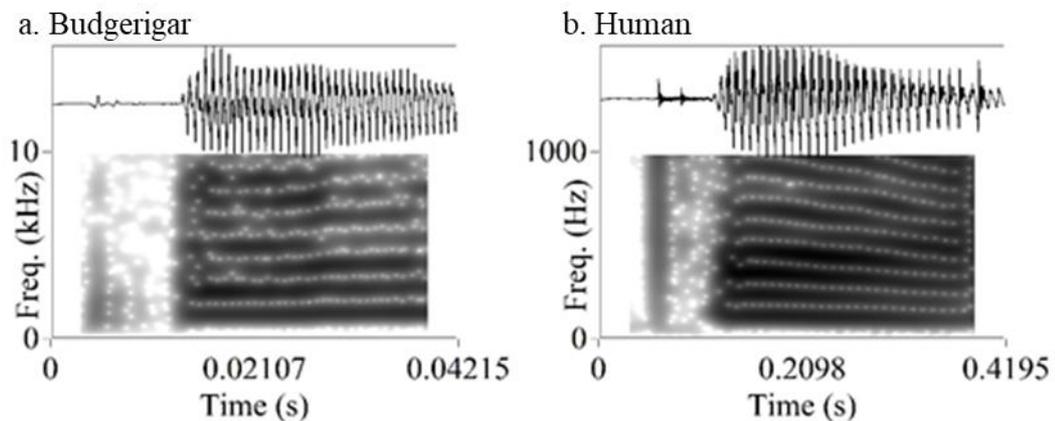
For the periodicity of segments, the full model performed better than the null model:  $X^2 = 10158$ ,  $df = 2$ ,  $p < 0.001$ . The model without group was worse than the full model:  $X^2 = 20.373$ ,  $df = 3$ ,  $p < 0.001$ .

All individuals shared the same basic “arch” pattern with respect to periodicity. Periodicity was lowest in initial position with slightly more than a third of segments having periodic vibration ( $n = 14$ ;  $\bar{x} = 34.3\%$ ,  $\sigma = 10.5\%$ ). Periodicity rose to its highest point in medial position ( $\bar{x} = 74\% \pm 4.9$ ) and then fell again in phrase-final segments ( $\bar{x} = 54.9\%$ ,  $\sigma = 10.5\%$ ).



**Figure 3-10. Periodicity by in budgerigar segments by phrase position.** The plot shows both a categorical and continuous representation of the position of segments within phrases. The violin plots and individual points show the data binned into initial (first segment), final (end segment), and medial (everything in between). Data points are mean periodicity at each position in the phrase. The lines are the data with position treated as continuous. Instead of binning the position data, I took segment’s ordinal position and divided it by the total number of segments in the phrase. I then applied a generalized additive model smoothing function to data: `geom_smooth()` in `ggplot2` (Wickham, 2016). For the continuous data, “Initial” is roughly 0 and “Final” is 1.

A surprising discovery was the strong tendency for the initial segments to be aperiodic, short, and quiet. As shown in Figure 3-11, many of the signals looked similar to human phrase-initial patterns.



**Figure 3-11. Budgerigar and human phrase initial patterns.** Both budgerigars (a) and humans (b) frequently begin phrases with an initial burst which is followed by a periodic signal. The two signals differ in scale – by an order of magnitude. The human example is the initial stop and vowel of an English speaker uttering the word “girl”.

### 3.5.3. Experiment 3: Discussion

As expected, there is a clear relationship between the relative position of segments within phrases and certain acoustic variables. For all four acoustic parameters, the best performing models included position. The null model, which included group but not position, was the worst performing model in all cases. Group had an effect for every variable except for duration.

On average, final segments are longer, quieter, lower in fundamental frequency, and more aperiodic than medial segments. Initial segments tend to be more aperiodic, shorter, and lower in intensity than either final or medial segments. I believe this is the first evidence for positional dependent differences at the level of the segment for any non-human vocal system. These patterns, however, are shared with humans. Segments in human language are generally longer when they precede a pause (Edwards, Beckman, & Fletcher, 1991; Fougeron & Keating, 1997; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Fundamental frequency and

intensity also tend to fall throughout a phrase resulting in a lower average F0 and amplitude in final position (Ladd, 1984; Vaissière, 1983, 1995). Humans also share a tendency for more aperiodicity in initial position. Sounds that generally have a clear F0 are often produced either without voicing or with irregular vocal-fold vibration when produced after a pause (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Garellek, 2014; Pierrehumbert, 1995; Westbury & Keating, 1986). Furthermore, a transient burst (“voiceless plosive”) preceding a loud, periodic signal (“vowel”) is the most common pattern across human languages and is almost universal among spoken languages (Hyman, 2008; Ladefoged & Maddieson, 1996; Prince & Smolensky, 2002). This aperiodic-periodic pattern is similar to that of budgerigars where the initial segment is also likely to be a transient burst.

While positional dependent differences at the level of the segment have not been described in any other species, there is evidence that some of these patterns are somewhat common at higher levels of analysis. For human language, segments are not the only units which are longer in final position as Wightman et al. (1992) found that final lengthening occurs in syllables, words, and phrases. The evidence for intensity and F0 decrease is most robust at the level of the sentence (Ladd, 1984; Vaissière, 1983, 1995). Outside of humans, Tierney, Russo, and Patel (2011) found that songbird notes (interpausal units which would be defined as “phrase” under my terminology) were on average longer at the end of a song bout than song medial notes. They also found that F0 tends to decrease at the end of a song. Hoeschele et al. (2010) found that the second note (also interpausal unit) in the two-note black-capped chickadee call was lower in intensity, especially for subordinate males. The aperiodic-periodic pattern, on the other hand, does not seem to have been described as a widespread pattern in any non-human species.

The fact that many species share final lengthening, F0 declination, and intensity declination provides insight into why budgerigars have these patterns. Final lengthening may

be the result of gradually slowing articulators in preparation for the end of a vocalization (J. Edwards et al., 1991; Lindblom, 1968). Humans and songbirds rapidly adjust their articulators during vocal production and abrupt termination of these movements may be more difficult than a gradual relaxation of articulators (Lindblom, 1968; Tierney et al., 2011). Budgerigar vocalizations, and complex phrases in general, are characterized by rapid acoustic modulations which are likely the result of rapid articulator movement, either in the vocal tract or of the syringeal muscles (Heaton, Farabaugh, & Brauth, 1995).

The mechanisms underlying lower  $F_0$  and amplitude in final position seem to be directly related (Vaissière, 1983). Air volume decreases throughout the production of vocal units, particularly longer and more complex units like utterances, songs, or phrases. A decrease in air volume will lead to a decrease in amplitude, all things being equal (Lieberman, 1958). This decrease in air volume will also affect subglottal pressure, which is one of the determining factors of the rate of vocal fold vibration (Titze, 1989; van den Berg, 1957). All available evidence on budgerigar vocal production, or that of closely related parrot species, suggests that the same mechanisms in humans and songbirds could explain why budgerigar segments are lower in amplitude and fundamental frequency (Brittan-Powell, Dooling, Larsen, & Heaton, 1997; Larsen & Goller, 2002). Budgerigars produce vocalizations by using a pair of membranes to constrict the expiratory airflow. The constriction creates a pressure build up which forces the membranes apart. The membranes recoil and pressure builds up once again resulting in a self-oscillating system, the acoustic result of which is a quasi-periodic signal. Like songbirds and humans, the rate of tissue vibration is partially affected by the amount of air volume, so the depletion of air over the duration of a phrase would cause both intensity and  $F_0$  to fall.

The burst-periodic pattern has not been described at any level in non-humans until now. In humans the underlying mechanism for this pattern is not largely agreed upon. The similarity

between budgerigars and humans suggests that species-specific explanations (e.g., Universal Grammar; Prince & Smolensky, 2004) are not warranted and that we should consider (one or more) widespread cognitive, functional, perceptual, or bio-mechanical mechanisms. For instance, one possibility is the maximization of contrastive units (Liljencrants & Lindblom, 1972; Lindblom & Maddieson, 1988). When taking the capabilities of the human vocal tract and ear into consideration, aperiodic bursts and periodic signals (with harmonic stacks) are the sound classes that are the most distinct from each other. By using sounds from these two classes, humans can maximize the amount of auditory space and minimize confusion for the listener. Furthermore, a burst followed by a periodic signal seems to be better than the reverse pattern at preserving the acoustic cues of both segments (Ohala & Kawasaki-Fukumori, 1997; Wright, 2004; though not for all stops, see Steriade, 1999a, 2001). For budgerigars, I do not know if this explanation is correct, but given the work I have conducted here, it could be tested directly by investigating budgerigar perception of their own segments. As in humans, bursts and periodic signals likely occupy opposite ends of the budgerigar articulatory-acoustic space. Presumably, whatever acoustic cues are relevant to the budgerigar system should be preserved with a burst-periodic pattern as well.

Why would budgerigars need to maximize the acoustic space? The function of warble song is not completely understood, though it is clearly relevant in courtship (Brockway, 1964, 1965, 1968). If females prefer more diverse signals, the males may use a burst-periodic pattern to maximize the chance that their diverse repertoire is noticed. However, the evidence for repertoire size being the result of sexual selection is less robust than once believed (Soma & Garamszegi, 2011). Specifically in budgerigars, Tobin, Medina-García, Kohn, and Wright (2017) found that female-directed song is actually more stereotyped than male-directed song with respect to phrase “syntax”. The social environment that budgerigars typically sing in also does not seem conducive to the preservation of the aperiodic bursts. Multiple individuals sing

at the same time in close proximity to each other, in the wild thousands of budgerigars could be in one area at the same time (Wyndham, 1980). Short, quiet segments could easily be lost in the surrounding noise. That being said, if the short, quiet bursts are likely to be lost in surrounding noise, the most optimal position would likely be phrase-initial. If the organization of human segments provides any support, the cues in plosives are best preserved before vowel (Wright, 2004).

Another (non-mutually exclusive) possibility is that budgerigars and humans share sound production mechanisms which lead to phrase initial bursts. In numerous languages, reduced periodicity and/or periodic bursts occur, even when speakers and listeners don't perceive the sound (*glottal stop insertion*; Blevins, 2008; Dilley et al., 1996; Fougeron & Keating, 1997; Garellek, 2014; Kingston & Diehl, 1994; Pierrehumbert & Talkin, 1992). This is particularly true at the beginning of a phrase and with emphasis (Dilley et al., 1996; Fougeron & Keating, 1997). For instance, English speakers will often produce a glottal stop before the vowel in a word like *apple* or *issue*, though most listeners don't recognize that it is present (Dilley et al., 1996). The budgerigar and human patterns could be the result of the difficulty in initiating voicing directly from non-phonation. With a closed vocal tract the pressure variation above and below the vibrating tissues is not sufficiently different for phonation; once the closure is released and the appropriate pressure differential can be reached then voicing can begin (Westbury & Keating, 1986). In humans, unless other measures are taken, vocalizations will naturally begin with a brief interval of aperiodicity (Kingston & Diehl, 1994; Westbury & Keating, 1986).

It is premature to assume that budgerigars and humans are the only species who would have a burst-periodic pattern. While this pattern hasn't been described as widespread or systematic for any other species, published descriptions or spectrograms suggest that at least some populations of species may have somewhat similar tendencies (Bartsch, Hultsch, Scharff,

& Kipper, 2016; Jansen et al., 2013). Without a segmental approach, like the one presented here, this pattern is likely to go unnoticed.

### **3.6. General discussion**

As far as I am aware, this is the first detailed acoustic analysis of inter-silence interval units in a complex system, outside of humans. Du & Troyer (2006) created an algorithm to segment zebra finch “notes” – equivalent to the level of the segment presented here – but they, nor anyone else, have analyzed zebra finch segments. Other analyses at the level of the segment have focused on simple call concatenations (Collier, Townsend, & Manser, 2017). Because the level of segment has been such an important area of research in human language, I believe that these findings will help to expand our understanding of complex acoustic communication and the similarities across vocal learning species. In budgerigars, I found that these units share important parallels with human segments: warble segments are tightly coordinated in articulatory/acoustic space (Experiment 1, 3.3), show less variation across individuals/populations than higher analyzable levels (Experiment 2, 3.4), and are affected by structural context (Experiment 3, 3.5).

Through a segmental approach, we can push comparative work even further and explore deeper questions related to the evolution of vocal learning and of human language. For example, evidence from some less vocally flexible songbirds suggest that large song or phrase repertoires can be costly, particularly during development (MacDougall-Shackleton & Spencer, 2012). Combining units may be a more efficient path to large repertoires. These data suggest that, all things being equal, we should expect to find segments in those species which have hundreds of unique phrases.

Previous research has found that warble can be organized at the level of the phrase (Tobin et al., 2017; Tu, 2009; Tu & Dooling, 2012). These data suggest that the organization of segments is also productive. Deeper analyses can reveal if intermediate levels of organization,

like syllables or metrical feet, exist in budgerigars or other non-human acoustic communication. Subsequent research can help us understand how segments are perceptually grouped. I found acoustic clusters in warble segments which will allow us to now address whether budgerigars have human-like “natural classes” and if they have population-specific rules that operate over those classes (Yip, 2006). This research revealed contextual biases in the production of segments, but I do not know the exact nature of the biases. In human language, phrase final lengthening is a within-type phenomenon, that is, the final unit may not be the longest in absolute terms, but rather it is longer in final position than it would be in a medial position (Fougeron & Keating, 1997). For example, the final [s] in *misplace* [mɪspleɪs] is longer than the one that precedes the [p], however, it is likely shorter than all of the vowels in the phrase.

Budgerigar warble seemingly lacks the referential aspects of human language, as such, it is an enticing model for a “bare phonology” and may even be able inform on more difficult questions related to human language evolution (Fitch, 2010). Studdert-Kennedy (1998) argues that the combinatorial properties of human phonology may have been an important step in the evolution of language. Humans can apply a label to practically any object, event, idea, or proposition. Without the combining and rearranging signals, the number of possible labels would be quickly exhausted. For budgerigars, however, no evidence exists to suggest that warble or its components are referential or are used for labeling objects. The pressure to create novel labels, it seems, is not a requirement for the development of particulate segment system.

Since budgerigar warble is sung mostly by males and, in part, during courtship (Brockway, 1964), the similar segmental systems of humans and budgerigars are, perhaps, more consistent with an alternative hypothesis of human language evolution: the “prosodic protolanguage” hypothesis. This hypothesis proposes that language first developed as a courtship display similar to the song of many whales and temperate zone songbirds (Darwin, 1871; Fitch, 2010;

Nowicki & Searcy, 2014). While modern human females and males are largely equal in linguistic abilities, the anatomical differences in the larynx that appears at puberty suggests sexual selection played a role in human vocal evolution (Fitch, 2000). Budgerigar segments provide us with the opportunity to understand how and if the vast productive abilities of human phonology could have arisen without needing to rely on propositional or referential meaning.

## **Chapter 4. Newfound variation: House finches learn canary trills**

### **4.1. Introduction**

Since behaviors rarely fossilize, comparing species with differing evolutionary histories is essential in understanding the functions, origins, and evolution of complex behaviors. Comparative research in vocal learning, for instance, has revealed that traits like bird song and human language share deeper similarities than one may expect. Vocal learning, or the ability to learn and mimic acoustic signals, is a central component of human language. It is a rare trait but it is not uniquely human as parrots, songbirds, hummingbirds, whales and dolphins, bats, seals, and elephants learn their vocalizations (Janik & Slater, 1997; Jarvis, 2006). All of these species, humans included, are constrained to some degree in their learning abilities, but the amount of vocal flexibility varies widely between them.

While much fruitful work has been done with the vocal learning mammals (Janik & Slater, 1997; Knörnschild, 2014; Reichmuth & Casey, 2014; Stoeger et al., 2012), bird acoustic communication has so far provided the most insights into vocal learning and has served as the best comparison for numerous components of human language (Bolhuis & Everaert, 2013; Doupe & Kuhl, 1999). Birds have a diversity of evolutionary histories, ecological niches, and learning abilities and they can be researched with relative ease both in the lab and in the wild. This diversity allows us to make more nuanced distinctions within the trait of vocal learning and to potentially discover the role that genes, habitat, breeding type, brain size, vocalization function, morphological attributes, etc. all have on vocal learning abilities. This diversity also allows us to find better models for human language and speech. Language is a complex, multi-component trait that no other species has *in toto*, but which many species will share in part (Fitch, 2010). While some avian species may share only the ability to learn sounds when they are young, other species may share a plethora of linguistic abilities with humans.

One species that has been understudied, despite its potential to serve as vocal learning model, is the house finch (*Haemorrhous mexicanus*). The house finch is a small, social oscine bird that, since the mid-20th century, has rapidly expanded its range from Mexico and the Southwest of the United States to most of the non-Arctic areas of North America. The species is non-territorial and typically occupies high areas, such as the tops of trees. Coloration is sexually dimorphic with males having bright red plumage while the females are brown. Both male and female members take part in care of nestlings (Badyaev & Hill, 2002). Populations nest semi-colonially, with many nests being separated by only a few meters (W. Thompson, 1960).

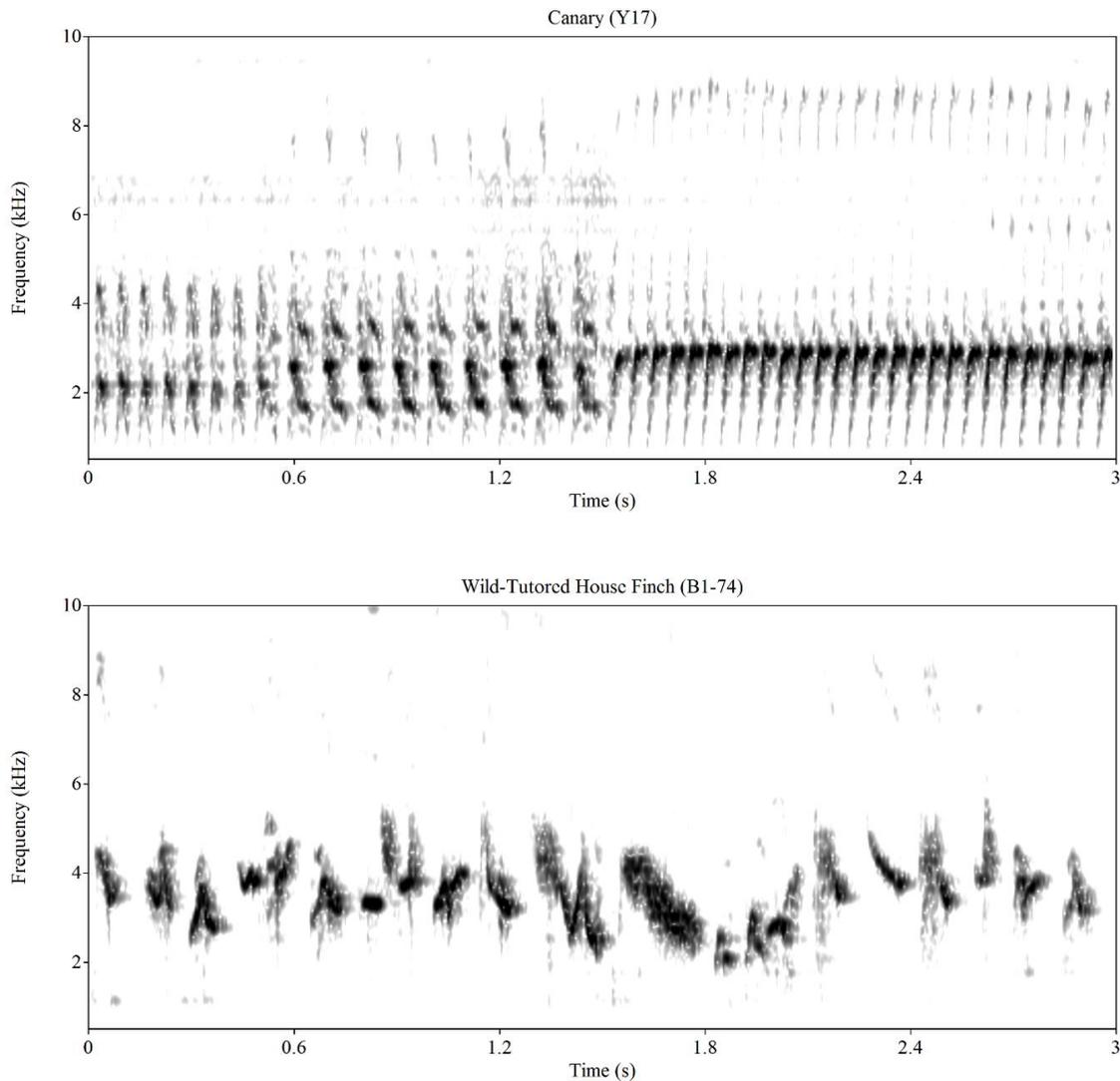
House finch song is moderately complex with each bird having a repertoire of roughly 3-5 song types and 20-30 unique phrases. Phrases are mostly frequency modulated pure tones or “buzzes” which are broadband noisy signals. Song typically lasts for about 2-3 seconds and is made up of 10-15 unique phrases, with little phrase repetition. Song varies between geographic regions with birds on the East Coast having more stereotyped song, more phrase sharing, and lower phrase repertoire size (Bitterbaum & Baptista, 1979; Mundinger, 1975; Pytte, 1997). Both males and females sing; while males sing more often, female song is more widespread than in most temperate, non-territorial songbirds (Bitterbaum & Baptista, 1979; Mundinger, 1975).

Research suggests that house finches, both male and female, learn their song. Song varies between populations, they can modify their song to counteract temporary environmental disruptions (Garcia, Bermúdez-Cuamatzin, Ríos-Chelén, & Gil, 2009), and they have been reported to mimic heterospecific song in the wild (Baptista, 1972; Payne, Payne, & Whitesell, 1998). However, no published work has yet confirmed this, nor tested the biases and extent of learning in house finches. In this study, I will close this gap by presenting data from house

finches who were tutored with canary song, house finches who were acoustically isolated, and house finches who had exposure to species-typical song.

Cross-fostering has been utilized frequently to test for vocal learning and the extent of learning permissivity (Owren et al., 1992; Zann, 1985). Cross-fostering removes species-typical input and replaces it with input from another species, “heterospecific”. The protocol has the benefit of providing an individual with a social environment, even if the social environment is atypical (Immelmann, 1975; Rowley & Chapman, 1986). Under this paradigm, a test subject which produces species-typical vocalizations provides evidence that vocal learning is not required for the development of that vocalization in that species, as has been found with non-human primates (Owren et al., 1992). The lack of heterospecific copying is not necessarily evidence for vocal learning, as mimicry could be affected by other factors, but those whose learning is more permissive are expected to more accurately reproduce the input of their fosters (Shizuka, 2014).

For house finches, canaries provide several benefits as a foster and tutor. Canaries do well in captivity, are very vocal, will foster heterospecific nestlings, and have songs that vary considerably from that of typical house finch song (see Figure 4-1).



**Figure 4-1. Canary and house finch song spectrograms.** Three seconds of waterslager canary (top) and house finch (bottom) song taken from individuals in the current study. Song defined as a sequence of vocalizations surrounded by 400 milliseconds of silence. While house finch song is typically 2-3 seconds, canary song can continue for much longer periods and are less stereotyped. The fundamental frequency of (waterslager) canaries is typically in the range of 1-4 kHz and is comprised of trills, rapid repetitions of the same unit (Dooling, Mulligan, & Miller, 1971; Güttinger, 1985). House finch fundamental frequency typically ranges from 2 to 6 kHz and has units as high as 8 kHz (Bitterbaum & Baptista, 1979; Mundinger, 1975). Trills are not common in house finch song.

As seen in Figure 4-1, Canary song is composed, primarily, of rapid repetitions of short units. House finch song has little within-song unit repetition. The stark differences between the natural song of the two species, in phrase structure and organization, provides us with an opportunity to see how permissive the house finch song learning program is. Because heterospecific mimicry has been reported in house finches, even mimicry of canaries (Payne

et al., 1998), I expect that they will be able to learn aspects of canary song. However, the similarities in house finch song across populations suggest that their song is constrained. Constraints will limit their ability to faithfully mimic. The errors in mimicry should provide insight into the nature of the limits of house finch learning.

## **4.2. Methods**

For this analysis, I analyzed acoustic recordings from twenty-two individuals from four groups. Recordings from eighteen of these subjects were collected from two unpublished experiments performed in the early 1970s by Paul C. Mundinger & Laura Waddick (PCM & LW, hereafter). The first experiment was a cross-fostering experiment wherein canaries fostered house finches. I analyzed three groups from this experiment: canary fosters, canary-tutored house finches, and quiet-reared house finches. The two house finch groups were fostered by canaries, but the latter group did not hear canary song due to acoustic masking (details below). The second experiment was a house finch song ontogeny experiment. In this study, PCM & LW recorded house finches during their development in order to track how song crystalizes. Recordings for the final four individuals were from PCM field recordings of wild house finches.

Before he passed away, PCM bequeathed the Lahti Lab all of his unpublished data so that the lab could continue his decades of research with house finch (and other songbird) song. LW gave permission for the use of these recordings, as well (Waddick, personal communication with David Lahti).

### **4.2.1. Cross-fostering subjects**

From PCM & LW's cross-fostering experiment I was able to locate quality recordings from nine house finches and six canaries. Recordings from more individuals were available, but I did not use these recordings for various reasons. While most recordings had clear information, some lacked details about the individual singing or what conditions the individual

was recorded under. Occasionally, different identification labels were used for the same individual. If information was unclear, lacking, or contradictory, I did not use the recording. I also excluded some recordings due to poor quality. These had poor signal-to-noise ratio, intrusive background noise, or clipping. One of the sound analysis programs that I used in this analysis, FinchCatcher (Ju, 2015; further details in 5.3.2), allows users to adjust parameters to overcome minor issues with sound quality (e.g., pass-band filters, variable amplitude thresholds), but some of the recordings were of such poor quality that reliable acoustic measurements could not be extracted using the program. Finally, I excluded recordings that did not address whether house finches would produce species-typical song in isolation or whether they could learn canary song. In their experimental design, PCM & LW included house finches who were deafened so that they wouldn't hear canary song or, later, their own song. They also included house finches who were rewarded with food for singing and house finches that were exposed to canary song only in the first month of life – to see when song learning would start. Birdsong research since the early 1970s has found that listening and producing song is its own reward (Adret, 1993), songbirds typically need to hear their own voice for proper sensorimotor development or for song to remain stable (Brainard & Doupe, 2000; Nordeen & Nordeen, 1992), and the sensory period of learning typically starts after the first few weeks post-hatching (Doupe & Kuhl, 1999; Marler, 1970a; Mooney, 1999). While these questions may not have been addressed in house finches, I wanted to focus on the question of “Do house finches learn?” and “What do house finches learn?”.

The house finches that I used from the cross-fostering experiment were from two experimental groups: canary-tutored and quiet-reared. Both groups were fostered by canaries, but the latter group was exposed to 100 dB white noise throughout fostering. The canary recordings were from foster canaries used in the cross-fostering experiment and related experiments.

The house finches were reared from eggs collected from the wild (Rye and Merrimack, NY) or from an aviary at Rockefeller University. To ensure that the birds received no acoustic exposure post-hatch, all house finches were hatched and reared in sound-proof chambers. The sex was not known before being put into the experimental groups. Five eggs were placed in the canary-tutored group. Of those five, three were male and two were female. Four eggs were put in the quiet-reared group, three of which were siblings (individuals B-5, D-3, & D-4). All quiet-reared individuals were female.

The canaries used for fostering were Belgian waterslager canaries. The use of live canaries allows for the fledgling house finches to have social interaction during song development, however, it comes with a cost: song input will vary between the foster canaries. To help mitigate this variability, the waterslagers were from an inbred strain purchased from a Belgian breeder. The breeder and strain were chosen because the population had little between-individual variability in song. I found quality recordings of three males and three females.

In searching for relevant files, I discovered that many of the recordings had gone missing since the initial experiment. I am missing some of the tutored foster parents, which would have allowed for a direct comparison of tutor and tutee song. All the canaries did come from the same in-bred stock and had similar initial acoustic exposure, which increases the likelihood of finding evidence of copying. The six canaries I analyzed (as well as another two I discarded because of lack of quality recordings) did share some song elements in common. The crystallized song of many of the cross-fostered house finches were also missing, though I do have printed spectrograms for some individuals. I compared the printed spectrograms to the files I had access to. I also compared subsong and crystallized song from the individuals where both were available. From these, song duration and the number of song variants differ from late subsong and crystallized song. As such, I did not analyze these acoustic parameters.

#### **4.2.1.1. Cross-fostering methods**

All nine of the house finches were fostered by canaries, even if they didn't receive acoustic input. All individuals were also exposed to a period of 100 dB white noise masking noise during rearing. The white noise served to mask auditory input while not causing long term damage (Marler, Konishi, Lutjen, & Waser, 1973). The canary-fostered group was exposed to white noise to try to control, somewhat, for any potential stress the white noise would cause. The stress associated with white noise, however, should be much less than that associated with cochlear removal and may have minimal stress potential (Crino, Johnson, Blickley, Patricelli, & Breuner, 2013). The white noise was produced by a Grayson-Stadler noise generator (Model 455C) and was amplified by a Lafayette amplifier (Model PA645A). Two to three 5'' speakers were attached to each chamber. Siblings were kept together until 1-2 months of age, before they began to produce subsong.

PCM & LW gave all of the house finches 10 mg pellets of testosterone propionate in late January, early March, and mid-April, to ensure high levels of testosterone throughout song development. They did this in order to compensate for the high number of females in the sample, as well as the possible decrease in motivation as a result of social isolation or deafening. The high number of females in the data set, and the fact that the quiet-reared group only has females, is not ideal. For many temperate zone songbird species, only the male produces song. However, female house finches do sing in the wild, typically, to solicit courtship or feeding (Mundinger, 1975). Female song was present in seven of the nine recording stations in Mundinger's 1975 analysis and they sang the same song types as the males. Bitterbaum & Baptista (1979) also found that female house finches injected with testosterone sang and produced song similar to that of males. Females had fewer song variants and phrase repertoires, but the acoustic structure of the phrase and phrase sequencing were not different.

The house finches were recorded in the isolation chambers on a voice actuated Tandberg tape recorder (Model 15-21) at 3 ¾ ips. Voice actuation was disabled when it was found to cut off the beginning of many songs. From the end of March, PCM & LW continuously recorded each individual for one to two hour sessions.

The quiet-reared group was comprised of four house finch females, which did not receive any acoustic tutoring during rearing, with white noise being played until they were isolated at 31-38 days. Once they were moved into individual quiet chambers, they were able to hear only their own vocalizations. Three siblings (D-3, D-4, and B-5) had noise removed before being isolated (noise ended at day 31 and they weren't moved until day 34). If the foster parents did sing, the birds were likely experiencing temporary deafness associated with chronic noise (Ryals et al., 1999).

The canary-tutored house finch group were exposed to white noise from hatching until 30-33 days old, at which point PCM & LW removed the noise and they were able to hear their canary fosters. PCM & LW implanted the male canaries with 10 mg of testosterone to ensure the house finches would have some acoustic input. Recordings showed that both foster parents sang during rearing, so the tutored birds heard song from both male and female canaries. Only one house finch individual (C-6) was not exposed to song of both sexes. After 60-68 days post-hatch, all the house finches were isolated into quiet acoustic chambers.

All birds had access to water and diet of millet, niger thistle seed, rape, lettuce, fruit (apple and orange slices), and, during the rearing season, egg food.

#### **4.2.2. House finch ontogeny subjects**

In July of 1973, PCM & LW captured three wild house finch juveniles, two females and one male, for a house finch song ontogeny study. They had an estimated hatching in mid-May or early June of 1973. At approximately 120 days of age PCM & LW began recording the house finch vocalizations. The birds were recorded from August until song crystallization in

late May of 1974. Unlike the previous year's experiment, the birds had acoustic contact with each other until at least March of 1974. And since they learned in the wild there was no acoustic masking.

These individuals had the same access to water and diet of millet, niger thistle seed, rape, lettuce, fruit (apple and orange slices), and egg food as the canaries and house finches in the cross-fostering experiment.

#### **4.2.3. Wild recorded subjects**

Because the house finch ontogeny experiment only used three individuals, I analyzed additional field recordings from PCM. PCM had recorded house finches at multiple sites across New York, New Jersey, and Connecticut in the 1970s. I only needed a few individuals to make sure that the house finch-tutored house finches were similar to wild individuals. I randomly selected a recording from Davids Island in June of 1972. The recording had song from four wild house finches, one female and three males. I used two distinct songs each from each individual. According to Mundinger (1975), house finches typically have two "themes", that is, songs that vary in at least a quarter of their constituent phrase types. All songs were recorded using a Narra III (3/4 ips) with Sennheiser 804 microphone. Further details can be found in Mundinger (1975).

Information on the individuals for whom I found recordings is shown in Table 4-1.

**Table 4-1. Individuals & number of quality recordings.**

Group	Individual	Sex	Sound files	Group	Individual	Sex	Sound files
Canary	O16	♀	10	House finch:	B1-74	♂	62
	OGW	♀	5	Ontogeny	B6-74	♀	9
	Peg	♀	18	experiment	C5-74	♀	8
	Red	♂	36	House finch: Field recordings	F1	♀	2
	Y17Y37	♂	33		M1	♂	2
	Y20	♂	14		M2	♂	2
House finch:	A5	♀	13	M3	♂	2	
Canary-tutored	B1	♂	100	House finch:	A6	♀	5
	C6	♂	39	Quiet-reared	B5	♀	28
	D2	♀	23	D3	♀	59	
	D5	♂	41	D4	♀	64	

#### 4.2.4. Song analysis

I defined “song” as a sequence of more than four phrases with less than 0.4 seconds between phrases, a broad definition meant to be as inclusive as possible when dealing with two different species, acoustically isolated (quiet-reared) birds, and cross-fostered birds. In Praat (Boersma & Weenink, 2016), I used the “Annotate: To TextGrid (silences)...” function to automatically annotate sections of the recordings find and label songs (minimum silent interval = 0.4 seconds, minimal sounding interval = 0.001 seconds, variable silence threshold settings depending on signal to noise ratio). In some cases, the automatic annotation failed to capture the whole song or captured background noise. I listened to each song and corrected the annotation by adjusting the Pratt textgrid interval boundaries. He used the same 0.4 second criteria. He hand-labeled each song by individual, date, and quality. Songs that had a poor signal to noise ratio or intrusive background noise were labeled as “poor” and excluded.

Songs from all groups were visually inspected to create a phrase inventory for every bird. To find phrase boundaries, I imported each song into FinchCatcher (Ju, 2015). I manually adjusted amplitude, filter, and interval settings for each song to so that phrases would match with the definition of “syllable” in Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra (2000), that is, a continuous sound with any amount of silence on both sides. Songs were cross-checked

by looking at the spectrograms in Praat. Praat allows for more flexibility in adjusting the number of samples per window so I could balance the spectral/temporal resolution trade-off.

With phrase boundaries defined, I classified phrases by frequency and spectral shape characteristics as seen on a spectrogram. For each individual bird, I labeled every phrase with a letter and cross-checked each new phrase with those that were previously labeled. Similar to Pytte (1997), I labeled phrases as the same type if I found continuous variation between them and different types if I found consistent discrete differences. I checked at least fifteen songs and at least a total minute of song for every bird, with the exception of the supplement house finches recorded in the wild. For these four individuals I used two themes for each. Most birds required less than twenty seconds before no new phrase type could be found.

Acoustic metrics were exported from FinchCatcher once the settings were appropriately adjusted. For each phrase, FinchCatcher calculates ten acoustic measures: mean fundamental frequency (F0), maximum and minimum F0, F0 range (the difference between the minimum and maximum F0), start and end F0, F0 slope (average frequency change), duration, concavity (the number of changes in the sign of the slope), and frequency-time excursion length (the sum of frequency modulations; Ju, 2015; Podos et al., 2016).

With these data, I used a Random Forest classification algorithm (Breiman, 2001) to test which canary-tutored house finch phrases would be classified with canary phrases. Using the *randomForest* package (Liaw & Wiener, 2002) in R (R Core Team, 2017), I used a random subset of 435 house finch and canary phrases to train the classification model (500 trees built with 3 parameters used at each split). I used 435 to make sure the numbers of phrases were equal between canaries and house finches, and I had 435 total phrases for house finches. I tested the model on all the phrases for each canary-tutored and quiet-reared house finch.

I also analyzed the following song-level parameters: trill length, trill rate, maximum trill rate, and sameness likelihood. Trills are defined as the same phrase repeated in sequence; trill

length is how many phrases are repeated before switching to another phrase type, while trill rate is number of phrases produced per second. Trill rate was measured for each unique phrase type in each bird's inventory that repeated three or more times. I took the median and maximum rate. I measured trill rate by finding the time from the energy peak of the first phrase to the energy peak of the last phrase and dividing by the number of inter-phrase silences. Sameness likelihood is the probability that a phrase will be followed by a phrase of the same type. I calculated sameness likelihood by giving a binary score to each phrase-phrase transition. If the phrase-phrase transition was composed of two phrases of the same type, I gave the transition a score of 1, e.g., AA = 1. I gave transitions with two distinct types a 0, e.g., AB = 0. I divided the sum by the total number of phrases transitions in the song to get the overall sameness likelihood score: e.g., AABB = { 1, 0, 1 } = 2/3 = 0.667.

Additionally, I present number of unique phrases and the percentage of phrases that were trilled (i.e., if a phrase type was ever repeated three or more times).

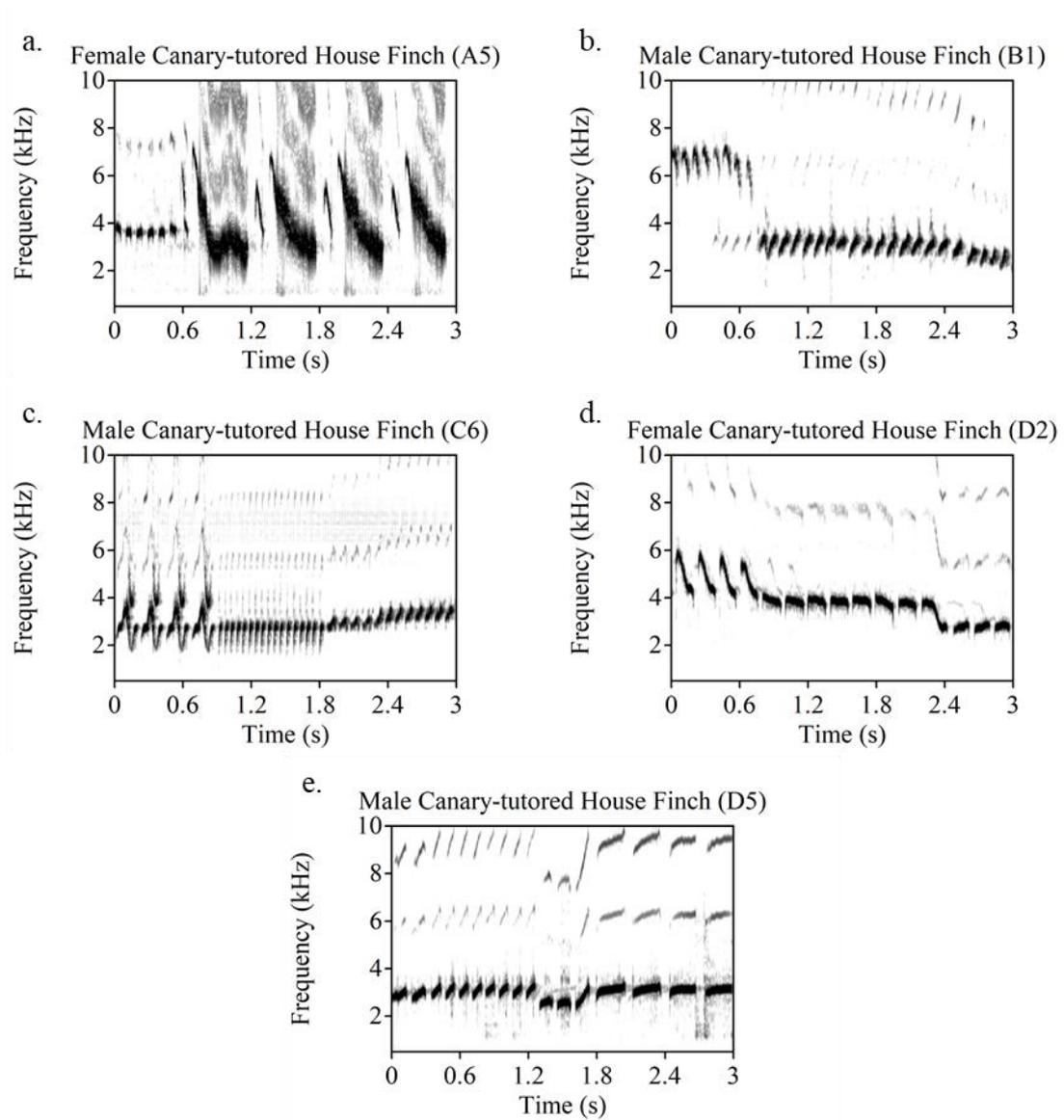
For descriptive and inferential statistics, I collapsed measurements for each individual down to an average value so that for each acoustic parameter each individual is represented by one data point. I used the *shapiro.test()* to check for normality. Because of the small sample sizes in each group and because much of the data were non-normal, I used non-parametric tests and median values. All data presented uses the median for the average and interquartile range for variance unless otherwise stated. Descriptive statistics and a visual inspection of the full dataset (i.e., multiple datapoints for individuals) suggested that variability within the quiet-reared and canary-tutored house finches was high enough that inferential statistics would not allow me to draw any firm conclusions with respect to house finch vocal learning. For inferential statistics, when comparing more than two groups I used the *kruskal.test* function in R *stats* package (R Core Team, 2017) to perform a Kruskal-Wallis rank sum test. For post-hoc comparisons, I used *pairwise.wilcox.test* with Benajmini & Hochberg (1995) false discovery

rate method (*p.adjust.method* = "BH") to correct for multiple comparisons. Where only two groups were tested, I used *wilcox.test*.

The mean was used for concavity and trill length because floor effects made the median unreliable. Concavity is the number of times the sign of the slope changes, that is, a change from a frequency fall to a frequency rise (or vice-versa). Because many phrases do not change frequency direction, concavity scores of zero are frequent. Trill length had a floor effect due to the most common repeated phrase length being two phrases. Median scores didn't distinguish between a house finch that had, exclusively, trills of two or three phrases and a house finch that had several trills of two or three phrases in addition to longer trills of seven or eight. So, I used the mean. The average trill length for individuals and average trill rate for groups were the only measures for which the mean-median difference mattered for the analysis. For trill rate I stayed consistent with the use of median because of the skewed distribution of the data. For group comparisons, I stayed with median for all parameters.

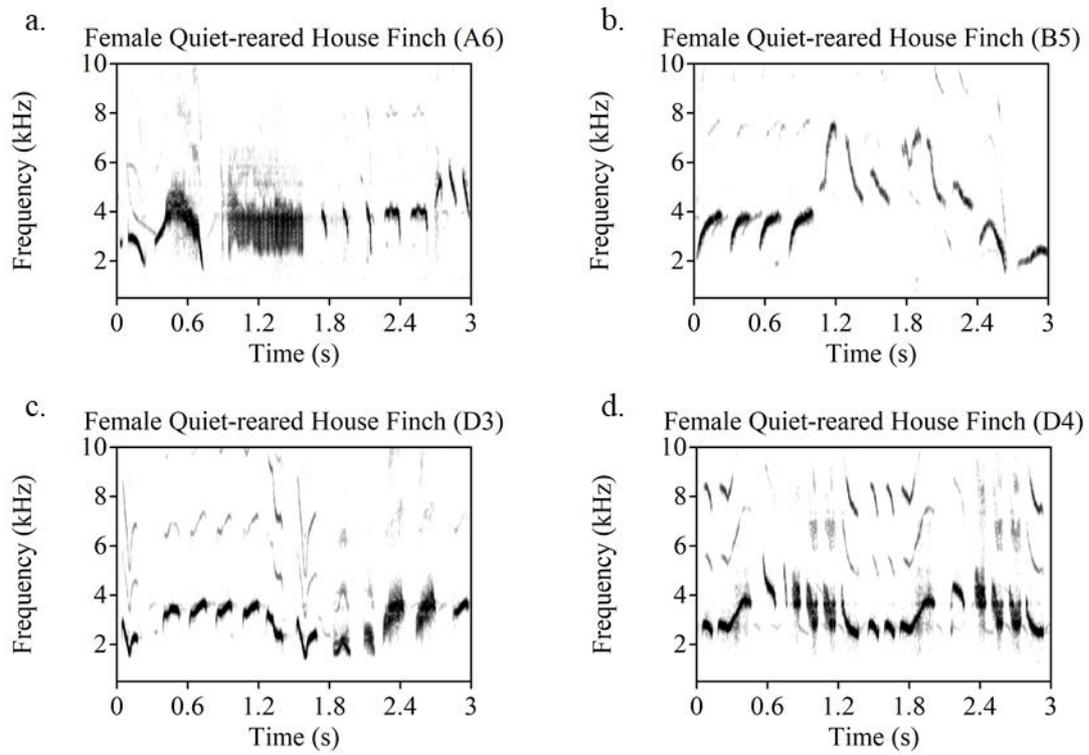
### 4.3. Results

Representative songs from the five canary-tutored house finches are shown in Figure 4-2.



**Figure 4-2. Spectrograms of canary-tutored house finches.** Three seconds of song from each canary-tutored house finch. In every case, the canary-tutored individuals produced trills.

Representative songs from the four quiet-reared house finches are in Figure 4-3.



**Figure 4-3. Spectrograms of quiet-reared house finches.** Three seconds of quiet-reared house finch song.

### 4.3.1. Phrase features

Median group values, with interquartile ranges, are presented in Table 4-2.

**Table 4-2. House finch and canary phrase level measures.** Median group values (median of individual medians) with interquartile ranges in parentheses.

	Canary	House finch		
		Canary-tutored	House finch-tutored	Quiet-reared
Unique phrases	16.5 ( $\pm 10.25$ )	8 ( $\pm 6$ )	21 ( $\pm 12.5$ )	10 ( $\pm 4$ )
F0 (Hz)	2079 ( $\pm 874.7$ )	3289 ( $\pm 484.4$ )	3614 ( $\pm 283.4$ )	3452 ( $\pm 695.5$ )
Start F0 (Hz)	2062 ( $\pm 796.9$ )	3000 ( $\pm 937.5$ )	3750 ( $\pm 281.2$ )	3281 ( $\pm 1383$ )
End F0 (Hz)	2062 ( $\pm 632.8$ )	3375 ( $\pm 187.5$ )	3562 ( $\pm 187.5$ )	3562 ( $\pm 656.2$ )
Minimum F0 (Hz)	1688 ( $\pm 796.9$ )	3000 ( $\pm 750$ )	3000 ( $\pm 281.2$ )	2719 ( $\pm 609.4$ )
Maximum F0 (Hz)	2438 ( $\pm 703.1$ )	3562 ( $\pm 375$ )	4312 ( $\pm 46.88$ )	3938 ( $\pm 609.4$ )
F0 range (Hz)	843.8 ( $\pm 468.8$ )	562.5 ( $\pm 375$ )	1125 ( $\pm 562.5$ )	1125 ( $\pm 93.75$ )
Duration (msec)	29.33 ( $\pm 16.67$ )	77.33 ( $\pm 45.33$ )	93.33 ( $\pm 31.33$ )	129.3 ( $\pm 75.33$ )
Slope	4.167 ( $\pm 3.094$ )	1.881 ( $\pm 0.89$ )	3.031 ( $\pm 1.548$ )	1.654 ( $\pm 0.845$ )
Concavity <sup>†</sup>	0.59 ( $\pm 0.315$ )	0.28 ( $\pm 0.83$ )	1.65 ( $\pm 0.98$ )	1.21 ( $\pm 0.73$ )
Excursion	0.135 ( $\pm 0.074$ )	0.176 ( $\pm 0.049$ )	0.298 ( $\pm 0.23$ )	0.325 ( $\pm 0.085$ )

<sup>†</sup> Uses mean scores for individuals. Value presented is the median of the means for individuals.

Between canaries and house finches tutored in the wild, values differ for duration and all F0 values except for F0 range. When compared to canaries, wild-tutored house finches produced phrases that have higher frequencies and are longer.

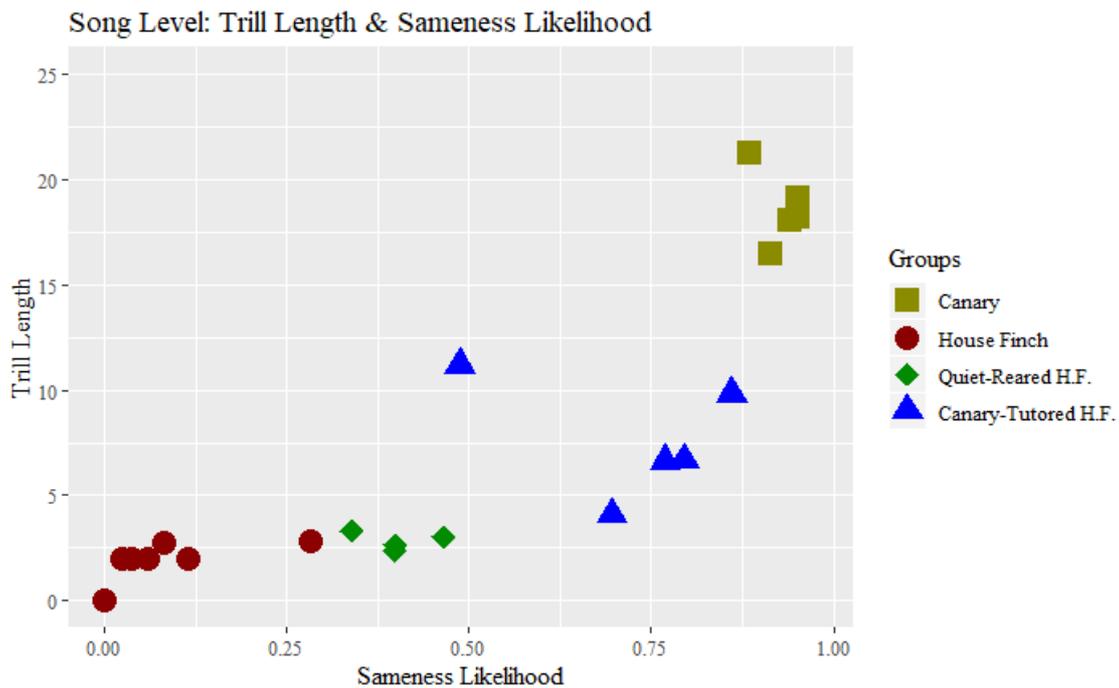
Among the three house finch groups, within-group variation obscures any obvious differences. While median scores differed for some acoustic parameters, interquartile ranges overlapped for every acoustic variable. Both the acoustically isolated group and the canary-tutored group were at the low end of the wild house finch values for slope, concavity, and number of unique syllables. The canary-tutored house finches fell to lower edge of wild house finch values for time-frequency excursion and F0 range.

The random forest classification of house finch and canary phrases produced an out-of-bag (OOB) estimate of error rate of 3.91%: 3.44% for canary phrases and 4.3% for house finch. As shown in Table 4-3, individuals from those two groups were tested on the model and members from both canary-tutored and quiet-reared groups had a majority of their phrases classified with house finch phrases. All quiet-reared individuals had over seventy percent of their phrases classified as house finch. One male from the canary-tutored group, D5, had a majority of his phrases classified with canary phrases, while another had almost half classified as canary.

***Table 4-3. Random Forest classification of house finch phrases.***

Group	Percent classified as house finch
Quiet-reared	A6 – 92.22
	B5 – 83.14
	D3 – 78.9
	D4 – 74.34
Canary-tutored	A5 – 87.03
	B1 – 62.25
	C6 – 52.98
	D2 – 85
	D5 – 13.04

### 4.3.2. Song features



**Figure 4-4. Length of repeated phrases by the sameness likelihood.** One canary (O16) removed from graph because of a very high median trill length (64.4).

As seen from Figure 4-4, the four groups were different with respect the length of repeated phrases and the likelihood that a phrase would be different from the phrase that preceded it. The Kruskal-Wallis rank sum test found a difference in sameness likelihood and trill length between the four groups (sameness: Kruskal-Wallis chi-squared = 19.565, df = 3, p-value = 0.0002; trill length: Kruskal-Wallis chi-squared = 18.77, df = 3, p-value = 0.0003).

The difference between canaries and house finches was statistically significant for both sameness likelihood and trill rate. Canaries have higher sameness likelihood scores (canary:  $0.94 \pm 0.03$  ~ house finch:  $0.06 \pm 0.07$ , adjusted p-value = 0.007) and longer trills ( $18.6 \pm 2.61$  ~  $2.0 \pm 0.4$ , adjusted p-value = 0.01).

The canary-tutored house finch group is statistically different from the other two house finch groups in sameness likelihood and trill length (Table 4-4). The canary-tutored house finch group had a sameness likelihood score of 0.77 ( $\pm 0.1$ ), which was higher than the 0.4 ( $\pm 0.03$ )

of the quiet-reared (adjusted p-value = 0.02) and the 0.06 ( $\pm 0.07$ ) of house finches (adjusted p-value = 0.007). The canary-tutored house finch group also had longer trills, with a median of 6.7 ( $\pm 3.18$ ) compared with 2.9 ( $\pm 0.49$ ) and 2.0 ( $\pm 0.4$ ) of the quiet-reared house finches and the wild-tutored house finches, respectively (adjusted p-values = 0.02; 0.01).

The canary-tutored house finches switched phrases more often and had shorter trills than the canary group (sameness: adjusted p-value = 0.009; trill length: adjusted p-value = 0.01).

The quiet-reared group was not statistically different from the wild-tutored house finch group in trill length (adjusted p-value = 0.07) but they did have a higher sameness likelihood score (adjusted p-value = 0.01). Sameness likelihood and percentage of phrases trilled are the only measures with a statistically significant difference between the quiet-reared group and wild-tutored house finches.

**Table 4-4. Group medians & p-values for song level.**

<b>Sameness likelihood</b>		<b>Quiet HF</b>	<b>Wild HF</b>	<b>Canary</b>
<b>Canary-tutored HF</b>	0.77 ( $\pm 0.1$ )	$p = 0.016$ *	$p = 0.008$ **	$p = 0.009$ **
<b>Canary</b>	0.94 ( $\pm 0.03$ )	$p = 0.011$ *	$p = 0.007$ **	-
<b>Wild HF</b>	0.059 ( $\pm 0.07$ )	$p = 0.009$ **	-	-
<b>Quiet HF</b>	0.4 ( $\pm 0.03$ )	-	-	-
<b>Trill length</b>		<b>Quiet HF</b>	<b>Wild HF</b>	<b>Canary</b>
<b>Canary-tutored HF</b>	6.7 ( $\pm 3.18$ )	$p = 0.019$ *	$p = 0.01$ **	$p = 0.01$ **
<b>Canary</b>	18.6 ( $\pm 2.57$ )	$p = 0.0143$ *	$p = 0.01$ **	-
<b>Wild HF</b>	2.0 ( $\pm 0.4$ )	$p = 0.066$	-	-
<b>Quiet HF</b>	2.9 ( $\pm 0.52$ )	-	-	-
<b>Trill rate</b>		<b>Quiet HF</b>	<b>Wild HF</b>	<b>Canary</b>
<b>Canary-tutored HF</b>	10.5 ( $\pm 5.42$ )	$p = 0.19$	$p = 0.035$ *	$p = 0.078$
<b>Canary</b>	13.3 ( $\pm 3.68$ )	$p = 0.038$ *	$p = 0.016$ *	-
<b>Wild HF</b>	0 ( $\pm 2.51$ )	$p = 0.110$	-	-
<b>Quiet HF</b>	5.65 ( $\pm 3.75$ )	-	-	-
<b>Max. trill rate</b>		<b>Quiet HF</b>	<b>Wild HF</b>	<b>Canary</b>
<b>Canary-tutored HF</b>	11.5 ( $\pm 3.77$ )	$p = 0.413$	$p = 0.013$ *	$p = 0.013$ *
<b>Canary</b>	31.5 ( $\pm 9.84$ )	$p = 0.309$	$p = 0.013$ *	-
<b>Wild HF</b>	0 ( $\pm 2.51$ )	$p = 0.138$	-	-
<b>Quiet HF</b>	5.78 ( $\pm 11.2$ )	-	-	-

*All p-values have been adjusted based on a Benjamani-Hochberg procedure for multiple tests.*

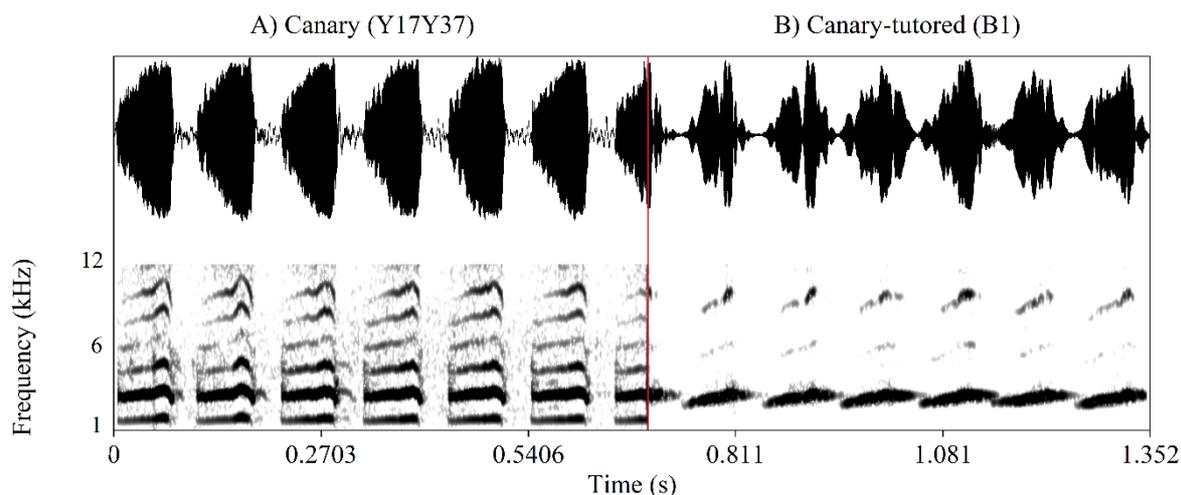
As with the previous metrics, trill rate and maximum trill rate differ between the groups (trill rate: Kruskal-Wallis chi-squared = 15.4, df = 3, p-value = 0.001; maximum: Kruskal-Wallis chi-squared = 14.612, df = 3, p-value = 0.002).

Trill rate and maximum trill rate are significantly different between canaries and house finches ( $13.3 \pm 2.41 \sim 0.0 \pm 2.51$ , adjusted p-value = 0.016;  $31.5 \pm 3.77 \sim 0 \pm 2.51$ , adjusted p-value = 0.013).

The canary-tutored group has a higher rate ( $10.5 \pm 5.42$ ) than that of the quiet-reared ( $5.65 \pm 3.75$ , adjusted p-value = 0.19) and a lower rate than that of canaries ( $13.3 \pm 3.68$ , adjusted p-value = 0.078) but in neither case is the difference statistically significant.

The difference between the canary and canary-tutored groups is significant in maximum trill rate ( $31.5 \pm 9.84 \sim 11.5 \pm 3.77$ , adjusted p-value = 0.013). Only one house finch, in any group, had a trill rate above 20 phrases per second. A6 produces a trill with a rate of 41 Hz, which was even higher than canary trills. However, this trill is an outlier in several respects: the longest exemplar is only 169 msec, it is in the upper range of the house finch frequency spectrum (7078 Hz), it's in the lower range of phrase duration (11 msec), and it has a higher phrase duration coefficient of variation than any other trill in the four groups (A6 trill  $C = 0.55 \sim \text{Overall} = 0.12 \pm 0.14$ ). This outlier is the reason why quiet-reared and canary maximum trill rates are not statistically significant. If removed, the interquartile range of the quiet-reared maximum trill rate falls from 11.2 to 3.47. Without the outlier trill, canaries and quiet-reared house finches are statistically different (adjusted p-value = 0.014) but quiet-reared and canary-tutored house finches are still not.

While I could not find the audio files of the exact canaries that tutored house finches, the canaries analyzed here did share common patterns with the canary-fostered house finches, as seen in Figure 4-5.



**Figure 4-5. Similar trill types in a canary and canary-tutored house finch.**

*I spliced the two trills together; the canary is the first half and after the red line is the canary-tutored individual. The fundamental of the canary-tutored house finch roughly corresponds with the 2<sup>nd</sup> harmonic of the canary. While the canary presented was not the direct tutor of the house finch many trills and phrase types were shared among the canaries.*

*A) Canary foster parent (male) Y17Y37. Trill type I. Trill rate = 9.24 Hz. F0 = 1680 Hz.*

*B) Tutored house finch B1. Trill type A. Trill rate = 9.3 Hz. F0 = 3130 Hz.*

I only found evidence of similar trills in the canary-tutored house finches and the canaries that had a trill rate below 25 Hz and a dominant frequency above 2 kHz (e.g., the second harmonic in Figure 4-5).

#### **4.4. Discussion**

This is the first experimental evidence that house finches learn their song. Without any acoustic input, house finches do not produce completely typical house finch song. Typical house finch song is produced with very little repetition in phrases, yet this was not the case for quiet-reared individuals. The failure to produce species-typical song is consistent with acoustic isolation experiments performed in other songbirds (Fehér, Wang, Saar, Mitra, & Tchernichovski, 2009; Konishi, 1965a). The other line of evidence for vocal production learning comes from the house finches who were tutored by canaries. All five of these individuals learned to trill. While acoustically isolated house finches produced repeated

phrases, canary-tutored individuals were much more likely to produce long strings of repetitions. Trilling is not a feature found in most house finch song, and when it is present it is at a slower rate and the trills are much shorter. All canary-tutored birds in this study trilled and some of the trills closely match canary tutor songs.

Similar to song sparrows, swamp sparrows, and white-crowned sparrows, house finches can learn phrase sequence patterns from other species, despite rarely using this ability in the wild (Kelley, Coe, Madden, & Healy, 2008; Marler, 1997; Marler & Peters, 1977; Soha & Marler, 2000).

While canary-tutored house finches learned to trill, the trilling is not completely canary-like. The limitation on faithful mimicry is in line with other work on learning biases in songbirds, which has found that song output deviates from heterospecific or atypical acoustic input. When exposed to trill rates faster than the species typical range, swamp sparrows insert breaks in the song, delete phrases, or reduce the trill rate (Podos, 1996). Lahti, Moseley, & Podos (2011) found that swamp sparrows also did not faithfully imitate artificially modified trill rates when those rates were slower than normal; instead they elevated the input rate to a more species-typical range. When waterslager canaries, which have been bred to produce low-frequency song, and border canaries, which have higher frequency song, were exposed to the same mixture of high and low frequency elements, each mimicked the phrases/tours in their breed-typical ranges (Mundinger & Lahti, 2014). Furthermore, Mundinger and Lahti found that the proportion of border or waterslager genes in border-waterslager hybrids and backcrosses predicted the number of high and low frequency elements produced (e.g., a bird with a higher proportion of border canary genes produces more high frequency phrases).

These house finch vocal learning data raise several questions related to the mechanisms and functions of house finch song: Why do house finches learn heterospecific song? If house

finches can trill, why do they not do so in the wild? And, why is heterospecific learning incomplete?

#### **4.4.1. Why do house finches learn heterospecific song?**

Vocal learning is a relatively rare trait, but learning the acoustic signal of other species is even rarer (Dobkin, 1979). Heterospecific mimicry is even less well understood than general vocal learning, particularly in terms of adaptive function (Kelley et al., 2008). Neither Kelley et al.'s (2008) review of heterospecific mimicry nor Garamszegi et al.'s (2007) comparative study of vocal mimicking European passerines were able to find strong support for any functional purpose for heterospecific mimicry, outside of a few narrow cases. One problem is that heterospecific mimicry varies greatly across avian species. Goller & Shizuka (2018) distinguish between two main types of vocal mimics: “flexible” and “incidental”. Flexible mimics, like mockingbirds (Gammon & Altizer, 2011) and European starlings (Hindmarsh, 1986), are those which are highly permissive in the sounds and species that they imitate. They typically have longer song learning windows and mimicry can be found in naturalistic settings. Incidental mimics, like chaffinches (Thorpe, 1958) or song sparrows (Marler & Peters, 1977), are those who only mimic in atypical situations, such as in captivity or in experimental settings. Whether this distinction is a difference of degree or kind, however, remains unclear. Some species seem to be very permissive, but, in the wild, do not include heterospecific sounds. The budgerigar is an adept mimic and is able to learn novel vocalizations throughout its entire life, yet no evidence exists of anything other than conspecific mimicry in wild budgerigar populations (Gramza, 1970). House finches song is consistent across populations and house finches do not typically integrate heterospecific sound into their repertoire (Bitterbaum & Baptista, 1979; Mundinger, 1975; Pytte, 1997; Tracy & Baker, 1999). However, more evidence for wild mimicry exists in house finches than in the more permissive budgerigar. Baptista (1972) reported a single wild individual mimicking a white-crowned sparrow song and R.

Payne et al. (1998) reported a small colony of house finches with canary tours. These data and the data presented in this study suggest that house finches may be more permissive in vocal learning than is evidenced from species-typical populations.

Garamszegi et al. (2007) suggested that Hindmarsh's (1986) "learning mistakes hypothesis" for heterospecific mimicry was the most well-supported, so far, by comparative work. Hindmarsh rejects a functional explanation for most heterospecific mimicry and argues that the use of different species vocalizations in song is a byproduct of complex song. As songs get more complex, the individual units which comprise song become less important. In support of this claim, he states that the acoustic structure of the phrase is more important for species-/individual-recognition in species with simpler songs, as is the case in swamp sparrows (Peters, Searcy, & Marler, 1980). For those species with complex song, the individual phrases matter less, and attention is focused on more global song patterns. For example, skylarks treat songs with synthetic phrases similar to natural song, provided there is enough frequency modulation over the whole bout (Aubin & Bremond, 1983). Because individual constituents become less important in complex song, Hindmarsh argues, cognitive filters which block non-species typical sounds will be reduced. The filter relaxation will cause more heterospecific sounds to pass through and be integrated into the song.

At first glance, the house finch data presented here do not seem to support the learning mistakes hypothesis. House finches learn canary sequencing while I found only limited, suggestive evidence that canary phrases were learned – e.g. a majority of D5's phrases were classified as canary. These data are seemingly in the opposite direction of Hindmarsh's (1986) hypothesis, which suggests that the filters on the acoustic structure of phrases are relaxed. However, the crux of the learning mistake hypothesis seems to be that the lack of cognitive filters will permit "mistakes" in song learning. Filters for song organization and for song composition may be distinct and may vary between species depending on the specific

functional needs of a species. Swamp sparrows learn song sparrow song organization as long as their own phrase types are used (Marler & Peters, 1977). Since swamp sparrow song organization differs between populations (Balaban, 1988), sequencing seems to be less genetically inherited than acoustic phrase structure. The data presented here suggest that filters on phrase patterning may be relaxed in house finches.

#### **4.4.2. Why don't wild house finches trill?**

Trilling is clearly learnable by house finches. Furthermore, when reared without song input, house finches repeat phrases. House finch trilling does occur in the wild, though it is not widespread, and it is usually at a much slower rate than canary-like trills. So why is it that trilling, or at least higher rates of repetitions, is not more common in typical house finch song?

The similarity in house finch song across North American populations may suggest some functional purpose for lack of trilling. Sexual selection is, at least for some species, a driver of song complexity and the number of unique units in a song is one metric for complexity (Byers & Kroodsma, 2009). In house finches, male song complexity predicts both female preference and male reproductive success (Mennill, Badyaev, Jonart, & Hill, 2006; Nolan & Hill, 2004). However, both Nolan & Hill (2004) and Mennill et al. (2006) found this to be the case for longer and faster songs, but not songs with a diversity of phrase types. Length and speed in song are more obviously linked to honest signaling than repertoire size. Longer songs require more energy and higher rates require more skill in vocal modulation (Gil & Gahr, 2002; Podos et al., 2016).

While the number of unique units may not be an honest cue, song could play a role in mate choice by acting as a cue to group identity. Nottebohm (1972) hypothesized that vocal learning could promote assortative mating by the causing local dialects to form. Females could use acoustic cues to recognize which males had the same song as their father and then mate with those individuals. House finch song does have dialects and the amount of similarity in song

organization decreases as a function of the distance between them (Bitterbaum & Baptista, 1979; Pytte, 1997; Tracy & Baker, 1999). If house finch song served to promote assortative mating, trilling may be less optimal than switching phrase types. Switching phrases would not only give an individual more unique phrases per unit of time but phrase transitions serve as an additional cue. However, with the exception of 1970s Northeastern U.S. populations (Mundinger, 1975), house finch song sharing is clinal. Assortative mating needs clear dialect boundaries to function, yet these are absent for most house finches. Furthermore, female house finches do not show a preference for songs they have heard early in life; though, they may prefer songs from individuals they are genetically related to (Hernandez & MacDougall-Shackleton, 2004)

The reason for which house finches do not trill more often is not clear. Female house finches do not show a preference for more phrase diversity nor do they show a preference for the dialect they were exposed to as a hatchling. Relatively little work has been done on female preference in house finch song, so female choice certainly cannot be ruled out as a potential driver of phrase-to-phrase diversity. Changing from one phrase to another may carry some fitness cue that is relevant for female house finches.

#### **4.4.3. Why is learning incomplete?**

While the house finches were able to learn some aspects of canary song, mimicry was not absolute. Canary-tutored house finches had significantly fewer phrases in their trills when compared to the canary tutors. Furthermore, maximum trill rate for canary-tutored house finches was slower than the median rate for canaries. I also found very little evidence that canary-tutored house finches learned the acoustic structure of phrases from canaries. A few individuals seemed to have some canary-like phrases, but most canary-tutored phrases looked more house finch-like than canary-like.

Serval potential, and non-mutually exclusive, explanations exist to account for the lack of learning. The developmental stress hypothesis (MacDougall-Shackleton & Spencer, 2012; Nowicki et al., 1998), the auditory template hypothesis (Konishi, 1964; Marler, 1970a; Marler & Sherman, 1983), and the vocal tract/motor constraints hypothesis (Podos, 1996) could potentially explain why the house finches in this study did not completely learn canary song.

#### **4.4.3.1. Developmental stress hypothesis**

Females of many songbird species show a preference for males with more complex song, but it is not completely clear why females should treat more complex or elaborate songs as an “honest” signal (Nowicki et al., 1998). In order for song to be an honest signal of mate quality it must incur a cost to the signaler; otherwise, any individual could produce the signal (Zahavi, 1975). Long or high amplitude songs are energetically costly (Searcy & Yasukawa, 1996), but other complex song features are less obviously tied to fitness. Complex song could, theoretically, allocate costly neural resources or increase the rate of predation. Evidence in support of these factors is limited, though, especially for the neural costs (Gil & Gahr, 2002). The developmental stress hypothesis, on the other hand, provides a potential mechanism for complex song to serve as an honest indicator of fitness (Nowicki et al., 1998).

The developmental stress hypothesis (Nowicki et al., 1998) suggests that the crystalized song of an adult is an honest indicator of how well that individual was able to cope with stress during its development or how well the parents were able to shield the individual from stress. Experimental evidence has confirmed that stress during development can produce smaller phrase repertoire size, lower peak frequencies, slower production rates, and less accurate copying (MacDougall-Shackleton & Spencer, 2012). For instance, restricting food results in reduced mimicry ability in swamp sparrows, parasite infections reduced canary phrase inventories, and corticosterone (a stress hormone) injections caused lower peak frequencies in zebra finches (MacDougall-Shackleton & Spencer, 2012).

The house finches in this study had somewhat similar patterns to stress-affected birds: maximum frequencies for some of the acoustically isolated house finches were lower than for the wild house finches, the rate of trilling for the canary-tutored house finches was slower than the canary tutors, and phrase inventories for the two canary-fostered house finches groups were smaller than the inventories for the foster canaries and wild house finches. While these data may be consistent with developmental stress, at least for some individuals, they are not convincing evidence. The average maximum frequency for the quiet-reared house finches was only slightly lower than the wild house finch group. The canary-tutored house finches might have had slower trill rates than canaries, but they were faster than wild house finches. Furthermore, trill rate has an alternative explanation which is more well-supported in the literature (Podos, Southall, & Rossi-Santos, 2004; discussed in 4.4.3.3). For inventory size, the quiet-reared house finches should be expected to have a small inventory since they were not exposed to song. The smaller phrase inventory of the canary-tutored house finches could suggest an effect of stress, but the evidence here is not convincing. Canaries have smaller inventories than wild house finches and unlike wild house finches, the canary-tutored house finches only had input from two acoustic models. The experimental design could have been better to allow for more parallel comparisons between groups. Had the house finches exposed to house finch song been tutored in a lab setting, rather than in the wild, I would have a much better idea if the environmental stress hypothesis was a factor – or really, if the lab affected song development in any way.

Further details suggest that developmental stress was not a strong factor in the imperfect learning of canary-tutored house finches. The house finches experienced none of the stressors that have been shown to reduce song quality; they were never restricted from food and water, they were not competing with a large brood, and they were healthy and parasite free. They even

had social interaction during the perceptual stage of song development since they were all fostered by a canary pair.

The introduction of white noise could have added a stressor, though the effects of noise on development and stress are unclear. Higher levels of corticosterone have been found in greater sage-grouse and chickens exposed to high levels of traffic noise (Blickley et al., 2012; Campo, Gil, & Dávila, 2005). However, Crino, Johnson, Blickley, Patricelli, & Breuner (2013) found that white-crowned sparrows exposed to traffic noise during development had lower stress levels than the control group. Domesticated chicks exposed to loud (80-95 dB) white noise for several days had no behavioral or hormonal indications of consistent with stress (McFarlane & Curtis, 1989; McFarlane, Curtis, Shanks, & Carmer, 1989). White noise seems to have fewer deleterious effects than traffic noise or other environmental noises. White noise is a spectrally consistent wall of noise, so, it may even have some benefits by masking stressful intermittent noise (Dooling & Popper, 2007; Rabat, 2007).

#### **4.4.3.2. The auditory template hypothesis**

The auditory template hypothesis states that birds are born with a song template that filters out acoustic signals that do not match with the template (Marler, 1970a). As there may be several species in a given area, the template serves to ensure the bird learns its species typical song and not an alien song. Most of the evidence for a template comes from the song of acoustically isolated individuals. The American robin, Mexican junco, and white crowned sparrow produce song even without acoustic input, though their structure was much simpler (Konishi, 1964, 1965a, 1965b). Some birds are able to learn their song with very little input, perhaps suggesting some preexisting knowledge. Nightingales, for instance, need as few as fifteen experiences with a song to be able to replicate it (Hultsch & Todt, 1989). Physiological changes after conspecific song exposure have also been noted. Dooling & Searcy (1980) found that swamp sparrows' heart rate slowed down more when exposed to conspecific song than to

song sparrow song or canary song. Interestingly, this difference was not found in the song sparrows who are better at learning heterospecific song than swamp sparrows (Marler & Peters, 1977).

Specific details as to the nature of the template that mediates between the genetic song structure and the environment remain theoretical and abstract, but the data presented are consistent with the model's predictions. All the isolated birds produced song, suggesting some canalized structure is present at birth. The acoustic structure of the phrases do not vary across individuals, either. Most phrases fall into three broad categories: flat periodic signals, sloping periodic signals, and aperiodic buzzes. None of these phrases have much frequency modulation and most are within a 100-250 millisecond range. In short, they look like a basic phrase template that could be lengthened, shortened, frequency modulated, and/or combined.

However, a template hypothesis does not seem to be able to account for the song patterning data as well. The organizational patterns of the canary-tutored house finches and the acoustically isolated individuals are very different from wild-type song. As mentioned in 4.4.1, filter reduction seems likely to explain why heterospecific song is learned in the first place.

#### **4.4.3.3. Vocal tract/motor constraints hypothesis**

The vocal tract/motor constraints hypothesis postulates that genetically inherited cognitive constraints are often not necessary to explain behavior, as much of song and song development are shaped by physical morphology (Podos, 1996). Several aspects of the tutored song suggest that motor constraints play a significant role in the inability to accurately mimic canaries. The median maximum trill rate for the canary-tutored individuals was a full 20 trills/second below the canary median. Trills were also shorter, suggesting that when the house finches did trill, they were limited in how long they could do so. Podos (2001) found a high correlation between song type and physical morphology in Darwin's finches, with larger beaked finches producing lower trill rates with narrower frequency ranges.

Trill rate also correlates with general body size. Songbirds trills can be produced by two mechanisms: pulsatile expiration or mini-breaths (Zollinger & Suthers, 2004). In the former, the songbird takes a breath before producing the trill. Air supply decreases and the individual must stop the trill to take another breath. In the latter, the songbird takes breaths in between each short unit in the trill. Mini-breaths allow for much longer trill durations, but as trill rates increase, mini-breaths become more difficult (Hartley & Suthers, 1989). Evidence from multiple songbird species suggest a relationship between body size and the rate at which a bird must switch from mini-breaths to pulsatile expiration (Hartley & Suthers, 1989; Suthers, Goller, & Hartley, 1994; Wild, Goller, & Suthers, 1998; Zollinger & Suthers, 2004). Zollinger & Suthers (2004) states that 18 gram canaries switch at higher rates of 30 trills/sec, 40 gram cardinals switch at 16 trills/sec, and the 50 gram mockingbirds switch at 10 trills/sec. House finches have a body mass of roughly 22 grams, Because house finches are slightly larger than canaries, roughly 22 grams (Badyaev, Whittingham, & Hill, 2001), precise mimicry of canary trills should be unexpected.

Differences in the fundamental frequency of canaries and the canary-tutored house finches likely has a physiological explanation as well. Podos, Southall, & Rossi-Santos (2004) and Palacios and Tubaro (2000) linked body size to song frequency. Bill size, as well, has an effect on frequency. Giraudeau et al. (2014) found that urban house finches have longer beaks, the result diet changes associated with an urban environment. The beak size has had an effect on song as the urban house finches have lower maximum frequencies. In this study, the canaries that tutored the house finches were all waterslager canaries which were bred for their low frequency song (Güttinger, 1985). Fundamental frequency for the canaries was 2.1 kHz. For all house finches, fundamental frequency was much higher, roughly 3.5 kHz. Median minimum frequencies were similarly much lower; canary minimum was 1688 Hz ( $\pm 796$ ) while minimum frequencies for the canary-tutored house finches were almost twice as high (3000 Hz  $\pm 750$ ).

House finch vocal morphology could prevent a lot of mimicry of canary song, especially when the phrase or sequence was more articulatorily demanding.

Interestingly, the canary-tutored house finches did seem to find ways of mimicking phrases with fundamental frequencies below their natural range. The canary sequence in Figure 4-5 has a fundamental around 1.7 kHz, which is toward the low end of house finch productive and auditory range (Dooling, Zoloth, & Baylis, 1978). The second harmonic, however, is within the house finch range and at least individual, B1, seems to have matched its fundamental with that of the canary's second harmonic. A higher pitched individual matching its fundamental with that of the second harmonic of a lower pitched individual is a strategy that is often employed in human music (known as "octave equivalence"). Humans tend to perceive a frequency (e.g., 440 Hz) and its double (880 Hz) as more similar than two frequencies that are close to each other (e.g., 440 & 466.16 Hz) (Hoeschele, Weisman, & Sturdy, 2012). This ability is potentially used to aid children in the mimicry of lower pitched adult speech (Peter, Stoel-Gammon, & Kim, 2008). Research with octave perception in avian species has so far failed to find an effect (Hoeschele, Weisman, Guillette, Hahn, & Sturdy, 2013; Wagner, Mann, Afroozeh, Staubmann, & Hoeschele, 2019).

To test some of these hypotheses it would be valuable to test house finch song development with border canary song. Border canary song structure is similar to waterslager song but is closer to the frequency range of house finches. Border canary song would provide a control for frequency and test the limits of house finch trilling. House finches are only slightly larger than canaries, they can learn a diversity of phrases, and, as this experiment has shown, they can learn trilling. I predict that house finches would mimic models more accurately with border canary song than waterslagers. A cross-generational tutoring experiment could also help clarify some of the questions raised by this study. Fehér, Wang, Saar, Mitra, & Tchernichovski (2009) used isolate zebra finches as first-generation tutors and found that song became more like the

wild-type song in each successive generation. The shift to the wild-type song in zebra finches suggests some narrowing mechanism which promote certain song features over others. House finches could, over time, introduce phrase-to-phrase variation. If trills remain stable across generations, the lack of trills in wild populations may be a random, historical accident. If trills remain in the population, but trill rates and trill lengths are modified this could suggest size and beak morphology are significant factors in limiting house finch trilling. Rapid introduction of phrase-to-phrase variability may suggest that the number of unique units per song serves a functional purpose in house finches.

## Chapter 5. Variation and uniformity in spoken language syllable onsets: the role of sonority in speech

### 5.1. Introduction

While human language is one of the most complex communication systems, it is also the system with the greatest wealth of research. Centuries of work has gone into documenting, describing, and analyzing languages across the world (Blevins, 2004, 2007; Blust & Trussel, 2013; Campbell, 1997; Croft, 2003; Greenberg, 1965; Ladefoged & Maddieson, 1996; Sapir, 1925; Schleicher, 1877; Wetzels & Mascaró, 2001). This research has clearly demonstrated two important facts: The first, generalizations can be made that hold for the majority of languages; the second, these generalizations, more often than not, have exceptions (Blevins, 2004; Greenberg, 1965). Because of the depth and breadth of research, human language serves as a profitable testing ground for any discussion of variability and uniformity.

Every language, that we know of, has restrictions on the how sounds are organized. For instance, English does not allow /ŋ/, the sound at the end of *sing*, to come at the start of a word. Under a null model of sound distribution there should be little asymmetry in the types of restrictions seen across languages. However, this is not the case. For instance, a majority of languages do not permit sequences like *rt* or *lb* to come at the beginning of a syllable (Greenberg, 1965). That is, speakers perceive these sequences as two syllables: *r.ta* and *l.ba* (Berent et al., 2007).<sup>9</sup> A simple reversal of the first two sounds, however, is treated differently: *tr* or *bl* are relatively common in syllable onsets and *tra* and *bla* are perceived as one syllable, not two (Berent et al., 2007; Greenberg, 1965).

Importantly, the generalization that *tra* and *bla* are perceived as one syllable while *rta* and *lba* are two holds for most languages, but not all. Russian has words like *rta* ‘mouth.gen’,

---

<sup>9</sup> Similar to how an English speaker may perceive words like *LeBron* and *return* as having two syllables even if the first vowel is heavily reduced as to be barely present.

which are treated as a single syllable by native speakers. Experimental data shows that these speakers perceive a sequence like *rta* as only having one syllable (Berent et al., 2007). Cross-linguistic research suggests languages like Russian are rare, though just how rare is not entirely clear.

The distributional patterns of syllable organization are not simply found with a random assortment of segments, but can be described in terms of category and acoustic properties (Ladefoged & Maddieson, 1996). In the case of *rta* and *lba* sequences and *tra* and *bla* sequences, the asymmetry is rooted in sonority and the role of sonority in the syllable (Blevins, 1995; Clements, 1990; Hooper, 1972; Parker, 2002). Sonority is a difficult to define concept (Ohala & Kawasaki-Fukumori, 1997; Parker, 2002). It has been defined in terms of articulation: e.g., degree of openness in the vocal tract; acoustics: e.g., loudness, and perceptual: e.g., if it functions as a syllable nucleus (for a comprehensive review see Parker, 2002). The most robust evidence for a physical correlate of sonority comes from Parker (2008). For English, Quechua, and Spanish, Parker found a strong correlation between theoretical sonority values and relative acoustic intensity.

Consistent across definitions of sonority is that vowel sounds are the most sonorous segments and plosives the least. How the other segment classes are divided and placed within the “sonority hierarchy” varies, though the model in Figure 5-1 is a standard organization:

Vowels	Glides	Liquids	Nasals	Obstruents
/a, i, u/	/j, w/	/l, ɹ, r/	/n, m/	/s, f, b, t, k/

**Figure 5-1. Sonority hierarchy.** (Clements, 1990).

Researchers divide the broad sonority classes up differently. Parker (2002) mentions that many don’t distinguish glides and vowels since the distinction follows from syllable structure. Blevins (1995) presents a binary branching model which allows for more fine-grained

distinctions. The top branch has vowels on one side and consonants on the opposite. Branching off from the vowel node, vowels with a more open oral tract, like [a], are more sonorous than those with more closed oral tracts, e.g., [i].

Despite the lack of a clear, widely agreed upon definition, sonority is still widely used because of its connection to the syllable. As far back as Sievers (1876), syllables have been linked to loudness. Extensions of this idea brought about the concept of sonority and that cyclical rises and falls in sonority could serve as a measurable basis for syllables. The link between sonority and syllabification is known as the Sonority Sequencing Generalization, or the Sonority Sequencing Principle (Blevins, 1995; Clements, 1990; Hooper, 1972). A recent definition, from Parker (2002), is stated in (5-1).

**5-1. Sonority Sequencing Generalization.** (Parker, 2002, p. 8)

- a. In every syllable there is exactly one peak in sonority, contained in the nucleus.
- b. Syllable margins exhibit a unidirectional sonority slope, rising toward the nucleus.

As can be seen in the sonority hierarchy in Figure 5-1, the sound *r* is classified as a liquid which has higher sonority than *t*, an obstruent. Therefore, the cross-linguistic rarity of syllables with an initial sequence of *rt* follows from the statement in (5-1): *rta* has two peaks and is therefore perceived as two syllables; *tra*, on the other hand, only has one peak, so it is only one syllable. An example in English would be with the word *train* /tɹeɪn/. In this word, the segment with the highest sonority is /eɪ/, so it functions as the nucleus. Sonority falls unidirectionally from the vowel nucleus to the /t/ in the onset and the /n/ in the coda, resulting in the perception of one syllable. The past tense of *trained* /tɹeɪnd/ is still perceived as monosyllabic since /d/, an obstruent, is lower in sonority than /n/. The fall is still unidirectional. Switching the /t/ and the /r/, /ɹteɪn/, no longer produces single unidirectional fall from the nucleus. /r/ is higher in sonority than /t/, so /ɹteɪn/ will be perceived as two syllables in languages that adhere to the Sonority Sequencing Generalization. English doesn't have a /rɹ/ onset, yet in fast speech it would not be uncommon to produce [ɹteɪn] for the word "retain".

The majority of known languages follow the Sonority Sequencing Generalization and no known language permits a “sonority reversal” onset, like /rt/, while not allowing a “sonority rise” onset, like /tr/ (Greenberg, 1965). As such, it requires an explanation.

Because some languages have clusters that do not adhere to the Sonority Sequencing Generalization, the generalization is better treated as a strong tendency rather than an inviolable cognitive constraint (Blevins, 1995, 2004, 2006b, 2007, 2015, 2017; Daland et al., 2011). Patterns which violate the Sonority Sequencing Generalization are clearly learnable. The rarity of Sonority Sequencing Generalization violations is likely due to phonetic factors and how those phonetic factors can lead to ambiguity and reanalysis of the signal. Sound patterns that are more confusable with other patterns are less stable over the course of repeated interactions. As such, they are predicted to be rarer than less ambiguous patterns (Blevins, 2004; Wedel, 2006). Sonority Sequencing Generalization violations may have properties which could give rise to misperception in the signal (Daland et al., 2011).

In this chapter, I approach the Sonority Sequencing Generalization from the perspective of Evolutionary Phonology (Blevins, 2004, 2006a, 2015, 2017). Evolutionary Phonology argues that many common sound patterns reflect common “natural”, phonetically motivated, sound changes. Natural sound changes are based in articulatory, perceptual, and general cognitive biases. As such, there is no need to represent these biases as universal phonological constraints which are specific to linguistic grammars. Within Evolutionary Phonology the initial assumption is that, like other phonological tendencies, the Sonority Sequencing Generalization is best explained in general aspects of human speech perception and production and exceptions to the Sonority Sequencing Generalization are expected under certain specific conditions. Evolutionary Phonology provides a framework for examining conformity to, and deviation from, the Sonority Sequencing Generalization in human speech.

A first step in assessing the role of the Sonority Sequencing Generalization in speech is to create an accurate description of the sound patterns. So, the first goal of this chapter is to take an inventory of languages which allow Sonority Sequencing Generalization violations and assess how rare these violations are. I focus specifically on Sonority Sequencing Generalization violations in the onset of a syllable. Onsets are more cross-linguistically common than codas and most languages seem to have more restrictions on sounds in the coda than they do in the onset (Blevins, 2006b). An inventory of Sonority Sequencing Generalization violations requires not only assessing whether the sound patterns exist, but also evaluating how they are perceived by speakers. A language may have a word like [rta] but still conform to the Sonority Sequencing Generalization if speakers treat the *r* as syllabic and syllabify the word as [r.ta]. Because common and rare sound patterns are often the result of common and rare sound changes, respectively, the second goal of this chapter is to analyze historical data from languages and language families which have Sonority Sequencing Generalization violations.

## 5.2. Methods

In this typological survey, I collected data from six unrelated languages that have sonorant + obstruent onset sequences which are described as belonging to a single syllable. Languages of interest have initial #RTV sequences, where R is a sonorant, T is an obstruent, and RT are tautosyllabic, that is, they belong to the same syllable. I define an onset cluster as a word-initial sequence of two or more tautosyllabic consonants without any intervening *phonological* material. Phonological, unlike phonetic, refers to perceptual categories and contrasts that are meaningful in a language whether involving segments, syllables, tone, or other features. Within a category, the acoustic signal can vary slightly, and the variations are often not perceptible to speakers; in these cases, the variations are phonetic. To illustrate, the sounds [ɹ] and [l] are phonological categories for English speakers. Switching one out for the other creates a lexical contrast: *rap* /ɹæp/ vs. *lap* /læp/. This contrast is not phonological in Japanese as the two sounds

do not distinguish meaning and they are often perceived as members of a single category, /t/ (Miyawaki et al., 1975). For onset clusters, a vowel could be present between two consonants without, necessarily, being a phonological vowel. During the production of consonant clusters, phonetic vowels can be inserted due to gestural mistiming or a language-specific features associated with consonantal release (Davidson & Stone, 2004). Likewise, phonological vowels may be reduced, devoiced, or lost due to gesture overlap (Chitoran & Iskarous, 2008). The acoustic signal itself is not conclusive evidence for or against a sequence being an onset cluster. In fast speech, English speakers may completely drop the [o] from *potato* /potaɪto/ > [p<sup>h</sup>tʰaɪt<sup>h</sup>oʷ], yet speakers still perceive a vowel. The opposite, where a non-phonological vowel is inserted within a cluster, could occur in hyperarticulated speech, as when an English speaker emphatically utters *please* /pliz/ > [p<sup>h</sup>əliːz].

The starting point of this survey was Greenberg's (1965) and Kreitman's (2008) typologies of onset clusters. I carefully examined each language in these earlier studies and defined a subset of the most plausible sonorant + obstruent onsets. From this group, I chose languages based on geography and historic unrelatedness. The language families included in the survey are Austronesian, Indo-European, Kartvelian, Maipurean, Oto-Manguean, and Penutian. Many of these language families have multiple languages that have sonority reversal onsets, but to have independent data points, I only used one language per family. In choosing the one language, I tried to choose a language that was well documented and for which strong conclusions could be made about syllable structure. For the six languages, I reviewed dictionaries and fieldwork descriptions to build the sonorant + obstruent onset inventory. I consulted secondary sources for information related to syllable structure, morphological alternations, and speaker intuitions. The six languages chosen are briefly described here.

Russian, a major world language, is a Slavic language and Slavic languages form part of the larger Indo-European family. Russian is the dominant language of Russia and is spoken as

a first language by roughly 154 million people. Because it was the dominant language of the former Soviet Union (1922-1991), it served as the *lingua franca* for much of Eastern Europe and Northeast Asia. Russia's influence diminished with the dissolution of the Soviet Union, but about 13 million people still speak it as a second language and in some countries, like Latvia, Russian is considered a useful skill in the job market (Sussex & Cubberley, 2006).

Among Slavic languages, complex consonant sequences are widespread, but not all languages have sonority reversals in initial clusters. Russian and Polish have sonority reversals while Macedonian and Slovenian syllabify #RTV as #R.TV with an initial syllabic sonorant. Czech shows a mixed pattern. It has syllabic /r/, but permits /l/ and nasals to be in sonority reversal clusters (Scheer, 2007; Sussex & Cubberley, 2006).

Georgian is a Kartvelian language primarily spoken in the Republic of Georgia. The Kartvelian language family has two main branches, Svan and Georgian-Zan. The former has only one language, Svan. The latter is divided into Georgian and two Zan languages, Mingrelian and Laz. All of the Kartvelian languages are spoken in in the Southwestern Caucasus, nestled between the Black Sea, Russia, Turkey, Armenia, and Azerbaijan (Klimov, 1998). All Kartvelian languages have a large number of initial consonant clusters, but Georgian is the best described and most widely spoken language in this small family.

Piro, or Yine, is a Maipurean, or Arawakan, language spoken in the Amazonian area of Brazil and Peru. Maipurean is the largest language family in the Americas with languages in the Caribbean, Central America, and throughout South America (Campbell, 1997). At one point, Maipurean languages spanned as far north as Belize and as far south as Paraguay. Piro belongs to the Southwestern branch of Maipurean. It is unique in that it is the only language in its family to have complex consonant sequences (Aikhenvald, 1999, p. 78).

Chatino is an Oto-Manguean language spoken in the Mexican state of Oaxaca. There are three main subdivisions of Chatino: Zenzontepec, Coastal, and Eastern Chatino, though the

languages are not mutually intelligible (Campbell, 2013, p. 395, fn. 2). The group which has had the most attention, particularly in the last decade, is the Eastern Chatino group. Within the Eastern Chatino languages there exist a variety of syllable structures (Campbell, 2013). In this language sub-group, pre-existing reconstructions and differences in syllable structure across languages provide insight into the historical origins of word-initial sonority reversals. Here, I focus on the variety of Chatino spoken in Panixtlahuaca (Pride & Pride, 2010). Sonorant-obstruent onset sequences seem to be present in many Chatino varieties, but the sequences seem to be most widespread in Panixtlahuaca Chatino (Campbell, 2013; Pride & Pride, 2010).

Klamath is an endangered language spoken in south central Oregon in the Klamath Basin. Klamath has been classified as part of the Plateau Penutian languages which also includes Molala and the two Sahaptian languages, Sahaptin and Nez Perce (Campbell, 1997, p. 318). A. L. Kroeber and Roland B. Dixon proposed a broad language family of Penutian which included many of the indigenous languages of the Pacific Coast of the United States, however the family grouping is still controversial (DeLancey & Golla, 1997).

Tsou is an indigenous Austronesian language of Taiwan. According to Blust (2009, p. 31), it belongs to the Tsouic subgroup of which Saaroa and Kananabu, two other languages of Taiwan, are also member. However, its within-Austronesian sub-grouping is still debated (Chang 2006). The dialects of Tsou are all mutually intelligible, though the Duhtu variety seems to have more sonorant + obstruent onset clusters than the Tfuca or Tapaŋ varieties. In the latter two, the rhotic /ɻ/ has been lost (Tscuchida, 1972). Wright (1994) states that the rhotic has become a glide in those dialects while Tscuchida (1972) and Chen (2002) state it is an /e/. Even in Duhtu, the rhotic may be falling out of the language and it is more common in older speakers than in younger speakers (Tscuchida, 1972; Wright, 1994).

Table 5-1 summarizes the languages used for this study and illustrates the word-initial clusters of interest.

**Table 5-1. Sonorant + obstruent onset languages.**

Family	Sub-group	Language	Sonorant + obstruent onset
Austronesian	Tsouic – proposed	Tsou	<i>ɬtuu</i> ‘loquat tree’
Indo European	Slavic	Russian	<i>rta</i> ‘mouth.gen’
Kartvelian	Karto-Zan	Georgian	<i>mgeli</i> ‘wolf’
Maipurean (Arawakan)	Southern Maipurean	Piro (Yine)	<i>wkata</i> ‘direction toward this place’
Oto-Manguean	Zapotecan	Panixtlahuaca Chatino	<i>lkaʔ</i> ‘room’
Penutian – proposed	Plateau Penutian	Klamath	<i>lkʻom</i> ‘charcoal’

In choosing these six languages, many factors were taken into account. I excluded languages that had *only* homorganic nasal-obstruent onset clusters because it is difficult to determine the phonological status of sounds that are often orthographically represented as a nasal + obstruent sequence. These forms could be true nasal + obstruent onset clusters, or they could be single segments – prenasalized obstruents or postploded nasals. Another possibility is that a nasal + obstruent series is a sequence of a syllabic nasal followed by a non-tautosyllabic obstruent (Cohn & Riehl, 2008). Downing (2005) and Ladefoged & Maddieson (1996) argue that there is no necessary or consistent cross-linguistic phonetic difference between prenasalized consonants and nasal + stop sequences, though there may be a phonological difference. In contrast, Cohn & Riehl (2008) contend that differences do exist in the durational ratios between prenasalization, postplosion, and true consonant clusters in some languages.

I focused most of my attention on clusters that contain either taps/flaps, trills, rhotic approximants, or laterals as the first member of the cluster and stops or fricatives as the second member. There is less risk of analyzing a liquid + obstruent sequence as anything other than

two segments. Pre-rhoticized and/or pre-lateralized obstruents are rare. Where syllabicity of the sonorant is in question, I discuss this on a case by case basis.

For second members of the clusters, I included sibilants. Sibilants are phonologically and phonetically exceptional which could make analysis more difficult (see Fleischhacker, 2001 for a detailed discussion). Sibilants are considered obstruents, yet in many languages they do not behave like other obstruents, particularly with respect to the Sonority Sequencing Generalization. Sibilant + obstruent onsets are found in languages which otherwise do not allow obstruent + obstruent onsets, sonority “plateaus” (Goad, 2016). English does not allow most #OO onsets where the first member is a non-sibilant, like \*\*/tkʌl/, \*\*/fpraj/ or \*\*/θtaɪ/, yet allows /skʌl/, /spraj/, and /staɪ/. For many of the world’s languages, if only one cluster violates the Sonority Sequencing Generalization, the cluster will include a sibilant (Fleischhacker, 2001). If the Sonority Sequencing Generalization is an active component of mental grammars, sibilant + obstruent clusters should be as rare as other obstruent + obstruent clusters, which is not the case (Fleischhacker, 2001; Goad, 2016). To get around the sibilant + obstruent onset under-generation problem of the Sonority Sequencing Generalization, some researchers have treated sibilants as an appendix to syllables or as a member of complex segments (discussed in Henke et al. 2012).

If the Sonority Sequencing Generalization is a descriptive generalization, the exceptionality of sibilants is less problematic (Henke et al., 2012). The differences between their behavior in consonant clusters and that of other obstruents may be rooted in phonetics (Fleischhacker, 2001; Henke et al., 2012; Wright, 2004). Like other fricatives, the source of energy for sibilants is turbulent air caused by a narrow constriction in the vocal tract. What sets them apart from other fricatives, however, is that most of the acoustic energy is not directly the result of the articulatory constriction. The stream of air that passes through the narrow gap between the tongue and roof of the mouth makes contact with the teeth which creates intense

noise at the higher end of the acoustic spectra (4-8 kHz; Ladefoged & Maddieson, 1996). Acoustic cues to place of articulation are more recoverable because of their high intensity (Wright, 2004). However, while sibilants are exceptional, I have yet to come across a language that allows sibilants to follow sonorants in a tautosyllabic onset cluster but does not allow stops or non-sibilant fricatives.

Finally, I do not consider glottal consonants, /h/ and /ʔ/. Parker (2002) reviews claims made about the sonority of glottal consonants and finds substantial variation in how they are treated. /h/ and /ʔ/ have been treated as sonorants, obstruents, simultaneously sonorants and obstruents, as segments completely lacking sonority, and as an intermediate between obstruents and sonorants. In a (preliminary) survey of thirty languages, Miller (2012) found that glottal continuants, /h/ and /ɦ/, patterned both with sonorants and obstruents, depending on the language. The glottal stop, /ʔ/, patterns more consistently with obstruents, though not universally (Miller, 2012). Due to this ambiguity, I do not consider sonorant + glottal or glottal + sonorant as sonority reversals.

### 5.3. Results

#### 5.3.1. Data

##### 5.3.1.1. Klamath

The obstruent and sonorant consonant inventory of Klamath is shown in Table 5-2.

*Table 5-2. Klamath consonant inventory.*

	Bilabial	Alveolar	Alveo-Palatal	Palatal	Velar	Uvular	Glottal
Stops	p p <sup>h</sup> p'	t t <sup>h</sup> t'			k k <sup>h</sup> k'	q q <sup>h</sup> q'	ʔ
Affricates			tʃ tʃ <sup>h</sup> tʃ'				
Fricatives		s					h
Nasals	m m'	n ɲ n'					
Liquids		l ɭ l'					
Glides				j j'	w w'		

Klamath has a large number of consonant phonemes with sixteen obstruents, thirteen sonorants, and two laryngeals: /h/ and /ʔ/. Of the sixteen obstruents, only one is a fricative: a voiceless alveolar sibilant /s/. Klamath has four places of articulation for stops: bilabial, alveolar, velar, and uvular. It has one place of articulation for affricates: alveopalatal. With the exception of the fricative, all of the obstruents have a three-way glottal distinction: voiceless, aspirated, and ejective (Blevins, 1993).

The large number of sonorants is due to a three-way laryngeal contrast. The sonorants, with exception of /m/, have a voiced, voiceless, and glottalized version. /m/ only has a voiced and glottalized pair. The nasals have two places of articulation: bilabial and alveolar. The glides also have two: velar and palatal. The three lateral phonemes have an alveolar articulation (Blevins, 1993).

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-3. Further examples can be found in 7.3.

**Table 5-3. Klamath sonorant + obstruent onsets.** (Barker, 1963)

	C1	Sonorants		
C2		Laterals	Nasals	Glides
Obstruents	Stops	<i>lpa</i> - ‘plant’ <i>lp<sup>h</sup>eʔ</i> - ‘roll in snow’ <i>ltok<sup>h</sup>w</i> - ‘to have a crush on’ <i>lt<sup>h</sup>ikʔ</i> - ‘lope pace’ <i>lt’oq’</i> - ‘thump with finger and thumb’ <i>lki</i> - ‘motion toward for a purpose’ <i>lk’om</i> - ‘charcoal’ <i>lqaw’a:w’al</i> ‘finger’ <i>lq’an</i> - ‘ripples’	<i>mpaw</i> ‘hoot’ <i>mp<sup>h</sup>ak’</i> ‘belch’ <i>mp’aq</i> ‘have spots’ <i>ntalk</i> ‘root’ <i>nt<sup>h</sup>iq’</i> ‘drip’ <i>nt’ak’</i> ‘to be stuck’ <i>nkat’</i> - ‘jump’ <i>nk<sup>h</sup>a</i> - ‘stomach’ <i>nqijf’</i> - ‘be tight’ <i>nq’ot’</i> - <i>scorch</i> <i>nq’aq</i> - ‘crown of head’	<i>wpe</i> ‘fringe’ <i>wp<sup>h</sup>up’a</i> ‘hits with a long instrument’ <i>wtu:t<sup>h</sup>k</i> ‘caterpillar’ <i>wt’am</i> ‘put a lid on’ <i>wk<sup>h</sup>ek’a</i> ‘strikes with long instrument’ <i>wk’al’a</i> ‘cut off with long instrument’ <i>wq<sup>h</sup>am</i> ‘plant’ <i>wq’a</i> ‘quartz’
	Affricates	<i>lf’wj</i> - ‘right up to’ <i>lf’i</i> - ‘project in a line’	<i>mf<sup>h</sup>n</i> - ‘to get’ <i>nf<sup>h</sup>eqi</i> - ‘become exasperated’ <i>nf’et</i> ‘inner bark’	<i>wf<sup>h</sup>uq’a</i> ‘washes’ <i>wf<sup>h</sup>aq</i> ‘reed’
	Fricatives		<i>msa</i> - ‘prairie dog’	<i>wsu</i> ‘chest’

As seen in Table 5-3, all sonorant manners of articulation appear as the first consonant (C<sub>1</sub>) and all obstruent manners appear in (C<sub>2</sub>). Of the three laryngeal settings for sonorants, only the voiced sonorants appear in sonorant + obstruent onsets. All laryngeal settings for obstruents appear in C<sub>2</sub>: a voiceless stop is in *nkat*'- 'jump', an aspirated stop is in *l<sup>h</sup>ikʔ*- 'lope pace', and an ejective appears in *wq'a* 'quartz'. Furthermore, each possible obstruent place of articulation is found in C<sub>2</sub>.

All attested initial lateral + obstruent sequences are shown in Table 5-3. The phoneme /l/ has a relatively broad distribution in C<sub>1</sub>. It occurs before all five points of articulation and all three glottal configurations in the Klamath obstruent inventory. /l/ does not occur before the single Klamath fricative, /s/. Lateral + obstruent onset sequences are less frequent than obstruent + lateral onset sequences. In Barker's (1963) dictionary, of the 66 morphemes which have both an obstruent and lateral in the onset, the obstruent precedes the lateral in 49 morphemes and follows it in 17 cases. /l/ has a wider distribution as the second member of a consonant cluster as it is found following every obstruent except for /p'/ and /k/.

All attested initial nasal + obstruent sequences are shown in Table 5-3. The nasals /m/ and /n/ have complementary distribution in nasal + obstruent onset sequences. /m/ appears before bilabial stops and /s/. /n/ occurs before every other obstruent. /m/ and /n/ are in complementary distribution in complex onsets, suggesting that in this environment they are allophones. Ohala (1990) found that when nasal + stop sequences come between two vowels, listeners perceive the nasal as having the same place cues as the stop, e.g., [anpa] is perceived as [ampa]. Even when the nasal is preceded by a vowel, listeners attend to the place cues in the stop-vowel transition rather than the vowel-nasal transition. In the Klamath data, there is no pre-nasal vowel, so the most robust place cues come from the obstruent release cues. Since [m], a bilabial segment, occurs before bilabial stops and [n] occurs before the other stops, the underlying

representation is likely to be an /n/. #NO clusters with an /s/ behave differently, with *msa*-‘prairie dog’, but I found only one example of a nasal + /s/ onset.

A subset of attested initial glide + obstruent sequences are shown in Table 5-3. The glide /w/ appears before all sixteen obstruents in initial position. The voiced, non-glottalized /w/ is the only glide that occurs before an obstruent in the onset. /j/ is absent in sonorant + obstruent clusters.

### 5.3.1.2. Georgian

The obstruent and sonorant consonant inventory of Georgian is shown in Table 5-4.

**Table 5-4. Georgian consonant inventory.**

	Bilabial	Denti-Alveolar	Alveo-Palatal	Palatal	Velar	Uvular	Glottal
Stops	b p <sup>h</sup> p’	d t <sup>h</sup> t’			g k <sup>h</sup> k’		
Affricates		ɬ ts <sup>h</sup> ts’	ɟʃ tʃ <sup>h</sup> tʃ’				
Fricatives	v	s z	ʃ ʒ		x ɣ	χ’	(h)
Nasals	m	n					
Rhotics		r					
Laterals		l					
Glides							

All Georgian stop and affricate phonemes have three laryngeal settings: voiced, voiceless-aspirated, and voiceless-ejective. Stops can be produced at the lips (bilabial), behind the teeth (denti-alveolar), and at the velum (velar). The affricates have either an alveolar or alveo-palatal articulation. There are four points of articulation for the fricatives: denti-alveolar, alveo-palatal, velar, and uvular. With the exception of the uvular fricative, the fricatives can be voiced or voiceless. The uvular fricative is glottalized. Georgian has four sonorants, a bilabial nasal, a denti-alveolar nasal, a denti-alveolar tap/trill, and a denti-alveolar lateral, and a glottal fricative /h/ (Butskhrikidze, 2002). /h/ is marginal, appearing primarily in loan words. It appears word-initially in native words, but it is often dropped (Butskhrikidze, 2002).

In addition to these 21 obstruents, four sonorants, and /h/, Georgian has a labio-velar phoneme. The sound is often represented with a /v/, but there is debate whether it functions as

a sonorant or obstruent in the Georgian system (Butskhrikidze, 2002; Ritter, 2006). When produced after an obstruent, /v/ is realized as a glide [w] or as a secondary articulation on the obstruent, e.g., [k<sup>w</sup>] (Butskhrikidze, 2002; Ritter, 2006). Butskhrikidze (2002) states that the glide realization is common after dorsal obstruents, however, she does not mention what the phonetic realization of /v/ is after a sonorant. Because of this ambiguity, I do not consider sonorant + /v/ clusters.<sup>10</sup>

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-5. Further data can be found in 7.3

**Table 5-5. Georgian sonorant + obstruent onsets.** (Klimov, 1998; Rayfield et al., 2006)

	C1	Sonorants		
C2		Rhotics	Laterals	Nasals
Obstruents	Stops	<i>rbili</i> ‘to run’ <i>rt<sup>h</sup>wa</i> ‘to spin’ <i>rt’o</i> ‘branch’ <i>rgav</i> ‘plant’ <i>rk<sup>h</sup>a</i> ‘remark’ <i>rk’a</i> ‘horn’	<i>lboba</i> ‘soak until soft’ <i>lp’eba</i> ‘rot’ <i>lt’olwilebi</i> ‘refugees’	<i>mgeli</i> ‘wolf’ <i>mteli</i> ‘whole’ <i>mt’ers</i> ‘enemy.DAT’ <i>ndoba</i> ‘faith’ <i>nt<sup>h</sup>xeva</i> ‘spilling’ <i>ngreva</i> ‘collapse’
	Affricates	<i>rɖe</i> ‘milk’ <i>rtsk<sup>h</sup>ila</i> ‘hornbeam’ <i>rts’eva</i> ‘to shake’ <i>rɖɖa</i> ‘toil’ <i>rɖeβa</i> ‘staying’		<i>mt’sɣeri</i> ‘quail’ <i>ntɖ<sup>h</sup>kleva</i> ‘shake liquid’ <i>nɖɖreva</i> ‘swaying’
	Fricatives	<i>ryveva</i> ‘to demolish’ <i>rxevit</i> ‘oscillate’ <i>rɣ’eva</i> ‘fluctuation’	<i>lxena</i> ‘joy’ <i>lyveβa</i> ‘to thaw’	<i>mze</i> ‘sun’ <i>myelvare</i> ‘state of agitation’

All attested rhotic + obstruent onset initial sequences are shown in Table 5-5. Georgian allows a wide range of consonant clusters word-initially. Rhotics, liquids, and nasals can all

<sup>10</sup> I checked the University of California at Los Angeles Phonetics Lab Archive (“The UCLA Phonetics Lab Archive,” 2007) and Forvo.com (Forvo Media SL, 2008) for examples of Georgian /rv/ onsets. To me, /v/ in the context of /rv/ sound more similar to its obstruent realization in cases like /navi/ [navi] ‘boat’ and /vask’vlavi/ [vask’vlavi] ‘star’ than its glide realization in the onset of /k<sup>h</sup>vevri/ [k<sup>h</sup>vevri] ‘wine jar’. However, the signal-to-noise ratio in all the files was low so there was no clear evidence of the fricative noise, which is usually low in labial fricatives to begin with.

appear before obstruents. Sequences of rhotics and obstruents are particularly frequent in Georgian. In monomorphemic word-initial sequences of two consonants, /r/ is the least restrictive phoneme in terms of which sounds can follow. Fourteen out of 22 Georgian obstruents are attested following /r/: /rb/, r<sup>h</sup>, r<sup>ʰ</sup>, rdʒ, rts<sup>h</sup>, rts<sup>ʰ</sup>, rdʒ, rʃ<sup>h</sup>, rg, rk<sup>h</sup>, rk<sup>ʰ</sup>, rx, ry, rχ<sup>ʰ</sup>/. The consonants that do not appear after /r/ in a consonant cluster do not seem to have much in common, so their non-presence in /r/ clusters may be a historical accident. Georgian does not have /rp<sup>h</sup>/ or /rp<sup>ʰ</sup>/ onsets, but does have /rb/, /r<sup>h</sup>/, /rk<sup>h</sup>/, /r<sup>ʰ</sup>/, and /rk<sup>ʰ</sup>/ so there is no restriction on /r/ + bilabials, /r/ + voiceless, or /r/ + ejective sequences. The only natural class restriction may be with sibilant fricatives as /s/, /z/, /ʃ/ and /ʒ/ never follow /r/. However, the rhotic + sibilant restriction seems to have less to do with the rhotic and more to do with the sibilant fricatives. Georgian sibilant fricatives do not occur after other coronal sounds. /l/ never co-occurs with sibilant fricatives either (“Lateral” column in Table 5-5; all attested initial lateral + obstruents combinations are shown).

Laterals never occur in C<sub>1</sub> with affricates in C<sub>2</sub>. Voiced stops and voiceless ejectives appear after /l/, but voiceless aspirated stops do not. It is unclear if these gaps are systematic since lateral + obstruents onsets are less common than rhotic + obstruent onsets. There are only five /lb/ onsets in Rayfield et al.'s (2006) dictionary, with only two independent lexemes. The reverse of the cluster, /bl/, has 69 entries, with roughly 43 independent words.

The denti-alveolar /n/ is even rarer in sonorant + obstruent onset clusters. Clusters of /n/ + obstruent occur in fewer than ten words. All attested /n/ + obstruent initial combinations are shown in Table 5-5. I found only one /n/ + obstruent onset in which the obstruent did not have an alveolar articulation: *ngreva* ‘collapse’.

Sequences of the nasal /m/ with a following obstruent are common in Georgian, as shown in the subset of /m/ + obstruent onset combinations in Table 5-5: “Nasal”. As far as I could find, /m/ could precede all obstruents in initial sonorant + obstruent sequences. Many of these

clusters are heteromorphic. Butskhrikidze (2015) states that consonant clusters with two bilabial segments are not permitted within the same morpheme. The sound /m/ is used in prefixes and confixes to mark both nouns and verbs, demonstrated in (5-2).

**5-2. /m/ affix**

- a. First person singular object prefix: *m-xat'-av* ‘you paint me’
- b. First person singular subject prefix: *m-fi-a* ‘I am hungry’
- c. Adjectivizer confix: *m-k'vax-e* ‘unripe’
- d. Agent marker confix: *m-tsodn-e* ‘expert’ (Butskhrikidze, 2015)

Georgian seems to have monomorphemic /m/ + obstruent onset clusters in words like *mɟ'adi* ‘cornbread’ and *mgeli* ‘wolf’. However, Butskhrikidze (2002) argues that all /m/ + consonant onsets are heteromorphic, though in the case of *mgeli* ‘wolf’ or *mt'red-i* ‘pigeon’ the morphological origin is not apparent.

**5.3.1.3. Piro**

The obstruent and sonorant consonant inventory of Piro is shown in Table 5-6.

**Table 5-6. Piro consonant inventory.**

	Bilabial	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
Stops	p	t			k	
Affricates		ts	tʃ	te		
Fricatives		s	ʃ	ç		ħ
Nasals	m	n				
Rhotics		r				
Laterals		(l)				
Glides				j	w	

As shown in Table 5-6, Piro has nine obstruent phonemes all of which are unaspirated and voiceless. It has bilabial, alveolar, and velar places of articulation for stops and alveolar, post-alveolar, and palatal contrast for fricatives and affricates. The Piro sonorant inventory is comprised of a rhotic, a lateral, two nasals, and two glides: /l/, /r/, /m/, /n/, /j/, and /w/. Finally, it has a glottal fricative (Matteson, 1965). /l/ only occurs in loanwords.

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-7. Further data can be found in 7.3.

**Table 5-7. Piro sonorant + obstruent onsets.** (Matteson, 1963; Nies, 1986)

	C1	Sonorants	
C2		Nasals	Glides
Obstruents	Stops	<i>mpiklehotu</i> ‘courage’ <i>mtiri</i> ‘infant’ <i>mka</i> ‘clothes’ <i>npika</i> ‘I am afraid’ <i>ntoçe</i> ‘cassabanana’ <i>nkaſit-ſa</i> ‘I caught’	<i>wpowratlu</i> ‘we clean it’ <i>wtiplata</i> ‘we sit down’
	Affricates	<i>mfira</i> ‘spider monkey’ <i>mtserkakati</i> ‘animals’ <i>ntspatate</i> ‘my guava’ <i>nteiwa</i> ‘black cricket’	<i>wtseriwna</i> ‘we grow’ <i>wſkotute</i> ‘our monkey’ <i>wteirika</i> ‘we kindle’
	Fricatives	<i>msapatneti</i> ‘barefoot’ <i>mfiri</i> ‘hair comb’ <i>mçihati</i> ‘clean’ <i>nso</i> ‘genipa’ <i>nſinikani</i> ‘almost’	<i>wsalwata</i> ‘we walk around’

From the data shown in Table 5-7, only nasals and glides are found in Piro #RO initial sequences. Both *w-* and *n-* are first person pronomial prefixes which can attach to nouns or verbs. As seen in the glosses from the “Glides” column, *w-* is a 3<sup>rd</sup> person plural marker. As seen in *npika* ‘I am afraid’, *nkaſit-ſa* ‘I caught’, and *ntspatate* ‘my guava’, *n-* marks 1<sup>st</sup> person singular.

All possible nasal + obstruent onsets are shown in Table 5-7. /m/ is the most common sonorant in tauto-morphemic sonorant + obstruent onset sequences. /m/ occurs before stops, affricates, and fricatives. With the exception of /tɛ/, /m/ occurs before all obstruents in the dataset. /n/ occurs in initial tauto-morphemic sonorant + obstruent sequences where the obstruent has an alveolar or postalveolar articulation: /t, tɛ, s, ʃ/.

All of the glide + obstruent onsets in Matteson (1963) and Nies (1986) are shown in Table 5-7. Matteson states that more glide + obstruent onsets are possible (1963, p. 45), however the clusters in Table 5-7 are the only attested combinations in either dataset. Nies’s (1986) data

provides mostly tauto-morphemic onsets and Matteson does not include her exhaustive dataset. Because *w-* is the 3<sup>rd</sup> person nominal marker it is likely that more combinations are possible than those found in either dataset. Matteson (1963) does not mention any restrictions on /w/ affixing to a consonant-initial noun. Another possibility for the difference between the list of possible clusters in Matteson (1963) and those that I found, is that I limited myself to word-initial cluster combinations. More combinations may exist word-internally. Matteson (1963) states that Piro syllables do not have codas so glide + obstruent sequences would be analyzed as onsets for her.

Matteson (1963) and Lin (1998) both state that Piro permits a wide range of initial clusters and those clusters are not subject to sonority restrictions. Both suggest some restrictions on heterorganic clusters, such as no fricative-fricative or affricate-affricate clusters. But even these restrictions only affect clusters in which both place and manner are similar, as Piro lacks /rl/, /nl/, /tʃ/, /tʃs/, and other similar clusters. However, I could find no attested example of /r/ + obstruent, /l/ + obstruent, or /j/ + obstruent in any of the work on Piro. Overall, /r/ and /l/ are somewhat rare in Piro to begin with. Matteson (1963) claims that /l/ primarily occurs in loan words and Matteson (1972) reconstructs the ancestor of Piro, Proto Piro-Apuriná, without a lateral. /r/ has a slightly wider distribution than /l/, because it is found in the masculine 3<sup>rd</sup> person singular prefix for class 1 nouns<sup>11</sup> and verbs, *r-*. Otherwise it is also found primarily in loan words (Matteson, 1963). /j/ + obstruent sequences occur word-internally, as in *rijkota* ‘he counsels’, but not in word-initial position.

#### 5.3.1.4. Panixtlahuaca Chatino

The obstruent and sonorant consonant inventory of Panixtlahuaca Chatino is shown in Table 5-8.

---

<sup>11</sup> Piro has three noun classes which take different pronominal prefixes. Class 1 nouns are those whose stem begins with *h-*. The prefix replaces *h-* (Hanson, 2010).

**Table 5-8. Panixtlahuaca Chatino consonant inventory.**

	Bilabial	Alveolar	Alveo-palatal	Palatal	Velar	Labio-velar	Glottal
Stops	(p)	t d	tʃ dʃ		k g	kʷ gʷ	ʔ
Affricates		ts dz	tʃ dʒ				
Fricatives		s	ʃ	hʲ		hʷ	h
Nasals	(m)	n	nʲ				
Rhotics		r					
Laterals		l	lʲ				
Glides				j		w	

Pride & Pride (2010) describe Panixtlahuaca Chatino as having 15 obstruents, 8 sonorants, and four glottal consonants: /h/, /hʲ/, /hʷ/, and /ʔ/. /p/ and /m/ are included in the inventory, though they are marginal. Apart from /r/, Chatino contrasts alveolar phonemes with alveo-palatal forms. The lack of an alveo-palatal rhotic is likely because there was no rhotic in Proto-Chatino; /r/ only came into the languages via contact with Spanish (Campbell, 2013; Rensch, 1966; Upson & Longacre, 1965). There is a contrast between plain velars, /k, g/ and labio-velars /kʷ, gʷ/.

Within the Eastern Chatino languages there exists subtle variations in the consonant inventories. Most of these differences seem to relate to the amount of Spanish influence. Panixtlahuaca Chatino and San Juan Quiahije Chatino have a voiced-voiceless contrast for all non-bilabial stops and affricates (Cruz, 2011; Pride & Pride, 2010). Teotepec Chatino lacks voiced affricates, velars, and labio-velars (McIntosh, 2015). Zacatepec Chatino, on the other hand, is described as only having marginal /d/ and /b/, otherwise it lacks voiced obstruents (Villard, 2015).

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-9. Further data can be found in 7.3.

**Table 5-9. Panixtlahuaca Chatino sonorant + obstruent onsets.** (Pride & Pride, 2010)<sup>12</sup>

	C1	Sonorants			
C2		Rhotics	Laterals	Nasals	Glides
Obstruents	Stops	<i>rta</i> ‘sweat’ <i>rkunʔ</i> ‘touch.PRET’ <i>rk<sup>w</sup>a</i> ‘apply.COMP’	<i>lti</i> ‘thin’ <i>lku</i> ‘food’ <i>lk<sup>w</sup>i</i> ‘to boil.HAB’	<i>mta</i> ‘seed’ <i>mt<sup>w</sup>i</i> ‘dry.PRET’ <i>md<sup>w</sup>ii</i> ‘cigarette’ <i>ngaʔa</i> ‘green’ <i>nk<sup>w</sup>i</i> ‘boil.COMP’	<i>w<sup>w</sup>i</i> ‘dry’ <i>jta</i> ‘small ditch’ <i>jka ke</i> ‘horn’ <i>jk<sup>w</sup>a fla</i> ‘chocolate atole’
	Affricates		<i>ltsaʔ</i> ‘wet’ <i>ltsuʔ</i> ‘sting’	<i>m<sup>w</sup>fanʔ</i> ‘hairy’ <i>m<sup>w</sup>sanʔ</i> ‘fringe’ <i>n<sup>w</sup>faʔ</i> ‘to get wet.HAB’	<i>w<sup>w</sup>se</i> ‘algae’ <i>w<sup>w</sup>fi</i> ‘lion’ <i>jtsaʔ</i> ‘advise.COMP’
	Fricatives	<i>rsun</i> ‘reason’ (loan)	<i>lsuʔ</i> ‘pompadour’	<i>mska</i> ‘tear.COMP’ <i>n<sup>w</sup>fen</i> ‘anis’	<i>w<sup>w</sup>faʔa</i> ‘floor’ <i>wsin</i> ‘beard’ <i>jsiin</i> ‘sand’

All attested rhotic + obstruent initial sequence combinations are shown in Table 5-9. The rhotic /r/ never appears before a voiced obstruent. /r/ appears before /t/ in a few nominal forms and before /s/ in at least one Spanish loan, /rsun/ from Spanish *razón* [razon] ‘reason’. More common is /r/ before /k/ and /k<sup>w</sup>/ in past tense verbs. In these words, /r/ functions as a completive aspect marker, but only for verbs whose root is of the form *tuku-* or *tuk<sup>w</sup>-*, as shown in (5-3).

**5-3. *tuk<sup>w</sup>-* > *rk<sup>w</sup>-*** (Pride & Pride, 2010)

	Base	Completive	gloss
a.	<i>tuk<sup>w</sup>a</i>	<i>rk<sup>w</sup>a</i>	apply
b.	<i>tukunʔ</i>	<i>rkunʔ</i>	close
c.	<i>tuk<sup>w</sup>i</i>	<i>rk<sup>w</sup>i</i>	run to another

Panixtlahuaca Chatino seems to be one of the few varieties which has rhotic + obstruent clusters. Zacatepec Chatino only has rhotics in Spanish loanwords. Some of the loans have a rhotic + vowel + obstruent sequence where Panixtlahuaca Chatino has a rhotic + obstruent

<sup>12</sup> Unless noted, tone information is not included. Tone information is kept in the full dataset in 7.3

sequence, as in ZAC *kurusi*: ~ PAN *krsi* which is borrowed from Spanish *cruz* [krus] ‘cross’. Zacatepec does have a couple of loans which have rhotic + obstruent clusters, though the rhotic is devoiced: [rkale] ‘mayor’ from Spanish *alcalde* [alkalde] and [rsũ:] ‘reason’ from Spanish *razón* [rason] (Villard, 2015). For all Chatino languages the presence of a rhotic is primarily due to borrowing from Spanish (Cruz, 2011; McIntosh, 2015; Rasch, 2002; Villard, 2015). Rhotics are rare enough outside of Spanish loans that Upson & Longacre (1965) did not reconstruct any rhotic for Proto-Chatino. In native Chatino words, San Juan Quiahije Chatino has an /r/ that functions as a locative marker (Cruz, 2011). Similarly, in Teotepec Chatino, the only native word with /r/ is the word *re* which means “here/this” (Rasch, 2002).

Like /r/, the sonorants /m/, /n/, /l/, and /j/ can function as aspectual markers on verbs which leads to sonorant + obstruent sequences. *l-* functions as a habitual marker, *j-* and *m-* as completive, and *n-* has multiple functions. Unlike /r/, the other sonorants typically attach directly to the stem: *k<sup>w</sup>an* ‘to be blessed’ > *lk<sup>w</sup>an* ‘to be blessed.HAB’; *nkwan* ‘to be blessed.COMP’. Sonorants + obstruent sequences occur in monomorphemic words as well.

All attested lateral + obstruent initial sequence consonants are shown in Table 5-9. Laterals do not occur before voiced obstruents or obstruents with a palatal articulation.

All attested glide + obstruent initial sequence consonants are shown in Table 5-9. Glides do not appear before voiced obstruents. There are restrictions on glide + obstruent clusters which are homorganic for place features. The palatal glide, /j/, doesn’t appear before obstruents which have a palatal place of articulation. Similarly, the bilabial /w/ doesn’t appear before labial or labialized obstruents.

A sample of the possible nasal + obstruent combinations are shown in Table 5-9. Few, if any, restrictions seem to exist for /m/, and /n/. I found no native /mɟ/ word-initial sequence but Chatino does have *mɟin* ‘monkey’ which derives from the Spanish *machín* ‘monkey’.

Sonorant + obstruent onset sequences exist in many other Chatino languages, but the segments involved differ from Panixtlahuaca Chatino. Yaitepec Chatino has a glide + obstruent onset in *jk<sup>wa</sup>* ‘flat’, whereas Panixtlahuaca Chatino has a lateral, *lk<sup>wa</sup>* ‘flat’. The lateral + obstruent onset *lkaʔ* ‘potato beetle’ in Panixtlahuaca Chatino is cognate with nasal + obstruent in San Juan Quiahije Chatino, *nkaʔ* (Campbell, 2013; Cruz, 2011). In fact, for some Chatino varieties, /l/ and /n/ are only minimally contrastive. Rasch (2002) argues that while there are good reasons to keep them as separate phonemes, the two are in near complementary distribution in Yaitepec Chatino.

### 5.3.1.5. Russian

The obstruent and sonorant consonant inventory of Russian is shown in Table 5-10.

**Table 5-10. Russian consonant inventory.**

	Bilabial		Labiodental		Denti-Alveolar		Palatal-Alveolar	Palatal	Velar
Stops	b p	b <sup>j</sup> p <sup>j</sup>			ɖ ʈ	ɖ <sup>j</sup> ʈ <sup>j</sup>			g k
Nasals	m	m <sup>j</sup>			ɳ	ɳ <sup>j</sup>			
Fricatives			v f	v <sup>j</sup> f <sup>j</sup>	ʐ ʂ	ʐ <sup>j</sup> ʂ <sup>j</sup>	ʃ ʃ̟		x
Affricates					ʧ ʧ̟		tʃ		
Rhotics					ɹ	ɹ <sup>j</sup>			
Laterals					ɭ	ɭ <sup>j</sup>			
Glides								j	

Russian has a large consonant inventory, including palatalized and non-palatalized pairs for all but the velars /g, k/ and palatal /j/. The stop series has bilabial, denti-alveolar, and velar primary articulations which can either be voiced or voiceless. The bilabials and denti-alveolars have palatalized and non-palatalized versions. Fricatives have a four-way contrast for primary place of articulation: labio-velar, denti-alveolar, post-alveolar, and velar. The velar fricative does not have a voiced-voiceless distinction. Russian has two voiceless affricates: /tʂ/, and /tʃ̟/. The sonorants /m/, /n/, /ɹ/, and /l/ also have palatalized and non-palatalized phonemes. There is one glide: /j/. /v/, as in Georgian, can be analyzed as obstruent or sonorant (Lulich, 2004).

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-11. Further data can be found in 7.3.

**Table 5-11. Russian sonorant + obstruent onsets.** (Russian National Corpus, 2003; Scheer, 2007)

	C1	Sonorants		
C2		Rhotics	Laterals	Nasals
Obstruents	Stops	<i>rdeŋ</i> ‘to glow’ <i>rtuŋ</i> ‘mercury’ <i>rta</i> ‘mouth.GEN’	<i>lba</i> ‘forehead.GEN’ <i>l̥da</i> ‘ice’ <i>lgaŋ</i> ‘to lie’ <i>l̥gota</i> ‘respice’	<i>mgl̥a</i> ‘haze’ <i>mknuŋ</i> ‘sudden, shocking movement’
	Fricatives	<i>rʒi</i> ‘rye’	<i>lʒi</i> ‘lie.GEN’ <i>l̥stiŋ</i> ‘flatter’ <i>l̥zja</i> ‘it is suitable to’	<i>mzda</i> ‘bribe’ <i>mstiŋ</i> ‘revenge’ <i>mʃit̥</i> ‘cover with moss’

All attested rhotic + obstruent onsets are shown in Table 5-11. Only the non-palatalized rhotic appears sonorant + obstruent initial sequences. /r/ appears before /t, d, ʒ/.

All attested lateral + obstruent initial sequences are shown in Table 5-11. Both the palatalized and non-palatalized laterals appear in #RO clusters. Laterals appear before all primary places of articulation for the Russian stop series but do not appear before palatalized or voiceless stops. The laterals occur before sibilants. Only one member of lateral + sibilant onsets seems to be able to have a palatal articulation. The palatalized /li/ occurs with the alveolar sibilants while the non-palatalized /l/ appears with the palatal-alveolar /ʒ/. For lateral + obstruent onsets, Russian seems to only permit one member to have a palatal articulation. The restriction holds in the other sonorant + obstruent onsets, but the number of Russian sonorant + obstruent onset clusters is small, so the pattern may be coincidental.

All attested nasal + obstruent onsets are shown in Table 5-11. The bilabial nasal appears with velar stops and sibilants. Russian has an alveolar nasal, it never seems to appear before obstruents in syllable initial position.

#RO clusters are somewhat rare in the lexicon. According to Proctor (2009), onset clusters with a liquid followed by an obstruent make up only 1.06% of all tautosyllabic clusters with

liquids. Because of their overall rarity, gaps in possible sonorant + obstruent onsets could simply be a historical accident rather than phonotactic restrictions.

### 5.3.1.6. Tsou

The obstruent and sonorant consonant inventory of Tsou is shown in Table 5-12.

*Table 5-12. Tsou consonant inventory.*

	Labial	Alveolar	Palatal	Velar	Glottal
Stops	p ɸ	t ɖ		k	ʔ
Affricates		ts			
Fricatives	f v	s z			h
Nasals	m	n		ŋ	
Rhotics		ɻ			
Glides			j <sup>13</sup>		

Tsou has three voiceless stops: /p t k/. It also has two implosives: /ɸ/ and /ɖ/. The implosives are somewhat marginal, though (Wright, 1994, p. 67). Tsou has four (non-glottal) fricatives, /f v s z/, and one affricate: /ts/. The sonorant inventory varies by dialect. All dialects have three nasals: a bilabial, alveolar, and velar. Only the Duhtu dialect has a rhotic, however. In the Tfuea dialect the rhotic has become a glide, /j/ (Tsuchida, 1972). There is disagreement about the status of glides in Tsou. Tsuchida (1972) does not include a glide in his initial inventory. Wright (1996) includes both /w/ and /j/. Zeitman (2005) includes /j/ but not /w/ because /w/ never occurs in initial position. Chen (2002) claims that /ɻ/ has become [e] rather than a glide in Tfuea.

A representative sample of Sonority Sequencing Generalization violating clusters are shown in Table 5-13. Further data can be found in 7.3.

---

<sup>13</sup> In the Tfuea dialect of Tsou, the rhotic has become /j/ (Tsuchida, 1972). The literature on Tfuea seems to disagree on the phonemic status of /j/.

**Table 5-13. Tsou sonorant + obstruent onsets.** (Blust & Trussel, 2013; Wright, 1996)

	C1	Sonorants	
C2		Rhotics	Nasals
Obstruents	Stops	<i>ɬuu</i> ‘loquat tree’	<i>mdɪŋi</i> ‘beautiful bell-like tone’ <i>mpitvihi</i> ‘seventy’ <i>mtoki</i> ‘to throw’ <i>mkameosi</i> ‘quick recovery of health’
	Affricates		<i>mtsoo</i> ‘eye’ <i>mtsoi</i> ‘die’
	Fricatives	<i>ɬvore</i> ‘flying squirrel’ <i>ɬsee</i> ‘tears’	<i>mfeɪʔsi</i> ‘to cover’ <i>mvore</i> ‘dry in the sun’ <i>msadi</i> ‘step on’ <i>ɲfuju</i> ‘antler’ <i>nsoo</i> ‘pond’

All attested rhotic + obstruent initial sequences are shown in Table 5-13. In the Duhtu dialect, the retroflex approximant rhotic /ɬ/ occurs before the voiced bilabial fricative /v/, the voiceless alveolar stop /t/, and the voiceless alveolar fricative /s/. No clear restriction exists due to the small number of words with rhotic + obstruent initial clusters.

All possible nasal + obstruent initial sequences are shown in Table 5-13. /n/ occurs before /f/, /s/, and /t/ while /ŋ/ occurs before /v/, /d/ and /t/.

### 5.3.1.7. Analysis of #RO clusters

All the sampled languages adhere to Greenberg's (1965) implicational statement that a language with a sonorant + obstruent onsets will have an obstruent + obstruent onsets, and a language with obstruent + obstruent onsets will have obstruent + sonorant onsets. If a language has a sonorant + obstruent cluster, it will also usually allow sonority plateaus, a tendency other typological surveys of onsets have found (Kreitman, 2008; Morelli, 1999).

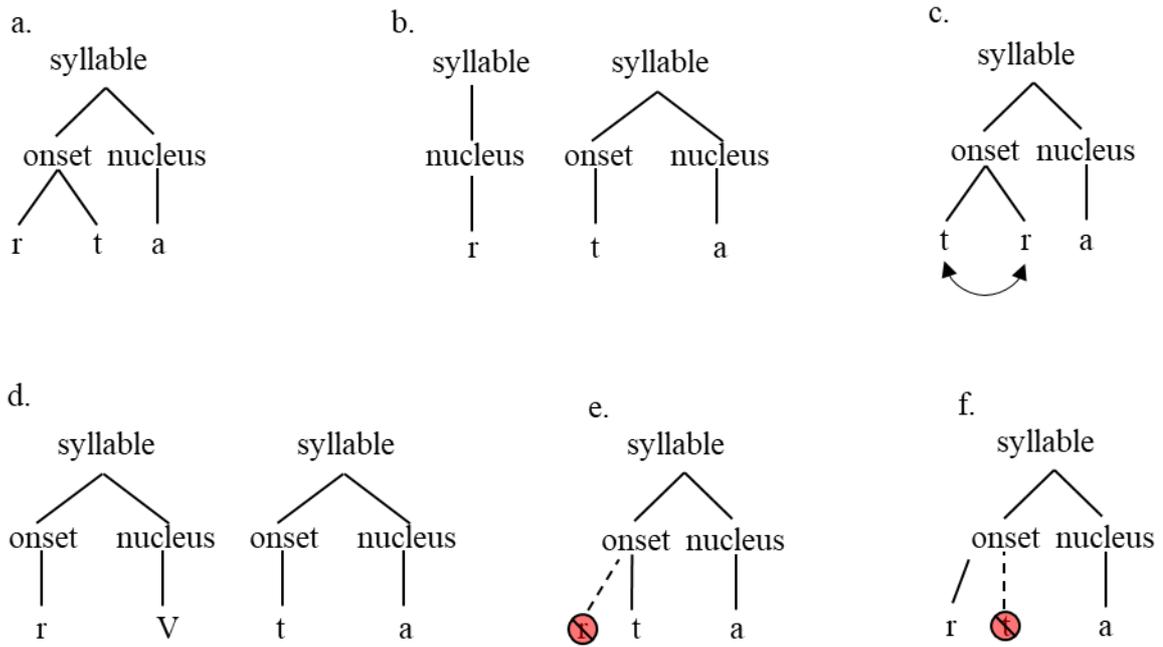
**Table 5-14. Implicational relationships.**

Language	Sonorant + Obstruent >>	Obstruent + Obstruent >>	Obstruent + Sonorant
Tsou	<i>ʎvo.ɬe</i> ‘flying squirrel’	<i>fkore</i> ‘a type of plant’	<i>tɬesi</i> ‘rope’
Russian	<i>lbom</i> ‘forehead.gen’	<i>kto</i> ‘who’	<i>dni</i> ‘days’
Georgian	<i>rbena</i> ‘to run’	<i>tbili</i> ‘warm’	<i>tla</i> ‘peel’
Piro (Yine)	<i>mtasa</i> ‘hollow’	<i>tpa</i> ‘curve’	<i>kna</i> ‘pole’
Panixtlahuaca Chatino	<i>yka</i> ‘tree’	<i>tka</i> ‘just; barely’	<i>tʃaʔ</i> ‘mouse’
Klamath	<i>nkalq</i> ‘mud’	<i>ktena</i> ‘spear’	<i>q’li:pa</i> ‘cricket’

Overall, the sonorant + obstruent onsets do not provide any evidence of systematic gaps. Sonorant + obstruent onsets are not only rare cross-linguistically, but for some of the languages in the dataset the onsets are rare language internally. Therefore, it is premature to claim that there are true gaps and/or tendencies. Many sonorant + obstruent onsets that do not occur may be attributable due to likely historical accident.

### 5.3.2. Tautosyllabicity

As I am arguing these languages are exceptions to the Sonority Sequencing Generalization, it is necessary to demonstrate that the clusters in question are, in fact, tautosyllabic and that they make up complex onsets. A framework which considers the Sonority Sequencing Generalization universal, inviolable component of grammar may have difficulty accounting for tautosyllabic clusters like that in Table 5-14.



**Figure 5-2. Sonorant + obstruent onset repairs.**

The illicit form could be repaired in several ways, all which would be in accordance with the Sonority Sequencing Generalization. The sonorant could be syllabified on its own, creating a syllabic sonorant, as in Figure 5-2b. The two consonants could undergo metathesis, yielding a single sonority rise in to the nuclear vowel as in Figure 5-2c. A vowel could be inserted, like in Figure 5-2d, creating two syllables. Or, either the sonorant or the obstruent could be deleted, as in Figure 5-2e-f, leaving a unidirectional sonority rise. In the following sections, evidence for syllabification supports #RO clusters are tautosyllabic onsets as in Figure 5-2a.

### 5.3.2.1. Klamath

Barker (1964) states that nasals and liquids are syllabic when they occur before a consonant at the beginning of a word. However, he also states that the sonorants “do not affect the stress pattern and thus ought not to be classified as vowel phones” (Barker, 1964, p. 26). It is not clear what is meant by “syllabic” in this context. Since sonorants do not affect the stress pattern, it would seem that they do not function as phonological syllables.

Perhaps the best evidence for treating word-initial sonorant + obstruent onset sequences as tautosyllabic comes from morphological reduplication. Reduplication involves the repetition

and affixation of a prosodic domain and syllables are a common target. In Klamath, the formation of the distributive involves the reduplication of the onset and nucleus of a base's initial syllable (Barker, 1963, 1964; Blevins, 1993). In (5-4a), the onset and nucleus of *q<sup>h</sup>lin* is /q<sup>h</sup>li/, this is then attached to the base to create *q<sup>h</sup>liqlən* [DIST *q<sup>h</sup>li*-[BASE *q<sup>h</sup>lən*]]. Vowel loss (5-4b & d) or reduction (5-4a & c) in the base is a secondary effect of this morphological process.

**5-4. Klamath reduplicative distributive prefix.**

	non-distributive	distributive	gloss
a.	<i>q<sup>h</sup>lin</i>	<u><i>q<sup>h</sup>li</i></u> . <i>q<sup>h</sup>lən</i>	choke
b.	<i>p<sup>h</sup>a.ka</i>	<u><i>p<sup>h</sup>a</i></u> . <i>pka</i>	bark
c.	<i>pon.wa</i>	<u><i>pə</i></u> . <i>pan.wa</i>	drink
d.	<i>la.ʃ'a</i>	<u><i>la</i></u> . <i>ʃ'a</i>	build a house

When a nasal or lateral is in initial position (5-5a:d), the morphological operation targets the CCV sequences the same as in non-sonority reversal sequences in (5-4a).

**5-5. Klamath reduplicative distributive prefix.**

	non-distributive	distributive	gloss
a.	<i>lt<sup>h</sup>ikʔ</i>	<u><i>lt<sup>h</sup>i</i></u> . <i>lt<sup>h</sup>ak<sup>h</sup>.tʃ<sup>h</sup>n'a</i>	go on tiptoes
b.	<i>lt<sup>h</sup>u:q'</i>	<u><i>lt<sup>h</sup>u</i></u> . <i>lt<sup>h</sup>u:q'at<sup>h</sup>k<sup>h</sup></i>	spotted
c.	<i>nq<sup>h</sup>ut'</i>	<u><i>nq<sup>h</sup>u</i></u> . <i>nq<sup>h</sup>t'a</i>	scorch
d.	<i>msa</i>	<u><i>msa</i></u> . <i>msk'a</i>	little prairie dogs

Reduplication does not target only the sonorant in these cases, suggesting that the initial sonorant + obstruent onsets are tautosyllabic.

**5.3.2.2. Georgian**

At first glance, the literature on Georgian sonorant + obstruent onsets seems to be divided as to whether initial sonorants in #RO sequences are syllabic. According to Nepveu (1994) and Robins & Waterson (1952), Georgian speakers don't perceive the acoustic peaks of sonorant consonants as syllable peaks. Native speaker syllable intuition only treats vowels as syllable

nuclei. Since disyllabic words receive initial stress (Butskhrikidze, 2002), the lack of stress on the sonorant in *rtwa* ‘to spin’ and *mze* ‘sun’ matches with the reported speaker intuition (Robins & Waterson 1952; Butskhrikidze 2002).

Contra Nepveu (1994) and Robins & Waterson (1952), Butskhrikidze (2002) and Cherchi (1999) refer to sonorants as syllabic. That being said, it seems that they are suggesting some sense of *phonetic* syllabicity, rather than phonological syllabicity. Butskhrikidze (2002) even uses the term “phonetically syllabic”. Cherchi (1999) states that the sonorants “sound syllabic” but also claims that they don’t count as nuclei for metrical purposes. The work of Pouplier & Beňuš (2011) demonstrate the difficulty in showing a clear acoustic difference between syllabic and non-syllabic sonorant consonants. Since the acoustic distinction between syllabic and consonantal sonorants has yet to be established, it is unclear what is meant by “sounds” syllabic. Here, I take the native speaker intuition and evidence from stress patterns as evidence that initial sonorants in #RO onsets are not syllabic, and they are prevocalic complex onsets.

### 5.3.2.3. Russian

Initial sonorant + obstruent sequences in Russian are widely accepted as being tautosyllabic based on experimental evidence, poetry, and comparisons with other Slavic languages that arguably have syllabic sonorants in the same position (Berent et al., 2007; Davidson, 2011; Kreitman, 2008).

Perhaps some of the clearest evidence for sonorant + obstruent cluster tautosyllabicity in forms like Russian *lgatʲ* ‘to lie’ and *mgla* ‘haze’ is the experimental work of Iris Berent and colleagues (Berent et al., 2007). Russian speakers provided clear-cut syllabic judgments for nonce words like *lba*.<sup>14</sup> Overwhelmingly, Russian speakers perceived *lba* as monosyllabic

---

<sup>14</sup> The purpose of the experiment was to show that English speakers had sonority-based intuitions about sonorant + obstruent onsets, despite their absence in the English lexicon. Russian speakers were used as a control.

while perceiving its epenthetic counterpart, *leba* as disyllabic. This pattern held for clusters with rhotics and nasals as well.

Unlike other languages in the dataset, the Slavic languages have well-documented and easily accessible poetic traditions. Poetry can be useful in assessing syllable counts and syllabification patterns as many poets organize the linear and prosodic structure of poems based on syllable count and syllable structure.

### ***5-6. Syllabification of Russian poem.***

*lgut-se-kra-na-ko-var-nu-je-poz-ne-ru*  
*u-ʃit-ziz-ni-nas-ksʲu-sha-sob-ʃak*  
*ras-poz-na-li-be-du-slif-kom-pozd-no-mu*  
*ʃto-ze-de-lat-nam-zit-dal-fe-kak*

Лгут с экрана коварные познеры,  
Учит жизни нас Ксюша Собчак.  
Распознали беду слишком поздно мы.  
Что же делать нам, жить дальше как?

(Степашинной, В.  
“Обращение к Русским  
Мужчинам 23-го  
Февраля”)

The poem in (5-6) by Vera Stepashina (only partially reproduced here), alternates lines of eleven and nine syllables. If *lgut* were syllabified as *l.gut*, with a syllabic lateral, the first line in (5-6) would have twelve syllables, not eleven, which would conflict with the rest of the poem. While poets often take liberties with syllable count, in this case, syllabification of *lgut* as a single syllable with a complex onset is consistent with earlier analyses.

Finally, when Russian is compared to Slavic languages that have syllabic sonorant consonants in these positions, there are clear differences. Czech, Serbian, Macedonian, and Slovenian all allow monosyllabic words, including those with syllabic sonorants. Compare forms in (5-7) where Russian cognates exhibit a corresponding vowel.

**5-7. Cognate forms in Slavic languages. (Proctor 2009)**

Slavic languages with syllabic sonorants	Russian	Gloss
a. Serbian <i>přst</i>	<i>p'erst</i>	'finger'
b. Czech <i>vřk</i>	<i>volk</i>	'wolf'
c. Macedonian <i>hřb</i>	<i>gorb</i>	'hump'
d. Slovenian <i>sřp</i>	<i>s'erp</i>	'sickle'

**5.3.2.4. Panixtlahuaca Chatino**

In Panixtlahuaca Chatino, and most other varieties of Chatino, the final syllable is the locus for suprasegmental contrasts with final syllables receiving stress and carrying contrastive tone (Campbell, 2013; Campbell & Woodbury, 2010; Pride, 1963; Pride & Pride, 2010). Words in (5-8) show no difference in #OR and #RO initial words. In both cases, stress falls on the only vowel in the word.

**5-8. Highland Chatino tone and stress.**

- a. *lk<sup>v</sup>i<sup>45</sup>* 'tepache (a type of liquor)'
- a' *da<sup>45</sup>* 'question marker'
- b. *lta<sup>21</sup>* 'armful'
- b'. *ʔa<sup>21</sup>* 'a lot'
- c. *lka<sup>23</sup>* 'be'
- c'. *bra<sup>23</sup>* 'time'
- d. *lti<sup>2</sup>* 'narrow'
- d'. *ku<sup>2</sup>* 'to eat'

At present, there seems to be little motivation for treating sonorants in initial #RO clusters as syllabic.

**5.3.2.5. Piro**

Like Georgian and Klamath, Piro sonorants have been called syllabic. Matteson (1965) claims there is a syllabic allophone, not just for sonorants, but for all consonants when they appear before another consonant in the onset of a word. Allophones are phonetic variants within phonological categories; that is, they are non-contrastive for linguistic meaning. In spite of using the term "syllabic", Matteson seems to be rejecting the syllabicity of consonants and

is describing a phonetic variation in which a vocalic element is present. Lin (1998) argues that the vocalic elements do not take part in any phonological processes or rules like word stress or phrase level rhythmic rules.

Further evidence comes from Piro syllable stress patterns. Piro syllables can have primary stress, secondary stress, tertiary stress, or they can be unstressed. The examples in (5-9), taken from Matteson (1965, p. 21), illustrate the stressed vowels in Piro words or “stress groups”.

**5-9. Piro stress patterns.** *Matteson (1965, p. 21)*

	Stress group	gloss
a.	<i>wá.lo</i>	‘rabbit’
b.	<i>ru.txí.txa</i>	‘he observes taboo’
c.	<i>ʃí.ja.há.ta</i>	‘he cries’
d.	<i>sá.lwa.je.hká.kna</i>	‘they visit each other’
e.	<i>pé.ʃi.ʃhí.ma.tló.na</i>	‘they say they stalk it’
f.	<i>rú.slu.nó.ti.nki.tká.na</i>	‘their voices already changed’
g.	<i>sá.ple.whí.ma.mtá.na.tná.ka</i>	‘they say he went along screaming again’

In Piro, the penultimate syllable receives primary stress (5-9). Any stress group over four syllables will have secondary stress on the initial syllable (5-9c:g). Stress groups over six syllables will have tertiary stress. As shown in the examples (5-9e:g), starting from the initial syllable, tertiary stress occurs on every other syllable. If the stress group has an odd number of syllables there will be two unstressed syllables before the penult, (5-9f). In (5-9f), if pre-obstruent sonorants were syllabic, the stress group should have eight syllables rather than the attested seven. The antepenultimate syllable would be *ki*, rather than *nki*, and the preceding nasal should receive stress. Similarly, in (5-9g) the stress group would have nine syllables rather than the attested eight. Furthermore, in the *mta* sequence the stress should fall on the nasal rather than the vowel. In both cases, if the nasal were the previous syllable’s coda then the syllables would adhere to the Sonority Sequencing Generalization. However, codas are otherwise unattested in Piro (Lin, 1998; Matteson, 1965).

Another example from Matteson (1965, p. 21), shown in (5-10), suggests that sonorant + obstruent onsets are tautosyllabic.

**5-10. Predicted stress patterns of syllabic consonants.**

- a. *tá.pa.lú.fka.ná.wa.thí.ma.ná.nu.mtá.na.tná.ka*  
 ó σ ó σ ó σ ó σ ó σ ó σ ó σ
- b. *tá.pa.lú.fka.ná.wa.t.hí.ma.ná.nu.m.tá.na.t.ná.ka*  
 ó σ ó σ ó σ ó σ ó σ ó σ ó σ ó σ ó σ
- c. *tá.pa.lú.fka.ná.wa.thí.ma.ná.nu.m.tá.na.tná.ka*  
 ó σ ó σ ó σ ó σ ó σ ó σ ó σ ó σ

In (5-10a), the attested stress pattern, the Sonority Sequencing Generalization violating syllable *mta* receives stress. Primary stress falls on the penultimate syllable *tna* and the initial syllable, *ta*, has secondary stress. Because tertiary stress falls on ever other syllable, for a stress group with 14 syllables, the 11<sup>th</sup> syllable, *mta*, is expected to receive tertiary stress. If consonants were syllabic, the stress pattern should follow (5-10b). Here, all clusters are broken up into syllabic consonants. Since Piro does not allow codas the consonant cannot be associated with the previous syllable. The result would be a stress group of eighteen syllables, but we would expect the syllable *wa* to receive stress and *na* to be unstressed, contra the attested form. In (5-10c), only the sonorants in sonority reversal clusters are considered syllabic, which again makes the wrong predictions. *m* would directly receive stress leaving *ta* unstressed.

Similarly, if sonorant consonants are syllabic, the attested stress patterns of words like *nsó* ‘genipa’ and *wpo.wrátlu* ‘we clean it’ would be exceptional and we should expect *\*\*ń.so* and *\*\*úpowrátlu*, respectively.

Matteson (1965, pp. 17 & 20) provides pitch contours, shown in (5-11), which further suggest that segments in these sonorant + obstruent clusters are not phonologically syllabic. Falls and rises in pitch are always associated with a full, phonological vowel, never a consonant:

**5-11. Piro tone contours.**

- a. 2 1 2 1 2 1  
txako nato prani
- b. 1 1 2 2 1 1 2 2  
kin-ʃikale, mʃir-ʃikale

**5.3.2.6. Tsou**

The Proto-Austronesian lexicon was dominated by disyllabic lexical bases and many modern languages have maintained this disyllabic preference (Blust, 2007). Modern Tsou seems to adhere to this preference too, as 90% of Tsou’s lexical bases in Blust’s comparative dictionary are disyllabic (5-12).

**5-12. Disyllabic lexical bases in Proto-Austronesian and Tsou.**

	Proto- Austronesian	Tsou	Gloss
a.	* <i>a.ki</i>	<i>ak.ʔi</i>	‘grandfather’
b.	* <i>da.ki</i>	<i>tsa.ʔi</i>	‘dirty’
c.	* <i>Ca.u</i>	<i>tso.u</i>	‘person’
d.	* <i>li.ma</i>	<i>ji.mo</i>	‘five’
e.	* <i>ba.tux</i>	<i>fa.tu</i>	‘stone’
f.	* <i>ku.ya</i>	<i>ku.zo</i>	‘bad’
h.	* <i>Ku.liC</i>	<i>ri.tsi</i>	‘tuber peelings’
i.	* <i>Si.Nuq</i>	<i>sku.u</i>	‘necklace’
j.	* <i>Su.ʔu</i>	<i>su.u</i>	‘thou’
k.	* <i>ta.naq</i>	<i>tno.o</i>	‘ <i>Aralia decaisneana</i> ’
l.	* <i>Ca.liS</i>	<i>tʃe.si</i>	‘rope’

Note that in (5-12i,j), Tsou permits vowel + vowel sequences, even when the vowel qualities are the same, like /uu/ or /oo/. Zeitman (2005) argues that stress assignment is more regular if these sequences are treated as two mora. In Tsou, stress is placed on the penultimate mora and shifts to the new penultimate mora after suffixation. For example, stress shifts from the first /o/ in *mtsó.o* ‘eyes’ to the second when the possessive *-ta* is affixed: *mtsó.ó.ta* (Zeitman, 2005, p. 262).

Many of the Tsou sonorant + obstruent sequences are historically derived from disyllabic bases, as shown in (5-13).

**5-13. Disyllabic sonorant + obstruent bases.**

	Proto-Austronesian	Tsou	Gloss
a.	*ma.Ca	<i>mtsoo</i>	‘eye’
b.	*lu.Seq	<i>ʎsee</i>	‘tears’
c.	*Ri.tu	<i>ʎtuu</i>	‘loquat tree’
d.	*la.waR	<i>ʎvore</i>	‘flying squirrel’

Where the vowel of the initial base was lost, and a form would be mono-moraic, the final vowel is lengthened, (5-13a:c), or a final vowel is added, as in (5-13d).

Clearer evidence for treating initial sonorant + obstruent clusters as belonging to a complex onset comes from Tsou reduplication. In (5-14), reduplication targets the head of the syllable, which includes the onset and nucleus, and suffixes it to the base:

**5-14. Tsou syllable reduplication.**

- a. *fɲúu* > *fɲúu-fɲu* ‘big head’
- b. *smúu* > *smúu-smu* ‘much due’
- c. *skúu* > *skúu-sku* ‘necklaces’

Sonorant + obstruent clusters in (5-15) are treated the same as the complex obstruent + sonorant onsets and the obstruent + obstruent clusters in (5-14):

**5-15. Tsou syllable reduplication in sonorant + obstruent onsets.**

- a. *mtsúu* > *mtsúu-mtsú* /\*\**mtsúu-m*/ \*\**mtsúu-mu* ‘ears of crops’
- b. *mtsóo* > *mtsóo-mtsó* /\*\**mtsóo-m*/\*\**mtsóo-mo* ‘big eye’
- c. *ɲsóo* > *ɲsóo-ɲso* /\*\**ɲsóo-ɲ*/ \*\**ɲsóo-ɲo* ‘much spring’

As there is no difference in how the clusters are treated there seems to be little justification for treating sonorants as syllabic in initial #RO sequences.

**5.3.2.7. Summary**

While initial sonorant + obstruent clusters are not common, there are several languages where there is a solid case that (i) they exist; (ii) they constitute complex onsets; and (iii) they violate the Sonority Sequencing Generalization. A phonological Sonority Sequencing

Generalization that is active in the grammars of all humans should require that all these languages correct the sonority reversal in some manner, yet none of these languages do so.

Furthermore, a phonological universal Sonority Sequencing Generalization makes no predictions about how these clusters arise or at what frequency they are expected.

### 5.3.3. The evolution of #RO- onsets

In order to understand why sonorant + obstruent onset clusters are typologically rare, we can examine the historical pathways that led to the development of these clusters. Evolutionary Phonology argues that many common sound patterns can be explained by common historical pathways (Blevins, 2004, 2017).

Potentially, tautosyllabic sonorant + obstruent onset clusters could have developed numerous ways, as shown in (5-16):

#### *5-16. Potential pathways to sonorant + obstruent onset clusters.*

---

a. Initial vowel loss:	arta	>	rta
b. Interconsonantal vowel loss:	rata	>	rta
c. Metathesis:	tra	>	rta
d. Sonorant insertion:	ta	>	rta
e. Obstruent insertion:	ra	>	rta

A vowel could have been lost in initial position or between the two consonants (5-16a,b). The loss of unaccented vowels is a common process (Taylor, 1994). Vowels that are unstressed are shorter and are associated with decreased subglottal pressure which typically devoices the vowel (Ohala, 1983). The shortness and the lack of voicing reduces the perceptibility of the segment which could lead to its eventual loss.

The two sounds could have changed order, “metathesis” (5-16c). Rhotics, laterals, and nasals have acoustic cues that can spread over multiple segments (Ohala, 1993). For instance, a lowered 3<sup>rd</sup> formant is one of the principle acoustic cues for rhotics and this cue can extend to adjacent segments, particularly vowels (Ladefoged & Maddieson, 1996, p. 244). Long duration acoustic cues can lead to perceptual confusion in cue localization which can lead to

sound metathesis (Blevins & Garrett, 1998, 2004). However, because obstruents are poor carriers of acoustic cues (Wright, 2004), #ORV > #ROV may be unexpected. Blevins & Garrett (2004, p. 129) provide an example of Armenian metathesis which could be informative. Classical Armenian \**trasu-* became *artasu-* ‘tear(s)’ in modern Armenian and *kran-* > *erkan-* ‘millstone’. The sonorant changed places with the obstruent resulting in RO, but a vowel was inserted in addition to metathesis.

Finally, a consonant could have been inserted as in (5-16d,e). The insertion of a sonorant (5-16d) or an obstruent (5-16e) seem to be the least likely in terms of natural sound changes. Laryngeal consonants are the most common insertions at prosodic boundaries and glides are the most common consonant insertions between vowels (Blevins, 2008). Lombardi (2002) cites examples of coronal insertions, though she admits these are rare. Blevins (2008) shows that in many cases where non-laryngeal or glide segments are inserted, the historical pathways are more complex and often involve multiple steps. In Chamorro, the synchronic grammar allows the insertion of the affricate [dz] between two vowels. Historically, however, the affricate developed from the glide [j] (Blevins, 2008, p. 11).

Despite all the potential pathways that could be possible, in the next section I show that the only attested pathway is interconsonantal vowel loss (5-16b).

### **5.3.3.1. Austronesian**

Proto-Austronesian is reconstructed as having mostly disyllabic lexical bases of CVCV(C). Proto-Austronesian had no syllable initial consonant clusters, so the clusters arose without structural support. In some daughter languages, unstressed vowels in #CVCV(C) were lost, creating a diversity of cluster types, including sonorant + obstruent clusters. The trajectory from Proto-Austronesian to Tsou for a range of initial clusters is shown in (5-17).

### 5-17. Proto-Austronesian > Tsou.

- a. \*baSaY > fsoi ‘a plant with a root like a sweet potato’
- b. \*Cumay > tsmoi ‘the Formosan black bear’
- c. \*baŋaS > fŋose ‘a tree: Melia azedarach’
- d. \*RamiS > ɲmisi ‘root’
- e. \*luSeq > ɲsee ‘tears’
- f. \*Ritu > ɲtuu ‘loquat tree’
- g. \*lawar > ɲvoɽe ‘flying squirrel’

In addition to morpheme-initial clusters, Proto-Austronesian languages also had #CV- prefixes. The loss of the vowel in the active voice affix *ma-* in Tsou led to sonorant + obstruent onset clusters:

### 5-18. Tsou prefix vowel loss.

- a. \*ma + \*puSaN > Tsou *m-pusku* ‘twenty’
- b. \*ma + \*aCay > Tsou *m-tsói* ‘die’

### 5.3.3.2. Indo European

Slavic sonorant + obstruent onset clusters come from what is commonly known as “the fall of the *yers*.” The *yers* were short high vowels that were lost when they did not receive primary stress (see the alterations in (5-19e) and (5-19f) below; also see Bethin, 1998; Proctor, 2009; Yearley, 1995; and references within for a more detailed discussion on *jer* loss). Like Austronesian vowel loss, the loss of *yers* created new consonant clusters in Russian, as shown in (5-19):

### 5-19. Proto-Slavic > Russian *jer* loss.

- a. \*sŭna > sna ‘sleep.gen’
- b. \*dŕne > dnja ‘day.gen’
- c. \*kŭto > kto ‘who’
- d. \*lŭgati > lga<sup>tʲ</sup> ‘lie.inf’
- e. \*lŭba > lba ‘forehead.gen’ (cf. *lŭbŭ* > *lob* ‘forehead.nom’)
- f. \*rŭta > rta ‘mouth.gen-sg’ (cf. \*rŭtŭ > *rot* ‘mouth.nom’)
- g. \*rŭtŭtŭ > rtu<sup>tʲ</sup> ‘mercury’
- h. \*mŭgla > mglā ‘haze’

Unlike Austronesian, however, consonant clusters were already widespread in Proto-Slavic with many directly inherited from Proto-Indo-European. Proto-Indo-European obstruent

+ sonorant onset clusters (e.g., \**k<sup>w</sup>rei-* ‘buy’) were widespread and the language even permitted obstruent + obstruent onsets (e.g., \**pter-* ‘wing’) and sonorant + sonorant onsets (e.g., *mreg<sup>h</sup>u-* ‘short’). Several Proto-Indo-European consonant clusters have persisted into the modern Slavic languages with minimal changes, as shown in (5-20):

**5-20. Proto-Indo-European clusters > Modern Slavic.**

- |    |     |   |   |                   |                              |   |      |             |           |
|----|-----|---|---|-------------------|------------------------------|---|------|-------------|-----------|
| a. | PIE | * <i>b<sup>h</sup>reh<sub>2</sub>tēr</i>  | > | PSI <sup>15</sup> | * <i>bratr<sub>2</sub>ja</i> | > | Rus. | <i>brat</i> | ‘brother’ |
| b. | PIE | * <i>sneig<sup>w<sup>h</sup></sup>-o-</i> | > | Rus.              | <i>sneg</i>                  |   |      |             | ‘snow’    |
| c. | PIE | * <i>k<sup>l</sup>ōwā</i>                 | > | Rus.              | <i>slava</i>                 |   |      |             | ‘fame’    |

**5.3.3.3. Kartvelian**

The historical data on Georgian is not as clear as the Austronesian or Indo-European data, yet for many #RO clusters there is a reconstructed #RVO sequence, suggesting Georgian underwent the same process of unstressed vowel loss, as shown in (5-21):

**5-21. Common Kartvelian/Georgian-Zan > Georgian.**

- |    |    |                |   |               |  |
|----|----|----------------|---|---------------|--|
| a. | CK | * <i>lag-</i>  | > | <i>rgav</i>   | ‘plant’                                  |
| b. | GZ | * <i>lab-</i>  | > | <i>lboba</i>  | ‘soaking until soft’                     |
| c. | CK | * <i>rekw-</i> | > | <i>rka</i>    | ‘remark’                                 |
| d. | CK | * <i>lexw-</i> | > | <i>lxoba-</i> | ‘to thaw, to melt’                       |
| e. | CK | * <i>rekw-</i> | > | <i>rk'uma</i> | ‘to say’                                 |
| f. | CK | * <i>reyw-</i> | > | <i>ryveva</i> | ‘demolish, collapse, to pour, pull down’ |

However, because consonant clusters are widespread in the Kartvelian languages the some of the proto forms for modern sonorant + obstruent onset clusters are reconstructed without an intervening vowel:

**5-22. Reconstructed Common Kartvelian/Georgian-Zan sonorant + obstruent onset.**

- |    |    |                 |   |               |                     |
|----|----|-----------------|---|---------------|---------------------|
| a. | GZ | * <i>rt-</i>    | > | <i>rtva-</i>  | ‘join’              |
| b. | GZ | * <i>rk'en</i>  | > | <i>rk'en-</i> | ‘to fight, wrestle’ |
| c. | CK | * <i>(s)ʒe-</i> | > | <i>rdze</i>   | ‘milk’              |
| d. | CK | * <i>mʒe-</i>   | > | <i>mze</i>    | ‘sun’               |

---

<sup>15</sup> PSI is Proto-Slavic, the ancestor to the Slavic languages.

Forms like those in (5-22a:d) lead Butskhrikidze (2002) to believe that Proto-Kartvelian had syllabic sonorants with the modern forms derived from syllabic sonorant + obstruent sequences. Vogt (1958), however, contends that the clusters are the result of vowel loss. It is possible that both are correct. As shown in (5-21), some clusters did arise from the loss of a vowel. Bell (1978) argues that all syllabic sonorants are the result of vowel loss. If this is true, syllabic sonorants could have been a stage in Georgian sonorant + obstruent onset cluster development.

Further evidence for a historical vowel in #RO comes from bases with multiple reconstructions. *rts'eva* 'to shake' and *rkvam* 'to cover, be covered' have both #RVO and #RO proto forms: GZ *\*ref<sup>h</sup>x-/\*rts'-* 'to shake; to swing' and CK *\*rekw-/ \*rkw-am-* 'to say', respectively. Finally, the earliest Kartvelian language to separate from the other languages, Svan, has a vowel where Common Kartvelian has a sonorant + obstruent onset sequence. Compare Svan *ləʒe* 'milk' and *məʒ/miʒ* 'sun' with the Georgian forms *rdze* and *mze* in (5-22c,d).

#### 5.3.3.4. Maipurean

Payne (1991) reconstructed Proto-Maipurean with a syllable shape of (C)V(C) with only /n/ and /h/ occurring in the coda. This is in stark contrast to Piro which allows sonority plateaus and sonority reversals (Matteson, 1965). Piro is the only Maipurean language that deviates so radically from the CV syllable shape (Aikhenvald, 1999, p. 78). So, like Austronesian, Piro shows evidence of clusters developing from the loss of an interconsonantal vowel *de novo*:

**5-23. Proto-Maipurean > Piro unstressed vowel loss.**

a.	<i>*kahit<sup>h</sup>i</i>	>	<i>ksiri</i>	‘moon’
b.	<i>*kiri</i>	>	<i>çri</i>	‘nose’
c.	<i>*mata</i>	>	<i>mta</i>	‘skin’
d.	<i>*nene</i>	>	<i>nne</i>	‘tongue’
e.	<i>*kafa</i>	>	<i>kfiwna</i>	‘armadillo’
f.	<i>*teni</i>	>	<i>tni</i>	‘breast (milk)’
g.	<i>*kasiukip<sup>h</sup>i</i>	>	<i>kfiyoçri</i>	‘cayman’
h.	<i>*mãka</i>	>	<i>mka</i>	‘clothing’
i.	<i>*matira</i>	>	<i>mfiira</i>	‘spider monkey’

**5.3.3.5. Oto-Manguean**

Rensch (1966) reconstructs Proto-Oto-Manguean nasal + obstruent clusters and palatal + consonant clusters, though the exact specifications of the segments are not described. Otherwise Proto-Oto-Manguean is constructed to have a syllable structure of CV(?). The Zapotecan branch, which includes Chatino, is reconstructed with a syllable shape of CV(?), without the palatal + consonant and nasal + obstruent onsets (Suárez, 1973; Swadesh, 1947). It is generally well-accepted that Chatino consonant clusters came about by the loss of unstressed interconsonantal vowels (Campbell, 2013, p. 401), as shown in (5-24):

**5-24. Proto-Zapotecan to Chatino unstressed vowel loss.**

a.	<i>*Laga?</i>	>	<i>lka<sup>23</sup></i>	‘leaf’
b.	<i>*yaga</i>	>	<i>yka<sup>2</sup></i>	‘tree’
c.	<i>*yuzi</i>	>	<i>ysin<sup>4</sup></i>	‘sand’
d.	<i>*luzu</i>	>	<i>wsin<sup>21</sup></i>	‘beard’
e.	<i>*Loba</i>	>	<i>lkwa<sup>45</sup></i>	‘sweep’
f.	<i>*lasi?</i>	>	<i>lti<sup>2</sup></i>	‘skinny’
h.	<i>*luze?</i>	>	<i>ltse<sup>23</sup></i>	‘tongue’

Oto-Manguean languages range from conservative varieties which maintain the Proto CV structure and allow polysyllabic forms to innovative varieties that have drifted toward monosyllabicity and allow few bases larger than two syllables. Compare the Panixtlahuaca Chatino forms in (5-25a:c) with their cognates in disyllabic varieties of Chatino in (5-25d:f):

**5-25. Monosyllabic and disyllabic cognates in Chatino.**

a.	<i>ltseʔ</i>	‘tongue’	~	d.	<i>lutzeʔ</i>	‘tongue’	Zenzontepec Chatino
b.	<i>wsaa</i>	‘weevil’	~	e.	<i>kusaa</i>	‘weevil’	Yaitepec Chatino
c.	<i>lkaʔ</i>	‘leaf’	~	f.	<i>likaʔ</i>	‘leaf’	San Juan Lachao

Like the Austronesian languages, some sonorant + obstruent onset clusters arose through vowel loss in an inflectional prefix. One example is habitual marker *ri-* > *r-* (Campbell, 2011, p. 226, fn. 11).

**5.3.3.6. Penutian**

A genetic relationship between Klamath-Modoc and other languages along the North American West Coast has yet to be widely accepted among most linguists (DeLancey & Golla, 1997). This makes any claim to the origin of sonorant + obstruent onset clusters in Klamath slightly weaker than that of Austronesian or Slavic. However, the efforts to connect the languages of the Southwest Canadian/Northwest US Coast has been fruitful (DeLancey, Genetti, & Rude, 1988; Tarpent, 1997). The available data demonstrates the same pattern that is seen with the aforementioned languages: sonorant + obstruent onset clusters in Klamath are cognate with sonorant + vowel + obstruent sequences in (purportedly) related languages (5-26):

**5-26. Klamath ~ Proposed Penutian cognates.**

a.	<i>lk'om</i>	‘coals, charcoal’	~	k.	<i>lak'im</i>	‘soot’	Sahaptian
b.	<i>ndan</i>	‘three’	~	l.	<i>mita-at; mātaat</i>	‘three’	Sahaptian; Nez Perce
c.	<i>ktena</i>	‘spear’	~	m.	<i>keetis</i>	‘spear’	Nez Perce
d.	<i>pk'isiisap</i>	‘mother’	~	n.	<i>pike</i>	‘(his) mother’	Nez Perce
e.	<i>ptisap</i>	‘father’	~	o.	<i>píst</i>	‘(his) father’	Nez Perce
f.	<i>wle</i>	‘run’	~	p.	<i>wîlé</i>	‘run, move quickly’	Nez Perce
h.	<i>nk'ey</i>	‘bullet, war arrow’	~	q.	<i>*nek'l; nok'</i>	‘draw a bow’; arrow	Proto-Yokuts Maidu
i.	<i>l'q<sup>h</sup>-</i>	‘down to the ground’	~	r.	<i>*léXw; láXw</i>	‘underside’	Proto-Tsimshianic; Nisqa'a
j.	<i>wden</i>	‘dream’	~	s.	<i>*wEq ; wé:wtuk</i>	‘to sleep’; ‘to camp overnight’	Proto-Tsimshianic; Nez Perce

Note that in (5-26k:s) the vowel quality is not predictable. As such, vowel loss is more likely than vowel insertion. Vowel insertion is rarely unpredictable (Hall, 2006). Certain vowels are more likely to be inserted, like schwas [ə], or the vowel quality is predictable from surrounding segments, as with vowel copying (Hall, 2006). Vowel loss, on the other hand, can target structural contexts and can affect most or all vowel types in the target context (Taylor, 1994).

### **5.3.3.7. Summary**

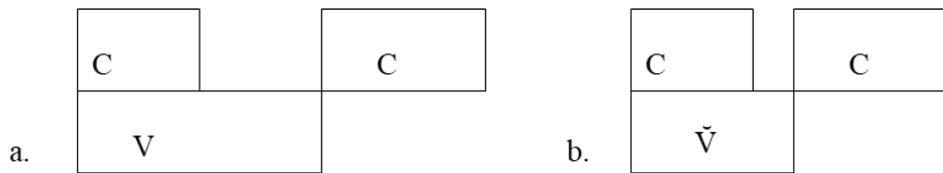
The only attested historical pathway for languages to develop tautosyllabic word initial sonorant + obstruent clusters is through interconsonantal unstressed vowel loss. Every language in this survey shows evidence of this sound change. Interconsonantal unstressed vowel loss can be explained by appealing to phonetic factors, suggesting that cross-linguistic sonority patterns do not require explanations rooted in complex, linguistic-specific mechanisms. In the next section, I discuss the naturalness of interconsonantal vowel loss.

## **5.4. General discussion**

### **5.4.1. Natural sound change?**

#### **5.4.1.1. Unstressed vowel loss**

Unstressed vowel loss has a clear phonetic base. Unstressed vowels are shorter than their stressed counterparts which makes them better candidates for eventual loss. Articulatory gestures for a consonant and vowel in a CV sequence are initiated synchronously. The consonantal gesture will always partially overlap with the vocalic gesture (Browman & Goldstein, 1992; Nam et al., 2009). The shorter a vowel is the more “hidden”, or less perceptible, it will be (Chitoran & Iskarous, 2008).



**Figure 5-3. Consonant + vowel gestural synchronicity.**

In Figure 5-3, the CV gestures are synchronous for both (a) and (b). The non-reduced vowel has less of its overall duration obscured by the consonantal gesture. The short vowel will be difficult to perceive because it is almost completely overlapped by the initial consonant. Furthermore, a short vowel requires movement of the tongue over a shorter period of time, which leads to “undershoot”, where the vowel does not reach its acoustic target (Lindblom, 1963).

#### **5.4.1.2. Phonologization of clusters as tautosyllabic**

The phonetic weakening of the vocalic gesture and a continuum of vocalic realizations along a hyper-hypo-articulated scale provides the conditions for a change in the phonological grammar (Blevins, 2004). The weakened vowel can be interpreted several ways. It could be interpreted as the presence of a vowel, consonantal release, or a gestural mistiming that allows for a brief period of an unobstructed vocal tract. A child, in acquiring their phonological system, may decide that there is no intervening phonological vowel. In the child’s phonology, the form with a consonant cluster is the underlying representation. If this happens, any presence of a vowel in hyper-articulated speech may be assumed to be an allophonic variant of the underlying #CCV structure, the opposite of the previous generations’ interpretation of a vowel being deleted in the output. This process would be an instantiation of CHOICE in EP.

The phonologization process could also be CHANGE. Sonorant + vowel + obstruent sequences, like *leba*, could be perceptually ambiguous with sonorant + obstruent onset sequences, like *lba*. Like vowels, sonorant consonants have continuous airflow, are naturally

voiced, and have formant structure (Chomsky & Halle, 1968; Ladefoged & Maddieson, 1996; Mielke, 2005). A listener may hear a sonorant + vowel + obstruent sequence as sonorant + obstruent. Berent et al. (2007) found that English speakers confused #RO sequences with #RVO at a high rate. Berent and colleagues interpreted these data as evidence for grammatical constraints on language learning, but the data are also consistent with ambiguity in the acoustic signals. Perceptual confusability of #RVO and #RO may even be the reason why #RO is relatively rare.

Vowel loss is a relatively common process, but the loss of a vowel is unlikely to give rise to the same ambiguity in all environments. For instance, voicing is naturally inhibited in obstruents which can cause a following vowel to be devoiced (Gordon, 1998). In this context, the perceptibility of the vowel will be reduced (Chitoran & Iskarous, 2008). Sonorants, however, are spontaneously voiced (Chomsky & Halle, 1968). Therefore, the voicing of the vowel will be preserved, making it more resistant to being perceived as a vowel-less sequence.

#RO onsets may be confusable for #RVO because of the multiple articulatory gestures associated with sonorant consonant production (Krakow, 1999; Pouplier & Beňuš, 2011; Sproat & Fujimura, 1993). Sonorant consonants are composed of consonant and vocalic gestures. The consonantal components are the gestures often associated with the segment: the tongue tip repeatedly contacting the alveolar ridge in the production of [r], the tongue creating partial closure at the alveolar ridge for [l], and the lips closing during the production of [m], etc.

The “vocalic element” could be a lack of oral constriction, a period of “relatively stable-state formant structure” (Ladefoged & Maddieson, 1996), dorsal retraction (Sproat & Fujimura, 1993), or velic lowering (Krakow, 1999). According to Sproat & Fujimura (1993), “light” and “dark” allophones of English /l/ differed in the relative sequencing of an apical and dorsal gestures. The retraction of the tongue dorsum, a gesture often associated with vowels (Krakow,

1999; Sproat & Fujimura, 1993), followed the apical constriction gesture when the /l/ was produced before the vowel. The dorsal gesture preceded the apical gesture post-vocally, suggesting the vocalic gesture moved toward the nucleus while the consonantal gesture moves toward the margin (Sproat & Fujimura, 1993). In Slovak, Pouplier & Beňuš (2011) found asynchronous apical and dorsal gestures for both syllabic [l] and syllabic [r]. Nasals do not have the same tongue retraction gesture as liquids, but velum lowering may function similar the dorsal gestures (Krakow 1999). In a sonorant + obstruent sequence, the vocalic element cannot be associated to any vowel and so could potentially be interpreted as a vowel.

#### **5.4.1.3. Conclusion**

Sonorant + obstruent onset clusters are typologically much rarer than obstruent + sonorant onsets clusters, a fact that is often held to be a result of the Sonority Sequencing Generalization. While the Sonority Sequencing Generalization has long been considered a phonological universal that explains phonotactic patterns and syllable shape (discussed in Parker, 2012), there are numerous problems with the proposal and the field remains divided on its relevance (Parker, 2012). Working within the framework of Evolutionary Phonology (Blevins, 2004, 2006a, 2007, 2015), I have found that onset clusters that violate the Sonority Sequencing Generalization evolved independently in at least six language families. In all six languages, sonorant + obstruent onset clusters developed from inter-consonantal unstressed vowel loss. Their rarity can be understood in terms of historical development and phonetics, without needing to rely on innate phonological constraints. Sonorants are continuous periodic signals with formant structure, are produced with vocalic gestures, and are less likely than obstruents to devoice a following vowels. Because of these intrinsic properties the full loss of a vowel in a #RVO sequence is expected to be rarer than loss in a #OVR.

## **Chapter 6. Conclusions**

### **6.1. Introduction**

In this dissertation, I discuss aspects of uniformity and variability in three evolutionary distinct species: budgerigars, house finches, and humans. Despite the differences between these species, I believe the work I present here has revealed important similarities, similarities which further our understanding of how acoustic systems can vary across populations, what factors might limit variability in a species, and how cross-population diversity can emerge from shared traits.

In chapter 2, I reviewed similarities in the neurological, genetic, perceptual, and production mechanisms of human and avian vocal behavior. Previous research has provided strong evidence in support of a comparative approach to behavior. The data from the rest of this dissertation build on that previous comparative work. The prior research in songbirds and humans provided an explanation for the cross-population patterns that I found in budgerigar segment organization. In turn, the findings of the budgerigar chapter (Chapter 3) and the house finch chapter (Chapter 4) can help refine linguistic theory. The CV preference in human languages has been explained by appealing to species-specific cognitive mechanism which are specific to language (Prince & Smolensky, 2002). The similar patterns that I found in budgerigars suggest that species-specific mechanisms are unwarranted. Furthermore, the house finch data suggest that caution should be taken when making claims about what is possible across languages. Trilling is absent in wild house finch populations, yet canary-tutored house finches were able to learn to trill. The organizational pattern of the canary-tutored song and the wild house finch song is remarkably different, yet both were learnable.

#### **6.1.1. Dissertation summary**

To review the findings chapters 3 - 5:

For budgerigars, I designed an algorithm to segment their species-typical song, warble, using acoustic transitions rather than intervals of silence. Because the vast majority of work in animal acoustic communication focuses on units separated by silence, this chapter provides methods which will permit us to delve deeper into the complexities of non-human vocal systems. The data which resulted from the algorithm reveal that budgerigars can achieve a high degree of complexity and variability by combining and arranging these small, more stereotyped units. Furthermore, the data show that budgerigar segment organization is not only consistent across independent budgerigar populations but is consistent with patterns found in human language. For example, I found patterns in budgerigar segment organization which appear to be similar to phrase-final lengthening (Edwards et al., 1991; Fougeron & Keating, 1997), F0 declination (Ladd, 1988), intensity declination (Vaissière, 1995), and the CV preference (Fougeron & Keating, 1997; Jakobson & Halle, 1956; Westbury & Keating, 1986).

For house finches, I found that individuals can reliably learn to trill, despite the lack of trilling in the song of wild populations. The house finch is not known to be an adept mimic and songs from populations across North America share the same basic description. These data suggest that house finches are more permissive in their song learning program than it would seem given their cross-population patterns.

For humans, I found that onset clusters with Sonority Sequencing Generalization violations have developed in at least six independent language families. In all six cases, the loss of a vowel between a sonorant and obstruent led to complex onsets of sonorant + obstruent. These data suggest that sonorant + obstruent onsets are rare because of phonetic factors. Namely, sonorants have intrinsic properties which are more likely to preserve the percept of a vowel.

While these projects are all different, evidence from all three support the following claims: (i) cross-population commonality does not reflect the full potential of variability; (ii) peripheral

mechanisms have a strong influence in limiting cross-population variability; (iii) high degrees of variation can emerge from uniform traits.

## **6.2. Universals and learnability**

Cross-linguistic research has played a central role in the formation of linguistic theory, though what these data tell us about the nature of human language has been debated (Blevins, 2004; Croft, 2003; Evans & Levinson, 2009; Newmeyer, 2005). Chomsky (1965, 1981, 2007) has argued that cross-linguistic patterns reflect innate properties of the human brain which guide and constrain language learning. Under this view, certain languages are impossible and unlearnable. Baker (2002) proposes that the lack of languages with a word order of auxiliary-subject-verb-object is evidence for their impossibility. Other frameworks suggest that recurrent patterns are not necessarily unlearnable or impossible but may be less common or non-existent due to communicative pressures and limitations on general cognitive abilities (Blevins, 2017; Evans & Levinson, 2009).

Innate learning constraints which are species- and behavior-specific cannot be ruled out, but the data presented in this dissertation suggest that the innate linguistic constraints hypothesis should be regarded with skepticism, at least in a strong form. In these three species, cross-population patterns are not reflective of what sounds and sound patterns are actually permissible by the learning machinery. Budgerigar warble is consistent across multiple independent populations (Chapter Chapter 3), yet previous research has found that budgerigars can mimic sounds from a diverse range of sources, sounds which are distinct from their own species-typical signal. House finches (Chapter Chapter 4) learn to trill when exposed to the song of a trilling species, the canary, even though trills are rare in wild populations. Sonority Sequencing Generalization violating onsets may be rare in human language, but they have evolved at least in at least six independent populations (Chapter Chapter 5). Furthermore, Sonority Sequencing Generalization violating onsets have remained stable for multiple

generations. In all three cases, rare patterns emerge only after being exposed to a relatively narrow set of circumstances.

### **6.3. Peripheral mechanisms of recurrent patterns**

The data presented here provide further evidence that universal tendencies in sound systems often have their roots in peripheral mechanisms of sound production and perception (Blevins, 2004; Ohala, 1983; Podos, 1996). The consistent patterns found in budgerigar segment organization (Chapter 3) all seem to be rooted in a phonetic basis. House finch trilling (Chapter 4) is likely limited by body size and beak shape. The development of sonorant + obstruent onset clusters from sonorant + vowel + obstruent sequences in human languages (Chapter 5) suggest that sonorant + obstruent clusters are limited by phonetic factors. Namely, the sonorant-obstruent transition, when the sonorant is in a word-initial position, has acoustic cues that are similar to vocalic cues. Because most members of a species will share vocal morphology and auditory capabilities, the physical form will be one of the primary sources of within-species uniformity.

### **6.4. Variation out of uniformity**

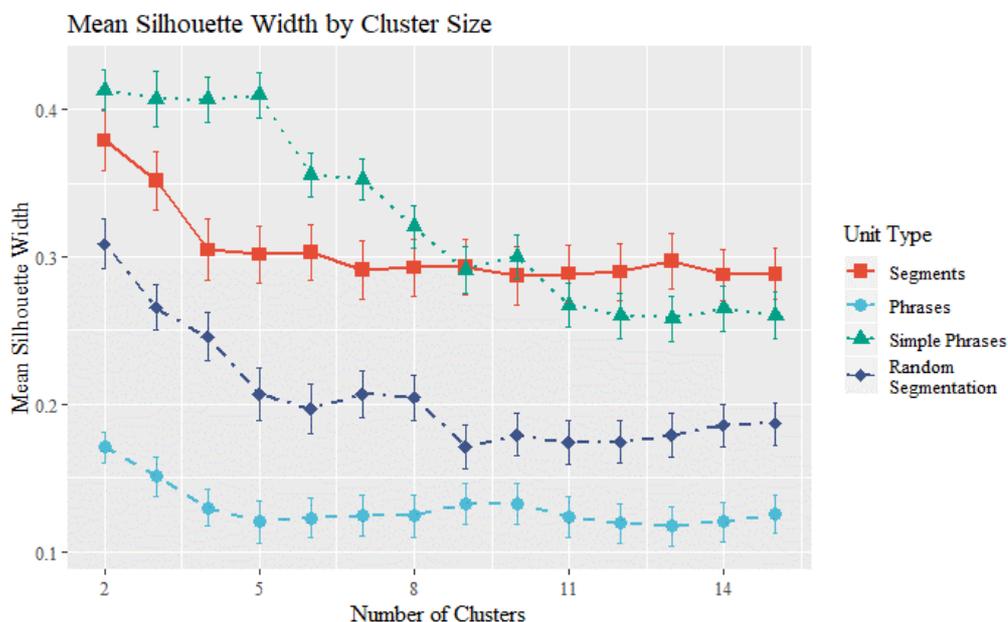
According to Abler (1989) a property of complex systems is that they can achieve high degrees of diversity by combining and rearranging smaller, more stereotyped units. This property is known as the particulate principle. Human language is the canonical example of a particulate system, but I found that budgerigar warble and house finch song adhere to this principle as well. In budgerigars (Chapter 3), I found that segments are more stereotyped and more consistent across independent populations of budgerigars than the phrases that segments build. The differences in stereotypy between phrases and segments is expected if budgerigars adhere to the particulate principle. As far as I am aware, this is the first evidence of the particulate principle in non-human segments. In house finches (Chapter 4), phrases were not clearly different across all house finch groups, suggesting basic phrase patterns are available

to house finches early in development. Variation emerges from the modification and organization of these basic elements. These data suggest that all complex acoustic communication systems may adhere to similar principles and that further research may be able to discover even more “low level” units in animal communication systems, such as distinctive features or articulatory gestures in human language (Browman & Goldstein, 1992; Chomsky & Halle, 1968; Mielke, 2004).

## Chapter 7. Appendix

### 7.1. Chapter 3 supplement

To check whether the segmentation algorithm produced reliable units, I compared segments to two other unit types, simple phrases and random segments. From a visual inspection, segments often seem to be shorter versions of the simple phrases found in budgerigar warble. As discussed in 3.1.2 and shown in Figure 3-2 and Figure 3-3, unlike complex phrases, simple phrases are more stereotyped and vary only in a few dimensions (such as duration). As such, simple phrases should produce tighter clusters than complex phrases and should cluster similar to segments. I also randomly segmented the complex phrases to ensure that the results were not an artefact of unit division or having shorter units. As shown in Figure 7-1, the random segmentation performed worse than my algorithm's segmentation. However, it was still better than the clustering of complex phrases. Simple phrases cluster as well, if not better, than segments.



**Figure 7-1. Silhouette width comparison for budgerigar warble units.** Comparison of silhouette widths for segments (red squares, solid line), complex phrases (blue circles, dashed line), simple phrases (green triangles, dotted line), and segments which were the byproduct of a random segmentation (purple diamonds, alternating line). Error bars are standard error of the mean.

## 7.2. Code

Code for this dissertation is posted on GitHub: <https://github.com/DanCMann/Dissertation-code>

### 7.3. Sonorant + obstruent onset list

Language	Word	Gloss	Data source
Georgian	rbili	soft; mild	Klimov 1998
Georgian	rt <sup>h</sup> wa	to spin	Klimov 1998
Georgian	rt <sup>h</sup> va-	join	Klimov 1998
Georgian	rt'x'ma	to hit	Rayfield 2006
Georgian	rt'o	branch	Klimov 1998
Georgian	rgoli	ring	Rayfield 2006
Georgian	rgav	plant	Rayfield 2006
Georgian	rg-	to plant	Klimov 1998
Georgian	rkvam	to cover, be covered	Klimov 1998
Georgian	rk'a	horn	Rayfield 2006
Georgian	rk <sup>h</sup> a	remark	Rayfield 2006
Georgian	rk'en-	to fight, wrestle	Klimov 1998
Georgian	rk'uma	to say	Rayfield 2006
Georgian	rdze	milk	Rayfield 2006
Georgian	rtskhila	hornbeam	Rayfield 2006
Georgian	rts'eva	to shake	Klimov 1998
Georgian	rdza	toil;trouble	Rayfield 2006
Georgian	rt'feba	staying	Rayfield 2006
Georgian	rxevit	oscillate	Rayfield 2006
Georgian	ryveva	demolish, collapse, to pour, pull down	Klimov 1998
Georgian	rx'eva	fluctuation	Rayfield 2006
Georgian	rva	eight	Klimov 1998
Georgian	lb-	to become tender	Klimov 1998
Georgian	lbili	soft	Rayfield 2006
Georgian	lboba	soaking until soft	Rayfield 2006
Georgian	lp'eba	rots	Rayfield 2006
Georgian	lxena	joy	Rayfield 2006
Georgian	lyveba	to thaw, to melt	Klimov 1998
Georgian	lt'olva	aspiration	Rayfield 2006
Georgian	mbayi	creator	Rayfield 2006
Georgian	mdagio	common people	Rayfield 2006
Georgian	mt <sup>h</sup> idan	mountain	Rayfield 2006
Georgian	mt'eri	enemy	Rayfield 2006
Georgian	mgeli	wolf	Rayfield 2006
Georgian	mk'a	crops	Rayfield 2006
Georgian	mp <sup>h</sup> aravi	protector	Rayfield 2006
Georgian	mp'ala	compost	Rayfield 2006
Georgian	mk <sup>h</sup> anri	rising river	Rayfield 2006
Georgian	mts <sup>h</sup> onara	lazy	Rayfield 2006
Georgian	mts'eri	insect	Rayfield 2006
Georgian	mtʃ <sup>h</sup> at'e	lightweight	Rayfield 2006
Georgian	mtʃ <sup>h</sup> evri	eloquent	Rayfield 2006
Georgian	mdzima	heavy	Rayfield 2006
Georgian	mdzobi	better	Rayfield 2006
Georgian	mze	sun	Klimov 1998
Georgian	msgasvad	like, similar to	Rayfield 2006

Georgian	mzawe	sour	Rayfield 2006
Georgian	mʃeva	starving	Rayfield 2006
Georgian	myaravi	dying	Rayfield 2006
Georgian	mxari	shoulder	Rayfield 2006
Georgian	mχ'esi	tendon	Rayfield 2006
Georgian	ndoba	faith	Rayfield 2006
Georgian	nt <sup>h</sup> xeva	spilling	Rayfield 2006
Georgian	ndzreva	swaying	Rayfield 2006
Georgian	ntʃ <sup>h</sup> kleva	shake liquid	Rayfield 2006
Georgian	ngreva	collapse	Rayfield 2006
Russian	rdeɫʲ	to blush	Scheer 2006
Russian	rdestʲ	bot. knotgrass	Scheer 2006
Russian	rtutʲ	mercury	Scheer 2006
Russian	rta	mouth (gen sg)	Scheer 2006
Russian	rzi	rye (Gen sg)	Scheer 2006
Russian	rʒavyj / rʒavetʲ	rust	Scheer 2006
Russian	rʒatʲ	neigh	Scheer 2006
Russian	rva	dig (gen sg)	Scheer 2006
Russian	rvatʲ	tear, rip	Scheer 2006
Russian	lba	forehead (gen)	Scheer 2006
Russian	lʲda	ice	Russian Corpus
Russian	lgatʲ	lie (inf)	Scheer 2006
Russian	lʲgota	respite	Scheer 2006
Russian	lʲstitʲ	cunning, ruse	Scheer 2006
Russian	lʲzja	it is suitable to	Scheer 2006
Russian	lzi	lie (gen sg)	Scheer 2006
Russian	lʲva	lion (gen sg)	Scheer 2006
Russian	mglɑ	mist	Scheer 2006
Russian	mknutʲ	sudden shocking movement	Scheer 2006
Russian	mzda	salary	Scheer 2006
Russian	mstitʲ	revenge	Scheer 2006
Russian	mʃitʲ	to cover with moss	Scheer 2006
Russian	mʃitsɑ	aphid	Scheer 2006
Piro	mpikleḥoti	courage	Nies 1986
Piro	mpiçkakakliti	unequal	Nies 1986
Piro	mpohiretkoti	peon	Nies 1986
Piro	mporo	donkey	Nies 1986
Piro	mtɑ	skin	Payne 1991
Piro	mtiriçi	nine	Nies 1986
Piro	mti-ri	infant	Nies 1986
Piro	mtɑ-ḥɑ	eyelid	Key 2015
Piro	mkoli	frog	Nies 1986
Piro	mka-lu	cloth, clothing	Matteson 1963
Piro	mkoçe	tuft, cloud, cluster	Nies 1986
Piro	mkaɫfri	fabric	Sebastian 2006
Piro	mkaɫi	spider web	Key 2015
Piro	mkaɫjaɫo	unmarried woman	Nies 1986

Piro	mtseriti	skinny	Nies 1986
Piro	mtserkakati	animals	Nies 1986
Piro	mʃira	spider monkey	Hanson 2010
Piro	msaçi	not black	Matteson 1963
Piro	msapatneti	barefoot	Nies 1986
Piro	mʃiri	hair comb	Nies 1986
Piro	mʃikoti	soft	Matteson 1963
Piro	mʃiko-tʃkiti	difficult to uproot	Nies 1986
Piro	mçihati	clean	Nies 1986
Piro	npika	I am afraid	Matteson 1963
Piro	ntoçe	cassabanana	Nies 1986
Piro	nkaʃit-ʃa	I caught	Matteson 1963
Piro	ntspatate	my guava	Sebastian 2006
Piro	nteiwa	black cricket	Nies 1986
Piro	nso	genipa	Matteson 1963
Piro	nʃinikani	almost	Nies 1986
Piro	wpowratlu	we clean it	Matteson 1963
Piro	wtiplata	we sit down	Matteson 1963
Piro	wseriwna	we grow	Matteson 1963
Piro	wʃkotite	our white monkey	Nies 1986
Piro	wteirika	we kindle	Matteson 1963
Klamath	lpa-	plant	Barker 1963
Klamath	lpawal-	round object lies on top of	Barker 1963
Klamath	lpo-	tunnel	Barker 1963
Klamath	lp <sup>h</sup> eʔ-	roll in snow	Barker 1963
Klamath	ltakal-	picks up a round object	Barker 1963
Klamath	ltok <sup>h</sup> w-	to have a crush on	Barker 1963
Klamath	lt <sup>h</sup> ew-	eat tules	Barker 1963
Klamath	lt <sup>h</sup> ikʔ-	lope, pace	Barker 1963
Klamath	lt <sup>h</sup> o:qʔ-	be dappled	Barker 1963
Klamath	lt <sup>h</sup> p-	away from against	Barker 1963
Klamath	ltʔoqʔ-	thump with finger and thumb	Barker 1963
Klamath	lki-	motion toward for a purpose	Barker 1963
Klamath	lkʔom-	charcoal	Barker 1963
Klamath	lqawʔa:wʔalʔ	finger	Barker 1963
Klamath	lqe-	has a stripe on the hair	Barker 1963
Klamath	lqe:j-	be cramped	Barker 1963
Klamath	lʔq <sup>h</sup> -	down to the ground	Barker 1963
Klamath	lqʔan-	ripples	Barker 1963
Klamath	lqʔaq <sup>h</sup> -	pole	Barker 1963
Klamath	lʃ <sup>ʔ</sup> i-	project in a line	Barker 1963
Klamath	lʃ <sup>ʔ</sup> wj-	right up to	Barker 1963
Klamath	mpaw	hoot	Barker 1963
Klamath	mpelj	to be cross-eyed	Barker 1963
Klamath	mputjʔ	wrinkle from exposure to water	Barker 1963
Klamath	mpa	cooking rock	Barker 1963
Klamath	mp <sup>h</sup> akʔ	gasp, belch	Barker 1963
Klamath	mp <sup>h</sup> etʔ	float	Barker 1963
Klamath	mpʔaq	have spots	Barker 1963
Klamath	mpʔaq	dry up	Barker 1963

Klamath	mʃ <sup>h</sup> ik-	[augmentative]	Barker 1963
Klamath	mʃ <sup>h</sup> n-	have, get, possess	Barker 1963
Klamath	mʃ <sup>h</sup> elq	feel, grasp, understand	Barker 1963
Klamath	msa-	prairie dog	Barker 1963
Klamath	npawi'a	burst with a round instrument	Barker 1963
Klamath	ntalk	root	Barker 1963
Klamath	ntuʃ <sup>ʔ</sup>	freeze	Barker 1963
Klamath	ntan	three	Delacey 1992
Klamath	ntuk	dote on	Barker 1963
Klamath	nt <sup>h</sup> ey	bow	Barker 1963
Klamath	nt <sup>h</sup> iq'	drip	Barker 1963
Klamath	nt <sup>h</sup> ul	flow	Barker 1963
Klamath	nt <sup>h</sup> up <sup>h</sup>	choke	Barker 1963
Klamath	nt'ak'	be stuck	Barker 1963
Klamath	nt'iw	fall	Barker 1963
Klamath	nt'up'	rot, spoil	Barker 1963
Klamath	nkak-	turtle	Barker 1963
Klamath	nkat'-	jump	Barker 1963
Klamath	nkeno:-	mons Veneris	Barker 1963
Klamath	nkoq <sup>h</sup> -	black helldiver	Barker 1963
Klamath	nk <sup>h</sup> a-	stomach	Barker 1963
Klamath	nk'ey-	bullet; war	Delancey 1997
Klamath	nk <sup>h</sup> ilik'-	be dusty	Barker 1963
Klamath	nk'af <sup>ʔ</sup> a	cuts off the head with a round instrument	Barker 1963
Klamath	nk'ejsʔa-	shoot pl objs	Barker 1963
Klamath	nqif <sup>ʔ</sup> -	be tight	Barker 1963
Klamath	nqol-	jackrabbit	Barker 1963
Klamath	nq <sup>h</sup> aq <sup>h</sup> ki-	give birth	Barker 1963
Klamath	nq <sup>h</sup> en-	shouts, yells	Barker 1963
Klamath	nq <sup>h</sup> ew'a	breaks with round instrument	Barker 1963
Klamath	nq <sup>h</sup> ot'-	scorch	Barker 1963
Klamath	nq'aq-	crown of the head	Barker 1963
Klamath	nq'iq'-	weigh	Barker 1963
Klamath	nq'oʃ <sup>ʔ</sup> a	bends with a round instrument	Barker 1963
Klamath	nt <sup>h</sup> alq	freshly	Barker 1963
Klamath	nt <sup>h</sup> ayak'	listen for	Barker 1963
Klamath	nt <sup>h</sup> eqi	becomes exasperated	Barker 1963
Klamath	nt <sup>h</sup> ik'	melt	Barker 1963
Klamath	nt <sup>h</sup> uq'	be deaf	Barker 1963
Klamath	nt <sup>ʃ</sup> ama:s	wipe	Barker 1963
Klamath	nt <sup>ʃ</sup> et	inner bark	Barker 1963
Klamath	nt <sup>ʃ</sup> iw	pop	Barker 1963
Klamath	nt <sup>ʃ</sup> ek'	small, little pieces	Barker 1963
Klamath	wpe	fringe	Barker 1963
		hits with a long instrument and bloodies	
Klamath	wp <sup>h</sup> up'a	someone's nose	Barker 1963
Klamath	wp'ak'a	smashes with a long instrument	Barker 1963
Klamath	wp'eq'a	hits in the face with a long instrument	Barker 1963
Klamath	wten	dream	Delacey et al 1988
Klamath	wtum	swim	Barker 1963

Klamath	wtu:t <sup>hk</sup>	catterpillar	Barker 1963
Klamath	w <sup>h</sup> ewʃ <sup>~</sup> a	snaps back on	Barker 1963
Klamath	wt'ek'a	breaks into pieces with a long instrument	Barker 1963
Klamath	wt'am'	put a lid on	Barker 1963
Klamath	wkip'	to be constipated	Barker 1963
Klamath	wk <sup>h</sup> ek'a	strikes on the back with a large instrument	Barker 1963
Klamath	wk'al'a	cuts off with a long instrument	Barker 1963
Klamath	wk'atʃ'a	cuts off the head with a long instrument	Barker 1963
Klamath	wqaw'	shine; moon	Barker 1963
Klamath	wqu:	act with wife's parents	Barker 1963
Klamath	wq <sup>h</sup> am	plant	Barker 1963
Klamath	wq <sup>h</sup> apq'	flap the wings	Barker 1963
Klamath	wq <sup>h</sup> a	fishpole	Barker 1963
Klamath	wq <sup>h</sup> iw	extend out into a plain	Barker 1963
Klamath	wq'a	quartz	Barker 1963
Klamath	wʃuq'a	washes	Barker 1963
Klamath	wʃ <sup>h</sup> ewa	jumps	Barker 1963
Klamath	wʃ <sup>h</sup> aq	reed	Barker 1963
Klamath	wʃ <sup>h</sup> ak	sucker	Barker 1963
Klamath	wsu	chest	Barker 1963
Chatino	rta <sup>21</sup>	brazada	Pride & Pride 2010
Chatino	rta <sup>45</sup>	sudor	Pride & Pride 2010
Chatino	rten <sup>32</sup>	nido	Pride & Pride 2010
Chatino	rkin <sup>32</sup>	quemó	Pride & Pride 2010
Chatino	rkunʔ <sup>21</sup>	tocó	Pride & Pride 2010
Chatino	rkunʔ <sup>4</sup>	cerró	Pride & Pride 2010
Chatino	rk <sup>w</sup> a <sup>23</sup>	aplicó	Pride & Pride 2010
Chatino	rk <sup>w</sup> a <sup>4</sup>	coló	Pride & Pride 2010
Chatino	rk <sup>w</sup> a <sup>45</sup>	se sentó	Pride & Pride 2010
Chatino	rk <sup>w</sup> a <sup>45</sup>	obedejo	Pride & Pride 2010
Chatino	rk <sup>w</sup> i <sup>2</sup>	corrió a otro	Pride & Pride 2010
Chatino	rk <sup>w</sup> iʔ <sup>43</sup>	excorió	Pride & Pride 2010
Chatino	rsun	razon	Pride & Pride 2010
Chatino	lta <sup>21</sup>	braza	Pride & Pride 2010
Chatino	ltsaʔ <sup>12</sup>	mojado	Pride & Pride 2010
Chatino	ltseʔ kiiʔ	llamas de fuego	Pride & Pride 2010
Chatino	lka <sup>23</sup>	to be (present)	Pride & Pride 2010
Chatino	lkaʔ <sup>23</sup>	cuarto	Pride & Pride 2010
Chatino	lkaʔ <sup>23</sup>	dorifora, catarina (an insect)	Pride & Pride 2010
Chatino	lkii <sup>23</sup>	jalar (gerund)	Pride & Pride 2010
Chatino	lku <sup>2</sup>	pecho (pos)	Pride & Pride 2010
Chatino	lku <sup>23</sup>	comida (pos)	Pride & Pride 2010
Chatino	lkuʔ <sup>23</sup>	vestir (present)	Pride & Pride 2010
Chatino	lk <sup>w</sup> a <sup>4</sup>	plano, estar plano	Pride & Pride 2010
Chatino	lk <sup>w</sup> a <sup>45</sup>	ser colado (pas)	Pride & Pride 2010
Chatino	lk <sup>w</sup> an <sup>43</sup>	temblar	Pride & Pride 2010
Chatino	lk <sup>w</sup> an <sup>43</sup>	bendecir	Pride & Pride 2010

Chatino	lk <sup>w</sup> an <sup>45</sup>	ser bendecido	Pride & Pride 2010
Chatino	lk <sup>w</sup> i <sup>43</sup>	hervir	Pride & Pride 2010
Chatino	lk <sup>w</sup> i <sup>45</sup>	tepache; licor	Pride & Pride 2010
Chatino	lk <sup>w</sup> i <sup>45</sup>	inclinado hacia abajo	Pride & Pride 2010
Chatino	lk <sup>w</sup> i <sup>45</sup>	bajando	Pride & Pride 2010
Chatino	lk <sup>w</sup> ii <sup>43</sup>	volar	Pride & Pride 2010
Chatino	lk <sup>w</sup> i <sup>?</sup> 43	excoriar	Pride & Pride 2010
Chatino	lk <sup>w</sup> in <sup>32</sup>	retrato	Pride & Pride 2010
Chatino	ltsu <sup>?</sup> 45	aguijon	Pride & Pride 2010
Chatino	lsu <sup>?</sup> 32	copete	Pride & Pride 2010
Chatino	mba <sup>45</sup>	friend.pos	Pride & Pride 2010
Chatino	mbi <sup>?</sup> ya <sup>4</sup>	buy.pret	Pride & Pride 2010
Chatino	mble <sup>?</sup> 43	pigheaded	Pride & Pride 2010
Chatino	mda <sup>?</sup> an <sup>23</sup>	walked	Pride & Pride 2010
Chatino	mdaa <sup>?</sup> 45	break.pret	Pride & Pride 2010
Chatino	mdo <sup>?</sup> o <sup>2</sup>	leave.pret	Pride & Pride 2010
Chatino	mta <sup>4</sup>	seed	Pride & Pride 2010
Chatino	mti <sup>45</sup>	trash	Pride & Pride 2010
Chatino	md <sup>?</sup> ii <sup>4</sup>	cigarret	Pride & Pride 2010
Chatino	mt <sup>?</sup> i <sup>32</sup>	dry.pret	Pride & Pride 2010
Chatino	mgin <sup>2</sup>	archbishop	Pride & Pride 2010
Chatino	mke <sup>?</sup> 23	heat.pret	Pride & Pride 2010
Chatino	mk <sup>?</sup> in <sup>12</sup>	burn.pret	Pride & Pride 2010
Chatino	mg <sup>w</sup> i <sup>?</sup> in <sup>23</sup>	hit.pret	Pride & Pride 2010
Chatino	mk <sup>w</sup> i <sup>32</sup>	boil.pret	Pride & Pride 2010
Chatino	mk <sup>w</sup> i <sup>?</sup> 32	sweep.pret	Pride & Pride 2010
Chatino	mdzu <sup>32</sup>	drizzle.pret	Pride & Pride 2010
Chatino	mtse <sup>?</sup>	small	Pride & Pride 2010
Chatino	mtsan <sup>?</sup> 43	fleco	Pride & Pride 2010
Chatino	mtse <sup>?</sup>	small	Pride & Pride 2010
Chatino	md <sup>?</sup> jin	monkey	Pride & Pride 2010
Chatino	m <sup>?</sup> fa <sup>32</sup>	broke	Pride & Pride 2010
Chatino	m <sup>?</sup> fa <sup>?</sup> 2	to get wet.pret	Pride & Pride 2010
Chatino	m <sup>?</sup> fan <sup>?</sup> 43	hairy	Pride & Pride 2010
Chatino	m <sup>?</sup> fa <sup>?</sup> 2	to get wet.pret	Pride & Pride 2010
Chatino	m <sup>?</sup> fan <sup>?</sup> 43	hairy	Pride & Pride 2010
Chatino	msa <sup>45</sup>	onza (medida)	Pride & Pride 2010
Chatino	msa <sup>?</sup> an <sup>23</sup>	filled	Pride & Pride 2010
Chatino	mska <sup>45</sup>	tear	Pride & Pride 2010
Chatino	msaa <sup>45</sup>	mesa	Pride & Pride 2010
Chatino	mse <sup>?</sup> 32	shattered	Pride & Pride 2010
Chatino	msk <sup>w</sup> a <sup>4</sup>	throw.pret	Pride & Pride 2010
Chatino	mstan <sup>45</sup>	peel.pret	Pride & Pride 2010
Chatino	m <sup>?</sup> i <sup>?</sup> 23	tomato	Pride & Pride 2010
Chatino	ndaa <sup>4</sup>	bean	Pride & Pride 2010
Chatino	nta <sup>21</sup>	wait.pret	Pride & Pride 2010
Chatino	nd <sup>?</sup> iaa <sup>4</sup>	roadrunner	Pride & Pride 2010
Chatino	nt <sup>?</sup> ia <sup>?</sup> 45	chew.pret	Pride & Pride 2010
Chatino	nga <sup>?</sup> a <sup>?</sup> 21	green	Pride & Pride 2010
Chatino	ngan <sup>32</sup>	coco	Pride & Pride 2010

Chatino	nkun? <sup>21</sup>	search	Pride & Pride 2010
Chatino	ng <sup>w</sup> aan <sup>43</sup>	weasel	Pride & Pride 2010
Chatino	nk <sup>w</sup> a <sup>4</sup>	count.pret	Pride & Pride 2010
Chatino	nk <sup>wi</sup> ? <sup>32</sup>	boil.pret	Pride & Pride 2010
Chatino	nɸa? <sup>32</sup>	to get wet.habitual	Pride & Pride 2010
Chatino	ndze <sup>4</sup>	suddenly	Pride & Pride 2010
Chatino	ntsuwa? <sup>45</sup>	holy grass (herb)	Pride & Pride 2010
Chatino	ndʒaa <sup>43</sup>	make a mistake.hab	Pride & Pride 2010
Chatino	ndʒuwi? <sup>32</sup>	embers	Pride & Pride 2010
Chatino	nɸga <sup>45</sup>	all	Pride & Pride 2010
Chatino	nsge <sup>4</sup>	guava	Pride & Pride 2010
Chatino	nsin? <sup>45</sup>	table	Pride & Pride 2010
Chatino	nsin? <sup>45</sup>	strip	Pride & Pride 2010
Chatino	nɸen <sup>12</sup>	anis	Pride & Pride 2010
Chatino	nɸen <sup>12</sup>	anise	Pride & Pride 2010
Chatino	wta <sup>4</sup>	cow	Pride & Pride 2010
Chatino	wɸi? <sup>23</sup>	dry	Pride & Pride 2010
Chatino	wke? <sup>23</sup>	to heat	Pride & Pride 2010
Chatino	wɸe <sup>43</sup>	alge	Pride & Pride 2010
Chatino	wɸi? <sup>32</sup>	lion	Pride & Pride 2010
Chatino	wsaa <sup>32</sup>	gorgojo (insect)	Pride & Pride 2010
Chatino	wsin <sup>21</sup>	beard	Pride & Pride 2010
Chatino	wɸa? <sup>21</sup>	floor	Pride & Pride 2010
Chatino	wɸe <sup>12</sup>	blister	Pride & Pride 2010
Chatino	jta <sup>4</sup>	zanja pequena	Pride & Pride 2010
Chatino	jka <sup>2</sup>	tree	Pride & Pride 2010
Chatino	jka ke	horn	Pride & Pride 2010
Chatino	jk <sup>w</sup> a ɸla	atole de chocolate	Pride & Pride 2010
Chatino	jtsa? <sup>43</sup>	advise.pret	Pride & Pride 2010
Chatino	jsiin <sup>4</sup>	sand	Pride & Pride 2010
Tsou	rsəə	tears	Blust & Trussel
Tsou	rtuu	loquat tree and fruit: Eriobotrya deflexa	Blust & Trussel
Tsou	rvorə	flying squirrel	Blust & Trussel
Tsou	mbaktsu	break	Chen 2002
Tsou	mpohsu	easy	Wright 1996
Tsou	mpitvihi	seventy	Wright 1996
Tsou	mpusku	twenty	Wright 1996
Tsou	mɸiŋi	beautiful bell-like tone	Wright 1996
Tsou	mtoki	to throw	Wright 1996
Tsou	mtujhu	thirty	Wright 1996
Tsou	mkameosi	quick recovery of health	Wright 1996
Tsou	mtsuu	ear of grain	Wright 1996
Tsou	mtso:	eye	Wright 1996
Tsou	mtsói	die	Blust & Trussel
Tsou	mɸei? <sup>si</sup>	to cover	Wright 1996
Tsou	mvore	dry in the sun	Blust & Trussel
Tsou	mvoe	to dry grain in sun	Wright 1996
Tsou	mza:	our	Wright 1994
Tsou	msapie	put on shoe	Wright 1994
Tsou	msaɸi	step on	Wright 1996

Tsou	msipɲi	to wedge	Wright 1996
Tsou	msua	to buy	Wright 1996
Tsou	mse:zutu	carry under clothing	Wright 1996
Tsou	nteʔo	conditional prefix	Wright 1996
Tsou	ntee	probably	Wright 1996
Tsou	ntosi	wish	Wright 1996
Tsou	nsoo	pond	Wright 1996
Tsou	ɲɖuju	cock's comb	Wright 1996
Tsou	ɲtosi	white hair	Wright 1996
Tsou	ɲfuju	antler	Wright 1996

## Chapter 8. References

- Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Systems*, 12(1), 1–13.
- Adret, P. (1993). Vocal learning induced with operant techniques - an overview. *Netherlands Journal of Zoology*, 43(1–2), 125–142.
- Aikhenvald, A. Y. (1999). The Arawak language family. In R. Dixon & A. Y. Aikhenvald (Eds.), *The Amazonian Languages* (pp. 64–106). Cambridge: Cambridge University Press.
- Araya-Salas, M., & Wright, T. (2013). Open-ended song learning in a hummingbird. *Biology Letters*, 9(5), 20130625.
- Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*, 320(5876), 630–634.
- Arriaga, G., 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), e46610.
- Aubin, T., & Bremond, J. (1983). The process of species-specific song recognition in the skylark *Alauda arvensis*. An experimental study by means of synthesis. *Zeitschrift Für Tierpsychologie*, 61(2), 141–152.
- Badyaev, A. V., & Hill, G. E. (2002). Paternal care as a conditional strategy: Distinct reproductive tactics associated with elaboration of plumage ornamentation in the house finch. *Behavioral Ecology*, 13(5), 591–597.
- Badyaev, A. V., Whittingham, L. A., & Hill, G. E. (2001). The evolution of sexual size dimorphism in the house finch. III. Developmental basis. *Evolution*, 55(1), 176–189.
- Baker, M. C. (2001). *Atoms of language*. New York City: Basic Books.
- Baker, M. C. (2002). Building and merging, not checking: The nonexistence of (Aux) -S-V-O languages. *Linguistic Inquiry*, 33(2), 321–328.
- Balaban, E. (1988). Bird song syntax: Learned intraspecific variation is meaningful. *Proceedings of the National Academy of Sciences*, 85(10), 3657–3660.
- Ball, M. J., Howard, S. J., & Miller, K. (2018). Revisions to the extIPA chart. *Journal of the International Phonetic Association*, 48(2), 155–164.
- Banta Lavenex, P. A. (1999). Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation. *The Journal of the Acoustical Society of America*, 106(1), 491–505
- Banta Lavenex, P. A. (2000). Lesions in the budgerigar vocal control nucleus NLC affect production, but not memory, of English words and natural vocalizations. *Journal of Comparative Neurology*, 421(4), 437–460.
- Baptista, L. F. (1972). Wild house finch sings white-crowned sparrow song. *Zeitschrift Für Tierpsychologie*, 30(3), 266–270.
- Baptista, L. F., & Morton, M. L. (1981). Interspecific song acquisition by a white-crowned sparrow. *The Auk*, 98(2), 383–385.
- Baptista, L. F., & Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, 32, 172–181.
- Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., & Cracraft, J. (2004). Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences*, 101(30), 11040–11045.
- Barker, M. A. R. (1963). *Klamath Dictionary* (Vol. 31). Berkeley and Los Angeles: University of California Press.
- Barker, M. A. R. (1964). *Klamath Grammar* (Vol. 32). Berkeley and Los Angeles: University

- of California Press.
- Barrington, D. (1773). Experiments and observations on the singing of birds. *Philosophical Transactions*, 63, 249–291.
- Bartsch, C., Hultsch, H., Scharff, C., & Kipper, S. (2016). What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. *Journal of Ornithology*, 157(1), 49–60.
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Beckers, G. J. L., Nelson, B. S., & Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Current Biology*, 14(17), 1592–1597.
- Beckman, M. E., De Jong, K., Jun, S., & Lee, S. (1992). The interaction of coarticulation and prosody in sound change. *Language and Speech*, 35(1–2), 45–58.
- Bell, A. (1978). Syllabic consonants. In J. H. Greenberg (Ed.), *Universals of human language 2, Phonology* (Vol. 2, pp. 153–201). Stanford: Stanford University Press.
- Benajmini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Berent, I. (2013). *The Phonological Mind*. Cambridge: Cambridge University Press.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition*, 104(3), 591–630.
- Berke, G. S., & Long, J. L. (2010). Functions of the larynx and production of sounds. In S. Brudzynski (Ed.), *Handbook of mammalian vocalization* (pp. 419–426). Amsterdam: Academic Press.
- Berwick, R. C., & Chomsky, N. (2017). Why only us : Recent questions and answers. *Journal of Neurolinguistics*, 43, 166–177.
- 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.
- Bethin, C. Y. (1998). *Slavic prosody: Language change and phonological theory*. Cambridge Studies in Linguistics (No. 86). Cambridge: Cambridge University Press.
- Bitterbaum, E., & Baptista, L. F. (1979). Geographical Variation in Songs of California House Finches. *The Auk*, 96(3), 462–474.
- Blevins, J. (1993). Klamath Laryngeal Phonology. *International Journal of American Linguistics*, 59(3), 237–279.
- Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Oxford/Cambridge, MA: Blackwell Publishing.
- Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Blevins, J. (2006a). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, 32(2), 117–166.
- Blevins, J. (2006b). Syllable typology. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd ed., pp. 333–337). Oxford: Elsevier.
- Blevins, J. (2007). The importance of typology in explaining recurrent sound patterns. *Linguistic Typology*, 11(1), 107–113.
- Blevins, J. (2008). Consonant epenthesis: natural and unnatural histories. In J. Good (Ed.), *Language universals and language change* (pp. 79–107). Oxford: Oxford University Press.
- Blevins, J. (2009). Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics*, 48(1), 264–273.
- Blevins, J. (2015). Evolutionary Phonology: A holistic approach to sound change typology. In P. Honeybone & J. Salmons (Eds.), *Handbook of historical phonology* (pp. 485–500).

- Oxford: Oxford University Press.
- Blevins, J. (2017). What are grammars made of? In B. D. Samuels (Ed.), *Beyond markedness in formal phonology* (pp. 47–68). Amsterdam/Philadelphia: John Benjamins.
- Blevins, J. and Garrett, A. (1998). The origins of consonant-vowel metathesis. *Language* 74: 508-56.
- Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 117–156). Cambridge: Cambridge University Press.
- Blickley, J. L., Word, K. R., Krakauer, A. H., Phillips, J. L., Sells, S. N., Taff, C. C., ... Patricelli, G. L. (2012). Experimental chronic noise is related to elevated fecal corticosteroid metabolites in lekking male greater sage-grouse (*Centrocercus urophasianus*). *PLoS ONE*, 7(11).
- Blust, R. (2007). Disyllabic attractors and Austronesian sound. *Phonology*, 24(1), 1–36.
- Blust, R. (2013). *The Austronesian languages* (Revised). Canberra: Research School of Pacific and Asian Studies.
- Blust, R., & Trussel, S. (2013). The Austronesian comparative dictionary: A work in progress. *Oceanic Linguistics*, 52(2), 493–523.
- Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer. Retrieved from praat.org
- Böhner, J. (1990). Early acquisition of song in the zebra finch, *Taeniopygia guttata*. *Animal Behaviour*, 39(2), 369–374.
- Bolhuis, J. J., & Everaert, M. (Eds.). (2013). *Birdsong, speech, and language: Exploring the evolution of mind and brain*. Cambridge, MA: The MIT Press.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135.
- Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K. H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, 7, 1–11.
- Bradbury, J. W., & Balsby, T. J. S. (2016). The functions of vocal learning in parrots. *Behavioral Ecology and Sociobiology*, 70(3), 293–312.
- Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behavior. *Nature Reviews Neuroscience*, 1(1), 1–10.
- Breen, G., & Pensalfini, R. (1999). Arrernte: A language with no syllable onsets. *Linguistic Inquiry*, 30(1), 1–25.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- Brenowitz, E. A., & Beecher, M. D. (2005). Song learning in birds: Diversity and plasticity, opportunities and challenges. *Trends in Neurosciences*, 28(3), 127–132.
- Brittan-Powell, E. F., Dooling, R. J., Larsen, O. N., & Heaton, J. T. (1997). Mechanisms of vocal production in budgerigars (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America*, 101(1), 578–589.
- Brockway, B. F. (1964). Ethological studies of the budgerigar: Reproductive behavior. *Behaviour*, 23, 294–323.
- Brockway, B. F. (1965). Stimulation of ovarian development and egg laying by male courtship vocalization in budgerigars (*Melopsittacus undulatus*). *Animal Behaviour*, 13(4), 575–578.
- Brockway, B. F. (1968). Influences of sex hormones on the loud and soft warbles of male budgerigars. *Animal Behaviour*, 16(3), 5–12.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3–4), 155–180.

- Brumm, H., & Zollinger, A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*.
- Brumm, H., Zollinger, S. A., & Slater, P. J. B. (2009). Developmental stress affects song learning but not song complexity and vocal amplitude in zebra finches. *Behavioral Ecology and Sociobiology*, 63(9), 1387–1395.
- Brush, A. H. (2000). Evolving a protofeather. *American Zoology*, 40, 631–639.
- Budka, M., & Osiejuk, T. S. (2013). Formant frequencies are acoustic cues to caller discrimination and are a weak indicator of the body size of corncrake males. *Ethology*, 119(11), 960–969.
- Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, 7(9), 10506.
- Butskhrikidze, M. (2002). *The consonant phonotactics of Georgian*. Utrecht: LOT.
- Butskhrikidze, M. (2015). The status of /m/ in /m/ + C clusters in Georgian. In *Typological Investigations VII* (pp. 130–147). Tsbili: G. Tsereteli Institute of Oriental Studies.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Byers, B. E., & Kroodsma, D. E. (2009). Female mate choice and songbird song repertoires. *Animal Behaviour*, 77(1), 13–22.
- Campbell, E. W. (2011). Zenzontepec Chatino Aspect Morphology and Zapotecan Verb Classes. *International Journal of American Linguistics*, 77(2), 219–246.
- Campbell, E. W. (2013). The internal diversification and subgrouping of Chatino. *International Journal of American Linguistics*, 79(3), 395–420.
- Campbell, E. W., & Woodbury, A. C. (2010). The comparative tonology of Chatino: A prolegomenon. In *SSILA*.
- Campbell, L. (1997). *American Indian languages: The historical linguistics of Native America*. Oxford: Oxford University Press.
- Campo, J. L., Gil, M. G., & Dávila, S. G. (2005). Effects of specific noise and music stimuli on stress and fear levels of laying hens of several breeds. *Applied Animal Behaviour Science*, 91(1–2), 75–84.
- Carbonell, K. M., & Lotto, A. J. (2014). Speech is not special... again. *Frontiers in Psychology*, 5, 3–6.
- Carr, P. (2008). *A glossary of phonology*. Edinburgh: Edinburgh University Press.
- Casey, R. M., & Gaunt, A. S. (1985). Theoretical models of the avian syrinx. *Journal of Theoretical Biology*, 116(1), 45–64.
- Catchpole, C. K., & Slater, P. J. B. (2008). *Bird song: Biological themes and variations*. Cambridge: Cambridge University Press.
- Catford, J. C. (1977). *Fundamental problems in phonetics*. Indiana University Press.
- Cavalli-Sforza, L. L., Feldman, M. W., Chen, K. H., & Dornbusch, S. M. (1982). Theory and observation in cultural transmission. *Science*, 218(4567), 19–27.
- Chamberlain, S. (2018). rphylopic: Get “Silhouettes” of “Organisms” from “Phylopic.” Retrieved from <https://cran.r-project.org/package=rphylopic>
- Chang, H. Y. (2006). Rethinking the Tsouic subgroup hypothesis: A morphosyntactic perspective. In H. Y. Chang, L. M. Huang, & D. Ho (Eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday* (pp. 565–583). Taipei: Institute of Linguistics, Academia Sinica.
- Chen, Y. (2002). *Tsou phonology: A study of its phonemes, syllable structure and loanwords (Masters thesis)*. National Tsing Hua University, Hsinchu.
- Cherchi, M. (1999). *Georgian*. LINCOM Europa.
- Chitoran, I., & Iskarous, K. (2008). Acoustic evidence for high vowel devoicing in Lezgi. In *8th International Seminar on Speech Production* (pp. 93–96).
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

- Chomsky, N. (1966). Explanatory Models in Linguistics. *Studies in Logic and the Foundations of Mathematics*, 44(C), 528–550
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.
- Chomsky, N. (2007). Of minds and language. *Biolinguistics*, 1, 9–27.
- Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.
- Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, 6(AUG), 1–18.
- Clayton, N. S. (1987). Song learning in cross-fostered zebra finches. A re-examination of the sensitive phase. *Behaviour*, 102(1–2), 67–81.
- Clements, N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech* (pp. 283–333). Cambridge: Cambridge University Press.
- Cohn, A., & Riehl, A. (2008). The internal structure of nasal-stop sequences: Evidence from Austronesian. *Laboratory Phonology*, 1–20.
- Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, 134, 257–269.
- Collins, J. (2016). Commentary: The role of language contact in creating correlations between humidity and tone. *Journal of Language Evolution*, 1(1), 46–52.
- Correia, S. P. C., Dickinson, A., & Clayton, N. (2007). Western scrub-jays anticipate future needs independently of their current motivational state. *Current Biology*, 17(10), 856–861.
- Crino, O. L., Johnson, E. E., Blickley, J. L., Patricelli, G. L., & Breuner, C. W. (2013). Effects of experimentally elevated traffic noise on nestling white-crowned sparrow stress physiology, immune function and life history. *The Journal of Experimental Biology*, 216, 2055–2062.
- Croft, W. C. (2003). *Typology and universals*. *Cambridge Textbooks in Linguistics* (2nd ed.). Cambridge: Cambridge University Press.
- Cruz, E. (2011). *The phonological patterns and orthography of San Juan Quiahije Chatino* (Masters dissertation). The University of Texas at Austin.
- Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, 31(5), 462–476.
- Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, 16(3), 564–570.
- Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, 28(02), 197–234.
- Darwin, C. (1871). *The descent of man and selection in relation to sex* (2nd ed.). London: John Murray.
- Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1).
- Davidson, L., & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL)*.
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38(6), 1199–1211.
- DeLancey, S., Genetti, C., & Rude, N. (1988). Some Sahaptian-Klamath-Tsimshianic lexical sets. In W. Shipley (Ed.), *In Honor of Mary Haas: From the Haas Festival Conference On Native American Linguistics* (pp. 195–224). Walter de Gruyter.
- DeLancey, S., & Golla, V. (1997). Penutian hypothesis: Retrospect and prospect. *International Journal of American Linguistics*, 63(1), 171–202.

- Dell, F. D., & Elmedlaoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics*, 7(2), 105–130.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444.
- Dobkin, D. S. (1979). Functional and evolutionary relationships of vocal copying phenomena in birds. *Zeitschrift Für Tierpsychologie*, 50(2), 348–363.
- Donohue, M. (2016). Commentary: Culture mediates the effects of humidity on language. *Journal of Language Evolution*, 1(1), 57–60.
- Dooling, R. J., & Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Perception & Psychophysics*, 47(6), 568–574.
- Dooling, R. J., Mulligan, J. A., & Miller, J. D. (1971). Auditory sensitivity and song spectrum of the common canary (*Serinus canarius*). *The Journal of the Acoustical Society of America*, 50(2), 700–709.
- Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction. *Perception & Psychophysics*, 46(1), 65–71.
- Dooling, R. J., Park, T. J., Brown, S. D., Okanoya, K., & Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals. *Journal of Comparative Psychology*, 101(4), 367–381.
- Dooling, R. J., & Popper, A. N. (2007). The effects of highway noise on birds. *Sacramento, CA: The California Department of Transportation Division of Environmental Analysis*, 74.
- Dooling, R. J., & Searcy, M. (1980). Early perceptual selectivity in the swamp sparrow. *Developmental Psychobiology*, 13(5), 499–506.
- Dooling, R. J., Soli, S. D., Kline, R. M., Park, T. J., Hue, C., & Bunnell, T. (1987). Perception of synthetic speech sounds by the budgerigar (*Melopsittacus undulatus*). *Bulletin of the Psychonomic Society*, 25(2), 139–142.
- Dooling, R. J., Zoloth, S. R., & Baylis, J. R. (1978). Auditory sensitivity, equal loudness, temporal resolving power, and vocalizations in the house finch (*Carpodacus mexicanus*). *Journal of Comparative and Physiological Psychology*, 92(5), 867–876.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567–631.
- Downing, L. J. (2005). On the ambiguous segmental status of nasals in homorganic NC sequences. In M. van Oostendorp & J. van de Weijer (Eds.), *The internal organization of phonological segments* (pp. 183–216). Berlin: De Gruyter Mouton.
- Dowsett-Lemaire, F. (1979). The imitative range of the song of the marsh warbler *Acrocephalus Palustris*, with special reference to imitations of African birds. *Ibis*, 121, 453–468.
- Du, P., & Troyer, T. W. (2006). A segmentation algorithm for zebra finch song at the note level. *Neurocomputing*, 69(10–12), 1375–1379.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Dusan, S., & Rabiner, L. (2006). On the relation between maximum spectral transition positions and phone boundaries. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 645–648).
- Eberhardt, C., & Baptista, L. F. (1977). Intraspecific and interspecific song mimesis in California song sparrows. *Bird-Banding Journal of Ornithological Investigation*, 48(3), 193–205.
- Eberhardt, L. S. (1994). Oxygen consumption during singing by male carolina wrens (*Thryothorus ludovicianus*). *The Auk*, 111(1), 124–130.

- Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, 89(1), 369–382.
- Edwards, T. (2012). Sensing the rhythms of everyday life: Temporal integration and tactile translation in the Seattle Deaf-Blind community. *Language in Society*, 41(01), 29–71.
- Elemans, C. P. ., Rasmussen, J. H., Herbst, C. T., Düring, D. N., Zollinger, S. A., Brumm, H., ... Švec, J. G. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nature Communications*, 6, 8978.
- Emery, N. J. (2006). Cognitive ornithology: the evolution of avian intelligence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1465), 23–43.
- 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.
- Erwin, D. H., & Davidson, E. H. (2002). The last common bilaterian ancestor. *Development*, 129(13), 3021–3032.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *The Behavioral and Brain Sciences*, 32(5), 429-48; discussion 448-494.
- Everett, C. (2017). Languages in drier climates use fewer vowels. *Frontiers in Psychology*, 8(JUL), 1–15.
- Everett, C., Blasí, D. E., & Roberts, S. G. (2016). Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution*, 1(1), 33–46.
- Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology*, 46(4), 621–646.
- Fant, G. M. (1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Farabaugh, S. M., Brown, E. D., & Dooling, R. J. (1992). Analysis of warble song of the budgerigar: *Melopsittacus undulatus*. *Bioacoustics*, 4(2), 111–130.
- Farabaugh, S. M., Brown, E. D., & Veltman, C. J. (1988). Song sharing in a group-living songbird, the Australian magpie. Part II. vocal sharing between territorial neighbors, within and between geographic regions, and between sexes. *Behaviour*, 104(1), 105–125.
- Farabaugh, S. M., Dent, M. L., & Dooling, R. J. (1998). Hearing and vocalizations of wild-caught Australian budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, 112(I), 74–81.
- Farabaugh, S. M., Linzenbold, A., & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108(1), 81–92.
- Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246), 564–568.
- Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals, 91, 173–190.
- Fischer, J., Wheeler, B. C., & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology*, 25(21), R1028–R1029.
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102(2 Pt 1), 1213–1222.
- Fitch, W. T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of Zoology*, 248, 31–48.
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4(7), 258–267.
- Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition*,

100(1), 173–215.

- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, 24(1), 3–33.
- Fitch, W. T. (2018). What animals can teach us about human language: The phonological continuity hypothesis. *Current Opinion in Behavioral Sciences*, 21, 68–75.
- Fitch, W. T., Boer, B. De, Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2, e1600723.
- Fitch, W. T., & Hauser, M. D. (2003). Unpacking “honesty”: Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, A. N. Popper, & R. R. Fay (Eds.), *Acoustic Communication* (pp. 65–137). New York: Springer-Verlag.
- Fitch, W. T., & Kelley, J. P. (2000). Perception of vocal tract resonances by whooping cranes *Grus americana*. *Ethology*, 106(6), 559–574.
- Fitch, W. T., & Mietchen, D. (2013). Convergence and deep homology in the evolution of spoken language. In J. J. Bolhuis & M. B. H. Everaert (Eds.), *Birdsong, speech, and language: Exploring the evolution of mind and brain2* (pp. 45–62). Cambridge, MA: The MIT Press.
- Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics*, 7, 77–116.
- Fortson, B. (2004). *Indo-European language and culture: An Introduction*. Malden, MA: Blackwell Publishing.
- Forvo Media SL. (2008). Forvo: All the words in the world. Pronounced. Retrieved November 1, 2018, from <https://forvo.com/>
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6), 3728–3740.
- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-bovy, G., ... Winsemius, D. (2011). *An R Companion to Applied Regression* (2nd ed.). Thousand Oaks, CA: Sage.
- Franz, M., & Goller, F. (2002). Respiratory units of motor production and song imitation in the zebra finch. *Journal of Neurobiology*, 51(2), 129–141.
- Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. D., & Koenig, L. L. (2015). Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics*, 52, 35–45.
- Gammon, D. E., & Altizer, C. E. (2011). Northern Mockingbirds produce syntactical patterns of vocal mimicry that reflect taxonomy of imitated species. *Journal of Field Ornithology*, 82(2), 158–164.
- Garamszegi, L. Z., Eens, M., Pavlova, D. Z., Aviles, J. M., & Moller, A. P. (2007). A comparative study of the function of heterospecific vocal mimicry in European passerines. *Behavioral Ecology*, 18(6), 1001–1009.
- Garcia, C. M., Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., & Gil, D. (2009). Strategies of song adaptation to urban noise in the house finch: syllable pitch plasticity or differential syllable use? *Behaviour*, 146(9), 1269–1286.
- Garcia, S. M., Kopuchian, C., Mindlin, G. B., Fuxjager, M. J., Tubaro, P. L., & Goller, F. (2017). Evolution of vocal diversity through morphological adaptation without vocal learning or complex neural control. *Current Biology*, 27(17), 2677–2683.e3.
- Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics*, 45(1), 106–113.
- Gil, D., & Gahr, M. (2002). The honesty of bird song: Multiple constraints for multiple traits. *Trends in Ecology and Evolution*, 17(3), 133–141.
- Giraudeau, M., Nolan, P. M., Black, C. E., Earl, S. R., Hasegawa, M., & McGraw, K. J. (2014). Song characteristics track bill morphology along a gradient of urbanization in house

- finches (*Haemorhous mexicanus*). *Frontiers in Zoology*, 11(83), 1–8.
- Giulivi, S., Whalen, D. H., Goldstein, L. M., Nam, H., & Levitt, A. G. (2011). An Articulatory Phonology account of preferred consonant-vowel combinations. *Language Learning and Development*, 7(3), 202–225.
- Goad, H. (2016). Sonority and the unusual behaviour of /s/. In M. J. Ball & N. Müller (Eds.), *Challenging sonority: Cross-linguistic evidence* (pp. 21–44). Sheffield: Equinox.
- Goller, F., Mallinckrodt, M. J., & Torti, S. D. (2004). Beak gape dynamics, during song in the zebra finch. *Journal of Neurobiology*, 59(3), 289–303.
- Goller, M., & Shizuka, D. (2018). Evolutionary origins of vocal mimicry in songbirds. *Evolution Letters*, 1–10.
- Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature*, 344(6268), 715.
- Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition*, 39(1), 1–50.
- Gordon, M. (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. In *Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Phonetics and Phonological Universals* (Vol. 24, pp. 93–105).
- Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science*, 306, 496–499.
- Gramza, A. F. (1970). Vocal mimicry in captive budgerigars. *Ethology*, 27(8), 971–983.
- Greenberg, J. H. (1962). Is the Vowel—Consonant Dichotomy Universal? *Word*, 18(1–3), 73–81.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (pp. 73–113). Cambridge, MA: MIT Press.
- Greenberg, J. H. (1965). Some generalizations concerning initial and final consonant sequences. *Linguistics*, 3(18), 5–34.
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, 20(4), 291–303.
- Güttinger, H. R. (1985). Consequences of domestication on the song structures in the canary. *Behaviour*, 94(3), 254–278.
- Haesler, S., Rochefort, C., Georgi, B., Licznarski, 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.
- Haesler, S., Wada, K., Nshdejan, A., Morrissey, E. E., Lints, T., Jarvis, E. D., & Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *Journal of Neuroscience*, 24(13), 3164–3175.
- Hale, K. (1964). Classification of Northern Paman Languages, Cape York Peninsula, Australia: A research report. *Oceanic Linguistics*, 3(2), 248–265.
- Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, 23(3), 387–429.
- Hanson, R. (2010). *A grammar of Yine (Piro) (Doctoral dissertation)*. La Trobe University, Bundoora, Victoria.
- Hartley, R. S., & Suthers, R. A. (1989). Airflow and pressure during canary song: direct evidence for mini-breaths. *Journal of Comparative Physiology A*, 165(1), 15–26.
- Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology*. Oxford: Oxford University Press.
- Hausberger, M., Jenkins, P. F., & Keene, J. (1991). Species specificity and mimicry in bird song: are they paradoxes. *Behaviour*, 117(1–2), 53–81.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.

- Hayes, B. (2009). *Introductory phonology*. Malden, MA: Blackwell Publishing.
- Heaton, J. T., & Brauth, S. E. (2000). Effects of lesions of the central nucleus of the anterior archistriatum on contact call and warble song production in the budgerigar (*Melopsittacus undulatus*). *Neurobiology of Learning and Memory*, 73(3), 207–242.
- Heaton, J. T., Farabaugh, S. M., & Brauth, S. E. (1995). Effect of syringeal denervation in the budgerigar (*Melopsittacus undulatus*): the role of the syrinx in call production. *Neurobiology of Learning and Memory*, 64(1), 68–82.
- Henke, E., Kaisse, E., & Wright, R. (2012). Is the Sonority Sequencing Principle an epiphenomenon. In S. G. Parker (Ed.), *The Sonority Controversy* (pp. 65–100). Walter de Gruyter.
- Hernandez, A. M., & MacDougall-Shackleton, S. A. (2004). Effects of early song experience on song preferences and song control and auditory brain regions in female house finches (*Carpodacus mexicanus*). *Journal of Neurobiology*, 59(2), 247–258.
- Hertz-Pannier, L., Chiron, C., Jambaqué, I., Renaux-Kieffer, V., Moortele, P. Van de, Delalande, O., ... Bihan, D. Le. (2002). Late plasticity for language in a child's non-dominant hemisphere. *Brain*, 125(2), 361–372.
- Hile, A. G., & Striedter, G. F. (2000). Call convergence within groups of female budgerigars (*Melopsittacus undulatus*). *Ethology*, 106(12), 1105–1114.
- Hindmarsh, A. M. (1986). The functional significance of vocal mimicry in song. *Behaviour*, 99(1), 87–100.
- Hock, H. H. (1991). *Principles of historical linguistics*. Walter de Gruyter.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3), 88–96.
- Hoeschele, M., & Fitch, W. T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Animal Cognition*, 19(3), 643–654.
- Hoeschele, M., Moscicki, M. K., Otter, K. A., van Oort, H., Fort, K. T., Farrell, T. M., ... Sturdy, C. B. (2010). Dominance signalled in an acoustic ornament. *Animal Behaviour*, 79(3), 657–664.
- Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition*, 16(4), 599–609.
- Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, and Psychophysics*, 74(8), 1742–1760.
- Hoh, J. F. Y. (2010). Laryngeal muscles as highly specialized organs in airway protection, respiration and phonation. In S. Brudzynski (Ed.), *Handbook of mammalian vocalization* (pp. 13–21). Amsterdam: Academic Press.
- Hooper, J. B. (1972). The syllable in phonological theory. *Language*, 48(3), 525.
- Howard, R. D. (1974). The influence of sexual selection and interspecific competition on mockingbird song (*Mimus polyglottos*). *Evolution*, 28(3), 428–438.
- Hultsch, H., & Todt, D. (1989). Memorization and reproduction of songs in nightingales (*Luscinia megarhynchos*): evidence for package formation. *Journal of Comparative Physiology*, 165(2), 197–203.
- Hyman, L. M. (2008). Universals in phonology. *Linguistic Review*, 25(1–2), 83–137.
- Hyman, L. M. (2011). Does Gokana really have no syllables? Or: what's so great about being universal? *Phonology*, 28(01), 55–85.
- Immelmann, K. (1975). Ecological significance of imprinting and early learning. *Annual Review of Ecology and Systematics*, 6(1), 15–37.
- International Phonetic Association. (2018). Full IPA chart. Retrieved October 24, 2018, from <https://www.internationalphoneticassociation.org/content/ipa-chart>
- Isaac, D., & Marler, P. (1963). Ordering of sequences of singing behaviour of mistle thrushes

- in relationship to timing. *Animal Behaviour*, 11(1), 179–188.
- Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, 7(2), 217–259.
- Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton & Co.
- Janik, V. M. (1999). Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. *Animal Behaviour*, 57(1), 133–143.
- Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, 28, 60–65.
- Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26(C), 59–99.
- Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, 60(1), 1–11.
- Jansen, D. A., Cant, M. A., & Manser, M. B. (2013). Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biology*, 10(1), 97.
- Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, 5(1), 5–14.
- Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology*, 148(SUPPL. 1).
- Jarvis, E. D. (2013). Evolution of brain pathways for vocal learning in birds and humans. In J. J. Bolhuis & M. B. H. Everaert (Eds.), *Birdsong, speech, and language: Exploring the evolution of mind and brain* (pp. 63–108). Cambridge, MA: MIT Press.
- Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., ... Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, 6(2), 151–159.
- Jarvis, E. D., Ribeiro, S., da Silva, M. L., Ventura, D., Vielliard, J., & Mello, C. V. (2000). Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature*, 406(6796), 628–632.
- Ju, C. (2015). *Cultural evolution in natural populations: A quantitative bioacoustic analysis (Doctoral dissertation)*. The City University of New York Graduate Center.
- Kassambara, A., & Mundt, F. (2017). Package “factoextra” type package title extract and visualize the results of multivariate data analyses. *R Package Version*, 1(3).
- Keating, P. A., MacEachern, P., Shryock, A., & Dominguez, S. (1994). A manual for phonetic transcription: Segmentation and labeling of words in spontaneous speech. *UCLA Working Papers in Phonetics*, 88, 91–120.
- Kelley, L. A., Coe, R. L., Madden, J. R., & Healy, S. D. (2008). Vocal mimicry in songbirds. *Animal Behaviour*, 76(3), 521–528.
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., ... Zamora-Gutierrez, V. (2014). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Reviews*, 91(1), 13–52.
- King, A. P., & West, M. J. (1983). Epigenesis of cowbird song - A joint endeavour of males and females. *Nature*, 305(5936), 704–706.
- King, J. R. (1972). Variation in the song of the rufous collared sparrow, *Zonotichia capensis*, in Northwestern Argentina. *Zeitschrift Für Tierpsychologie*, 30, 344–373.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3), 419–454.
- Klatt, D. H. (1974). How does a mynah bird imitate human speech? *The Journal of the Acoustical Society of America*, 55(4), 822.
- Klimov, G. A. (1998). *Etymological dictionary of the Kartvelian languages. Trends in Linguistics Documentation*. Berlin, New York: Mouton de Gruyter.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic

- categories. *Science*, 237(4819), 1195–1197.
- Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology*, 28, 80–85.
- Konishi, M. (1963). The role of auditory feedback in the vocal behavior of the domestic fowl. *Zeitschrift Für Tierpsychologie*, 20(3), 349–367.
- Konishi, M. (1964). Effects of deafening on song development in two species of juncos. *The Condor*, 66(2), 85–102.
- Konishi, M. (1965a). Effects of deafening on song development in American robins and black-headed grosbeaks. *Zeitschrift Für Tierpsychologie*, 22(5), 584–599.
- Konishi, M. (1965b). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift Für Tierpsychologie*, 22(7), 770–783.
- Koopman, H. (2014). Recursion restrictions: Where grammars count. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in Cognition* (pp. 17–38). Springer International Publishing.
- Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, 27(1), 23–54.
- Kreitman, R. (2008). *The phonetics and phonology of onset clusters: The case of modern Hebrew (Doctoral dissertation)*. Cornell University.
- Kroodsma, D. E. (1972). Variations in songs of vesper sparrows in Oregon. *The Wilson Bulletin*, 84(2), 173–178.
- Kroodsma, D. E., Hamilton, D., Sánchez, J. E., Byers, B. E., Fandiño-Mariño, H., Stemple, D. W., ... Powell, G. V. N. (2013). Behavioral evidence for song learning in the suboscine bellbirds (*Procnias spp.*; *Cotingidae*). *The Wilson Journal of Ornithology*, 125(1), 1–14.
- Kroodsma, D. E., & Parker, L. D. (1977). Vocal virtuosity in the brown thrasher. *The Auk*, 94(4), 783–785.
- Kroodsma, D. E., & Pickert, R. (1980). Environmentally dependent sensitive periods for avian vocal learning. *Nature*, 288(5790), 477–479.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190(4209), 69 LP-72.
- Kumar, S., & Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. *Nature*, 392(6679), 917–920.
- Kuo, J., & Wang, H. (2006). A minimum boundary error framework for automatic phonetic segmentation. In *Chinese Spoken Language Processing, 5th International Symposium, ISCSLP* (pp. 399–409). Berlin: Springer.
- Kuypers, H. G. J. M. (1958). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *The Journal of Comparative Neurology*, 110(2), 221–255.
- Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology*, 124(1), 92–108.
- Ladd, D. R. (1984). Declination: a review and some hypotheses. *Phonology*, 1(1), 53–74.
- Ladd, D. R. (1988). Declination “reset” and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, 84(2), 530.
- Ladd, D. R. (2016). Commentary: Tone languages and laryngeal precision. *Journal of Language Evolution*, 1(1), 70–72.
- Ladefoged, P., & Johnson, K. (2011). *A Course in Phonetics*. Boston, MA: Wadsworth.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Cambridge, MA: Blackwell Publishers.
- Lahti, D. C., Moseley, D. L., & Podos, J. (2011). A tradeoff between performance and accuracy

- in bird song learning. *Ethology*, 117(9), 802–811.
- Larsen, O. N., & Goller, F. (1999). Role of syringeal vibrations in bird vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, 266(1429), 1609–1615.
- Larsen, O. N., & Goller, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *Journal of Experimental Biology*, 205, 25–35.
- Lemon, R. E. (1975). How birds develop song dialects. *The Condor*, 77(4), 385–406.
- Li, G., Wang, J., Rossiter, S. J., Jones, G., & Zhang, S. (2007). Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE*, 2(9).
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
- Lieberman, P. (1958). *Intonation, perception, and language (Doctoral dissertation)*. Massachusetts Institute of Technology.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4), 839–862.
- Lin, Y. (1998). Syllabic and moraic structures in Piro. *Phonology*, 14(3), 403–436.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(5), 783–783.
- Lindblom, B. (1968). Temporal organization of syllable production. *STL-QPSR*, 9(2–3), 1–5.
- Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. *Language, Speech, and Mind*.
- Lionnet, F. (To appear). Paralinguistic use of clicks in Chad. In B. Sand (Ed.), *The Handbook of Click Languages*. Leiden: Brill.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.
- Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology*, 19(2), 219–251.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Animal models of speech perception phenomena. *Chicago Linguistic Society*, 33, 357–367.
- Lowenstamm, J. (1996). CV as the only syllable type. *Current Trends in Phonology: Models and Methods*, (2), 419–441.
- Lulich, S. (2004). Russian [v]: An acoustic study. *Folia Linguistica*, 38(1–2), 63–85.
- MacDougall-Shackleton, E. A., & MacDougall-Shackleton, S. A. (2001). Cultural and genetic evolution in mountain white-crowned sparrows: Song dialects are associated with population structure. *Evolution*, 55(12), 2568–2575.
- MacDougall-Shackleton, S. A. (1997). Sexual selection and the evolution of song repertoires. *Current Ornithology*, 14, 81–124.
- MacDougall-Shackleton, S. A., & Spencer, K. A. (2012). Developmental stress and birdsong: Current evidence and future directions. *Journal of Ornithology*, 153(SUPPL. 1), 105–117.
- Mack, A. ., & Jones, J. (2003). Low-frequency vocalisations by Cassowaries. *American Ornithological Society*, 120(4), 1062–1068.
- Macmahon, M. K. C. (1986). The International Phonetic Association: the first 100 years. *Journal of the International Phonetic Association*, 16, 30–38.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *The Behavioral and Brain Sciences*, 21(4), 499–511.
- Maddieson, I. (1983). The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics*, 14(3), 285–310.
- Maddieson, I. (1984). *Patterns of sounds. Cambridge studies in speech science and communication*. Cambridge: Cambridge University Press.

- Maddieson, I. (2013a). Absence of common consonants. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <https://wals.info/chapter/18>
- Maddieson, I. (2013b). Presence of uncommon consonants. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <https://wals.info/chapter/19>
- Maddieson, I., & Precoda, K. (1989). Updating UPSID. *The Journal of the Acoustical Society of America*, 86(S1), S19–S19.
- Marler, P. (1970a). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, 71(2), 1–25.
- Marler, P. (1970b). Birdsong and speech development: Could there be parallels? *American Scientist*, 58(6), 669–673.
- Marler, P. (1990). Innate learning preferences: Signals for communication. *Developmental Psychobiology*.
- Marler, P. (1991). Song-learning behavior: The interface with neuroethology. *Trends in Neurosciences*, 14(5), 199–206.
- Marler, P. (1997). Three models of song learning: Evidence from behavior. *Journal of Neurobiology*, 33(5), 501–516.
- Marler, P. (2000). Origins of music and speech: Insights from animals. In S. Brown, B. Merker, & C. Wallin (Eds.), *The Origins of Music* (pp. 31–48). Cambridge: The MIT Press.
- Marler, P., Konishi, M., Lutjen, A., & Waser, M. S. (1973). Effects of continuous noise on avian hearing and vocal development. *Proceedings of the National Academy of Sciences of the United States of America*, 70(5), 1393–1396.
- Marler, P., & Peters, S. (1977). Selective vocal learning in a sparrow. *Science (New York, N.Y.)*, 198(4316), 519–521.
- Marler, P., & Sherman, V. (1983). Song structure without auditory feedback: emendations of the auditory template hypothesis. *The Journal of Neuroscience*, 3(3), 517–531.
- Matteson, E. (1963). *The Piro (Arawak) language (Doctoral dissertation)*. University of California at Berkeley.
- Matteson, E. (1965). Piro phonology. In *The Piro (Arawakan) Language* (pp. 13–37).
- Matteson, E. (1972). Proto-Arawakan. In E. Matteson (Ed.), *Comparative studies in Amerindian languages* (pp. 161–242). The Hague: Mouton.
- McCracken, K. G., & Sheldon, F. H. (1997). Avian vocalizations and phylogenetic signal. *Proceedings of the National Academy of Sciences*, 94, 3833–3836.
- McFarlane, J. M., & Curtis, S. E. (1989). Multiple concurrent stressors in chicks. 3. Effects on plasma corticosterone and the heterophil:lymphocyte ratio. *Poultry Science*, 68(4), 522–527.
- McFarlane, J. M., Curtis, S. E., Shanks, R. D., & Carmer, S. G. (1989). Multiple concurrent stressors in chicks. 1. Effect on weight gain, feed intake, and behavior. *Poultry Science*, 68(4), 501–509.
- McIntosh, J. (2015). *Aspects of phonology and morphology of Teotepac Eastern Chatino (Doctoral dissertation)*. The University of Texas at Austin.
- Mennill, D. J., Badyaev, A. V., Jonart, L. M., & Hill, G. E. (2006). Male house finches with elaborate songs have higher reproductive performance. *Ethology*, 112(2), 174–180.
- Meyer, J. (2008). Typology and acoustic strategies of whistled languages: Phonetic comparison and perceptual cues of whistled vowels. *Journal of the International Phonetic Association*, 38(01).
- Michonneau, F., Brown, J. W., & Winter, D. J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution*, 7(12), 1476–1481.
- Mielke, J. (2004). *The emergence of distinctive features (Doctoral dissertation)*. The Ohio

State University.

- Mielke, J. (2005). Ambivalence and ambiguity in laterals and nasals. *Phonology*, 22(02), 169.
- Miller, B. (2012). Sonority and the Larynx. In S. G. Parker (Ed.), *The Sonority Controversy* (pp. 257–288). De Gruyter Mouton.
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340.
- Mol, C., Chen, A., Kager, R. W. J., & ter Haar, S. M. (2017). Prosody in birdsong: A review and perspective. *Neuroscience and Biobehavioral Reviews*, 81, 167–180.
- Mooney, R. (1999). Sensitive periods and circuits for learned birdsong. *Current Opinion in Neurobiology*, 9(1), 121–127.
- Moran, S. (2016). Commentary: Issues of time, tone, roots and replicability. *Journal of Language Evolution*, 1(1), 73–76.
- Morelli, F. (1999). *The phonotactics and phonology of obstruent clusters in Optimality Theory (Doctoral dissertation)*. Rutgers University.
- Mundinger, P. C. (1975). Song dialects and colonization in the house finch. *Condor*, 77(4), 407–422.
- Mundinger, P. C., & Lahti, D. C. (2014). Quantitative integration of genetic factors in the learning and production of canary song. *Proceedings of the Royal Society B-Biological Sciences*, 281(March), 20132631.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (Eds.), *Approaches to Phonological Complexity* (pp. 299–328). Berlin, New York: Mouton de Gruyter.
- Nelson, D. A., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: birdsong. *Science*, 244(4907), 976–978.
- Nelson, D. A., Whaling, C., & Marler, P. (1996). The capacity for song memorization varies in populations of the same species. *Animal Behaviour*, 52, 379–387.
- Nepveu, D. (1994). *Georgian and Bella Coola: Headless syllables and syllabic obstruents (Masters thesis)*. University of California at Santa Cruz.
- Newmeyer, F. J. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.
- Nies, J. (1986). *Diccionario Piro*. Lima: Instituto Lingüístico de Verano.
- Nolan, P. M., & Hill, G. E. (2004). Female choice for song characteristics in the house finch. *Animal Behaviour*, 67(3), 403–410.
- Nordeen, K. W., & Nordeen, E. J. (1992). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behavioral and Neural Biology*, 57(1), 58–66.
- Nottebohm, F. (1972). The origins of vocal learning. *The American Naturalist*, 106(947), 116–140.
- Nottebohm, F. (1976). Phonation in the Orange-winged Amazon parrot, *Amazona amazonica*. *Journal of Comparative Physiology*, 108(2), 157–170.
- Nottebohm, F. (1977). Asymmetries in neural control of vocalizations in the canary. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), *Lateralization in the nervous system* (pp. 23–44). New York: Academic Press.
- Nottebohm, F., Nottebohm, M. E., & Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and Neural Biology*, 46(3), 445–471.
- Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere. *Nature*.
- Nowicki, S., & Capranica, R. R. (1986). Bilateral syringeal coupling during phonation of a songbird. *The Journal of Neuroscience*, 6, 3595–3610.

- Nowicki, S., Peters, S., & Podos, J. (1998). Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, 38(1), 179–190.
- Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, 28, 48–53.
- O’Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., ... Cirranello, A. L. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science*, 339(6120), 662–667.
- Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag.
- Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology: Vol. 1, Between the grammar and the physics of speech* (pp. 258–275). Cambridge: Cambridge University Press.
- Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.
- Ohala, J. J. (1997). Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL]*, (January 1997), 1–6.
- Ohala, J. J., & Kawasaki-Fukumori, H. (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In S. Eliasson & E. H. Jahr (Eds.), *Language and its ecology: Essays in memory of Einar Haugen* (pp. 343–365). Berlin: Mouton de Gruyter.
- Ohala, J. J., & Kawasaki, H. (1984). Prosodic phonology and phonetics. *Phonology Yearbook*, 1, 113–127.
- Ohms, V. R., Beckers, G. J. L., ten Cate, C., & Suthers, R. A. (2012). Vocal tract articulation revisited: the case of the monk parakeet. *The Journal of Experimental Biology*, 215(Pt 1), 85–92.
- Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Nĕmec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences*, 113(26), 201517131.
- Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Dorothy, L. (1992). “Food” calls produced by adult female Rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour*, 120(3/4), 218–231.
- Palacios, M. G., & Tubaro, P. L. (2000). Does beak size affect acoustic frequencies in woodcreepers? *The Condor*, 102(3), 553.
- Paradis, C., & Prunet, J.-F. (Eds.). (1991). *Phonetics and phonology (Volume 2) The special status of coronals: Internal and external evidence*. San Diego: Academic Press.
- Parker, S. G. (2002). *Quantifying the sonority hierarchy (Doctoral dissertation)*. University of Massachusetts Amherst.
- Parker, S. G. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics* 36: 55-90.
- Parker, S. G. (Ed.). (2012). *The Sonority Controversy. Phonetics and Phonology* (Vol. 18). De Gruyter Mouton.
- Patterson, D. K., & Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: acoustic and articulatory correlates of vowels. *The Journal of the Acoustical Society of America*, 96(2 Pt 1), 634–648.
- Patterson, D. K., & Pepperberg, I. M. (1998). Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *The Journal of the Acoustical Society of America*, 103(4), 2197–2215.
- Payne, R., Payne, L., & Whitesell, S. (1998). Interspecific learning and cultural transmission of song in house finches. *The Wilson Bulletin*, 100(4), 667–670.

- Pepperberg, I. M. (1990). Cognition in an African grey parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology*, 104(1), 41–52.
- Pepperberg, I. M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language*, 115(1), 81–91.
- Peter, B., Stoel-Gammon, C., & Kim, D. (2008). Octave equivalence as an aspect of stimulus-response similarity during nonword and sentence imitations in young children. In *Speech Prosody* (pp. 731–734).
- Peters, S. S., Searcy, W. A., & Marler, P. (1980). Species song discrimination in choice experiments with territorial male swamp and song sparrows. *Animal Behaviour*, 28(2), 393–404.
- Pierrehumbert, J. B. (1995). Prosodic effects on glottal allophones. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology: Voice quality control* (pp. 39–60). San Diego: Singular Press.
- Pierrehumbert, J. B., & Talkin, D. (1992). Lenition of /h/ and glottal stop. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*.
- Pinker, S. (1994). *The language instinct: The new science of language and mind*. Penguin Books. New York: Harper Perennial Modern Classics.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(04), 707–727.
- Pinker, S., & Jackendoff, R. (2009). The reality of a universal language faculty. *Behavioral and Brain Sciences*, 32(5), 465–466.
- Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of Acoustical Society of America*, 55(2), 328–333.
- Podos, J. (1996). Motor constraints on vocal development in a songbird. *Animal Behaviour*, 51, 1061–1070.
- Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, 409(6817), 185–188.
- Podos, J., Moseley, D. L., Goodwin, S. E., McClure, J., Taft, B. N., Strauss, A. V. H., ... Lahti, D. C. (2016). A fine-scale, broadly applicable index of vocal performance: Frequency excursion. *Animal Behaviour*, 116, 203–212.
- Podos, J., Southall, J. A., & Rossi-Santos, M. R. (2004). Vocal mechanics in Darwin's finches: Correlation of beak gape and song frequency. *The Journal of Experimental Biology*, 207, 607–619.
- Poupier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology*, 2(2).
- Pride, L. (1963). Chatino tonal structure. *Anthropological Linguistics*, 5(2), 19–28.
- Pride, L., & Pride, K. (2010). *Diccionario Chatino de la Zona Alta* (2nd ed.). Tlalpan, D.F.: Instituto Lingüístico de Verano,.
- Prince, A., & Smolensky, P. (2002). *Optimality Theory: Constraint interaction in Generative Grammar* (ROA). Oxford, UK: Blackwell Publishing Ltd.
- Proctor, M. (2009). *Gestural characterization of a phonological class: The liquids* (Doctoral dissertation.). Yale University.
- Progovac, L., & Ratliff, M. (2016). Commentary: Beyond tone and climate: broadening the framework. *Journal of Language Evolution*, 1(1), 77–79.
- Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526(7574), 569–573.
- Pytte, C. L. (1997). Song organization of house finches at the edge of an expanding range. *The*

- Condor*, 99, 942–954.
- Pytte, C. L., Ficken, M. S., & Moiseff, A. (2004). Ultrasonic singing by the blue-throated hummingbird: A comparison between production and perception. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 190(8), 665–673.
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Rabat, A. (2007). Extra-auditory effects of noise in laboratory animals: the relationship between noise and sleep. *Journal of the American Association for Laboratory Animal Science : JAALAS*, 46(1), 35–41.
- Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, 445(7130), 919–921.
- Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 63(5), 1050–1056.
- Raphael, L. J., Borden, G. J., & Harris, K. S. (2011). *Speech science primer: Physiology, acoustics, and perception of speech*. (6th ed.). Philadelphia: Lippincott Williams & Wilkins.
- Rasch, J. (2002). *The basic morpho-syntax of Yaitepec Chatino (Doctoral dissertation)*. Rice University.
- Rayfield, D., Apridonidze, S., Broers, L., Chanturia, A., Chkhaidze, L., & Margalitadze, T. (2006). *A comprehensive Georgian-English dictionary*. London: Garnett Press.
- Read, A. F., & Weary, D. M. (1992). The Evolution of Bird Song: Comparative Analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 338(1284), 165–187.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences*, 272(1566), 941–947.
- Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, 28, 66–71.
- Rensch, C. R. (1966). *Comparative Otomanguean phonology (Doctoral dissertation)*. University of Pennsylvania.
- Riede, T., Eliason, C. M., Miller, E. H., Goller, F., & Clarke, J. A. (2016). Coos, booms, and hoots: The evolution of closed-mouth vocal behavior in birds. *Evolution; International Journal of Organic Evolution*, 70(8), 1734–1746.
- Riede, T., Suthers, R. A., Fletcher, N. H., & Blevins, W. E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proceedings of the National Academy of Sciences*, 103(14), 5543–5548.
- Ringe, D. (1999). How hard is it to match CVC-roots? *Transactions of the Philological Society*, 97(2), 213–244.
- Ritter, N. A. (2006). Georgian consonant clusters: The complexity is in the structure, not the melody. *The Linguistic Review*, 23(4), 429–464.
- Robins, R. H., & Waterson, N. (1952). Notes on the phonetics of the Georgian word. *Bulletin of the School of Oriental and African Studies*, 14(1), 55–72.
- Rosen, S. M., & Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 30(2), 156–168.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Rowley, I., & Chapman, G. (1986). Cross-fostering, imprinting and learning in two sympatric

- species of cockatoo. *Behaviour*, 96(1–2), 1–16.
- Russian National Corpus. (2003). Russian National Corpus. Retrieved August 21, 2013, from <http://www.ruscorpora.ru/en/corpora-intro.html>
- Ryals, B. M., Dooling, R. J., Westbrook, E., Dent, M. L., MacKenzie, A., & Larsen, O. N. (1999). Avian species differences in susceptibility to noise exposure. *Hear Res*, 131(1–2), 71–88.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals* (Vol. 1).
- Sapir, E. (1925). Sound patterns in language. *Language*, 1(2), 37–51.
- Scanlan, J. (1999). The functional significance of inter-species acoustic cues in the transformation of Budgerigar (*Melopsittacus Undulatus*) sounds into “speech.” *International Journal of Comparative Psychology*, 12(3), 111–152.
- Scheer, T. (2007). On the status of word-initial clusters in Slavic (and elsewhere). In *Annual Workshop on Formal Approaches to Slavic Linguistics. The Toronto Meeting 2006* (pp. 346–364).
- Schleicher, A. (1877). *A compendium of the comparative grammar of the Indo-European, Sanskrit, Greek and Latin languages*. Trübner & Company.
- Schutz, A. J. (1981). A reanalysis of the Hawaiian vowel system. *Oceanic Linguistics*, 20(1), 1.
- Seddon, N. (2005). Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution*, 59(1), 200–215.
- Senghas, A., & Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12(4), 323–328.
- Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and Receivers in Animal Communication. *Annual Review of Psychology*, 54(1), 145–173.
- Shizuka, D. (2014). Early song discrimination by nestling sparrows in the wild. *Animal Behaviour*, 92, 19–24.
- Shubin, N., Tabin, C., & Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature*, 388(6643), 639–648.
- Shubin, N., Tabin, C., & Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, 457(7231), 818–823.
- Sievers, E. (1876). *Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Druck und Verlag von Beritkopf und Härtel.
- Soha, J., & Marler, P. (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour*, 60(3), 297–306.
- Soma, M., & Garamszegi, L. Z. (2011). Rethinking birdsong evolution: Meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, 22(2), 363–371.
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21(3), 291–311.
- Steriade, D. (1999a). Alternatives to syllable-based accounts of consonantal phonotactics. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of the 1998 Linguistics and Phonetics Conference* (pp. 205–242). Prague: Karolinum Press.
- Steriade, D. (1999b). Phonetics in phonology: the case of laryngeal neutralization. In M. Gordon (Ed.), *UCLA Working Papers in Linguistics* (Vol. 3, pp. 25–146). Los Angeles: Department of Linguistics, University of California.
- Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *Perception in Phonology* (pp. 1–26). New York: Academic Press.
- Stern, T. (1957). Drum and whistle “languages”: An analysis of speech surrogates. *American*

- Anthropologist*, 59(3), 487–506.
- Stoeger, A. S., & Manger, P. (2014). Vocal learning in elephants: Neural bases and adaptive context. *Current Opinion in Neurobiology*, 28, 101–107.
- Stoeger, A. S., Mietchen, D., Oh, S., De Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, 22(22), 2144–2148.
- Studdert-Kennedy, M. (1998). The particulate origins of language generativity: from syllable to gesture. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (pp. 202–221). Cambridge: Cambridge University Press.
- Suárez, J. A. (1973). On Proto-Zapotec phonology. *International Journal of American Linguistics*, 39(4), 236–249.
- Sussex, R., & Cubberley, P. (2006). *The Slavic languages. Handbook of Comparative Syntax*. New York: Cambridge University Press.
- Suthers, R. A. (2001). Peripheral vocal mechanisms in birds: Are songbirds special? *Netherlands Journal of Zoology*, 51(2), 217–242.
- Suthers, R. A., Goller, F., & Hartley, R. S. (1994). Motor dynamics of song production by mimic thrushes. *Journal of Neurobiology*, 25(8), 917–936.
- Suthers, R. A., Rothgerber, J. R., & Jensen, K. K. (2016). Lingual articulation in songbirds. *Journal of Experimental Biology*, 219(4), 491–500.
- Swadesh, M. (1947). The Phonemic Structure of Proto-Zapotec. *International Journal of American Linguistics*, 13(4), 220–230.
- Szekely, T., Catchpole, C. K., Devoogd, A., Marchl, Z., & Devoogd, T. J. (1996). Evolutionary changes in a song control area of the brain (HVC) are associated with evolutionary changes in song repertoire among European Warblers (*Sylviidae*). *Proceedings of the Royal Society B: Biological Sciences*.
- Tarpen, M.-L. (1997). Tsimshianic and Penutian: Problems, methods, results, and implications. *International Journal of American Linguistics*, 63(1), 65–112.
- Taylor, M. (1994). *The interaction of vowel deletion and syllable structure constraints (Doctoral dissertation)*. Simon Fraser University.
- Tchernichovski, O., Nottebohm, F., Ho, C., Pesaran, B., & Mitra, P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59(6), 1167–1176.
- The UCLA Phonetics Lab Archive. (2007). Retrieved from <http://archive.phonetics.ucla.edu/>
- Thompson, N. S., LeDoux, K., & Moody, K. (1994). A system for describing bird song units. *Bioacoustics*, 5(4), 267–279.
- Thompson, W. (1960). Agonistic behavior in the House Finch. Part II: Factors in aggressiveness and sociality. *The Condor*, 62(5), 378–402.
- Thorpe, W. H. (1958). The learning of song patterns by birds, with special reference to the song of the chaffinch *Fringilla coelebs*. *Ibis*, 100(4), 535–570.
- Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, 108(37), 3–8.
- Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, 85(2), 901–906.
- Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.
- Tobin, C., Medina-García, A., Kohn, G. M., & Wright, T. F. (2017). Does audience affect the structure of warble song in budgerigars (*Melopsittacus undulatus*)? *Behavioural Processes*.
- Tracy, T. T., & Baker, M. C. (1999). Geographic variation in syllables of house finch songs. *The Auk*, 116(3), 666–676.
- Tsuchida, S. (1972). The origins of The Tsou phonemes /b/ and /d/. *Journal of the Linguistic Society of Japan*, 62, 24–35.
- Tu, H.-W. (2009). *The structure and perception of budgerigar (Melopsittacus undulatus)*

- warble songs. University of Maryland, College Park.
- Tu, H.-W., & Dooling, R. J. (2012). Perception of warble song in budgerigars (*Melopsittacus undulatus*): Evidence for special processing. *Animal Cognition*, *15*(6), 1151–1159.
- Tu, H.-W., Smith, E. W., & Dooling, R. J. (2011). Acoustic and perceptual categories of vocal elements in the warble song of budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, *125*(4), 420–430.
- Upson, B. W., & Longacre, R. E. (1965). Proto-Chatino phonology. *International Journal of American Linguistics*, *31*(4), 312–322.
- Vaissière, J. (1983). Language-independent prosodic features. In D. R. Ladd & A. Cutler (Eds.), *Prosody: Models and Measurements* (pp. 53–65). Berlin: Springer Verlag.
- Vaissière, J. (1995). Phonetic explanations for cross-linguistic prosodic similarities. *Phonetica*, *52*(3), 123–130.
- van den Berg, J. W. (1957). Subglottic pressures and vibrations of the vocal folds. *Folia Phoniatrica et Logopaedica*, *9*(65), 65–71.
- Vargha-Khadem, F., Gadian, D. G., Copp, A., & Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Nature Reviews Neuroscience*, *6*(2), 131–138.
- Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, *92*(3), 930–933.
- Vaux, B., & Samuels, B. (2005). Laryngeal markedness and aspiration. *Phonology*, *22*(3), 395–436.
- Vernes, S. C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review*, *24*(1), 111–117.
- Villard, S. (2015). *The phonology and morphology of Zacatepec Eastern Chatino (Doctoral dissertation)*. University of Texas at Austin.
- Vogt, H. (1958). Structure phonémique du géorgien. *Norsk Tidsskrift for Sprogvidenskap*, *18*, 5–90.
- Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires - hard vs. soft classification approaches. *PLoS ONE*, *10*(4), 1–16.
- Wagner, B., Mann, D. C., Afroozeh, S., Staubmann, G., & Hoeschele, M. (2019). Octave equivalence perception is not linked to vocal mimicry: Budgerigars fail standardized operant tests for octave equivalence. *Behavior*, *1*, 1–26.
- Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*(3), 247–274.
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.
- West, M. J., & King, A. P. (1988). Female visual displays affect the development of male song in the cowbird. *Nature*, *334*(6179), 244–246.
- West, M. J., Stroud, A. N., & King, A. P. (1983). Mimicry of the human voice by European starlings: The role of social interaction. *The Wilson Bulletin*, *95*(4), 635–640.
- Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166.
- Wetzels, W. L., & Mascaro, J. (2001). The typology of voicing and devoicing. *Language*, *77*(2), 207–244.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Retrieved from <http://ggplot2.org>
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717.

- Wild, J. M. (1997). Functional anatomy of neural pathways contributing to the control of song production in birds. *European Journal of Morphology*, 35(4), 303–325.
- Wild, J. M., Goller, F., & Suthers, R. A. (1998). Inspiratory muscle activity during bird song. *Journal of Neurobiology*, 36, 441–453.
- Williams, H. (2004). Birdsong and singing behavior. *Annals of the New York Academy of Sciences*, 1016, 1–30.
- Williams, H., Crane, L. A., Hale, T. K., Esposito, M. A., & Nottebohm, F. (1992). Right-side dominance for song control in the zebra finch. *Journal of Neurobiology*, 23(8), 1006–1020.
- Wright, R. (1994). Fieldwork studies of targeted languages II. *Working Papers in Phonetics*, (87).
- Wright, R. (1996). *Consonant clusters and cue preservation in Tsou (Doctoral dissertation)*. University of California at Los Angeles.
- Wright, R. (2004). A review of perceptual cues and cue robustness. In B. S. Bronson (Ed.), *Phonetically based phonology*. Cambridge: Cambridge University Press.
- Wyndham, E. (1980). Diurnal cycle, behaviour and social organization of the Budgerigar *Melopsittacus undulatus*. *Emu*, 80, 25–33.
- Yearley, J. (1995). Jer Vowels in Russian. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *University of Massachusetts Occasional Papers in Linguistics 18* (pp. 533–571). Amherst, MA: GSLA.
- Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, 10(10), 442–446.
- Zahavi, A. (1975). Mate selection-A selection for a handicap. *Journal of Theoretical Biology*, 53(1), 205–214.
- Zann, R. (1985). Ontogeny of the Zebra Finch Distance Call: I. Effects of Cross-fostering to Bengalese Finches. *Zeitschrift Für Tierpsychologie*, 68(1), 1–23.
- Zann, R., & Cash, E. (2008). Developmental stress impairs song complexity but not learning accuracy in non-domesticated zebra finches (*Taeniopygia guttata*). *Behavioral Ecology and Sociobiology*, 62(3), 391–400.
- Zeitman, E. (2005). Tsou. In K. A. Adelaar & N. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar*. Psychology Press.
- Zhang, G., Li, C., Li, Q., Li, B., Jarvis, E. D., Gilbert, M. T. P., & Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215), 1311–1321.
- Zollinger, S. A., Riede, T., & Suthers, R. A. (2008). Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. *The Journal of Experimental Biology*, 211, 1978–1991.
- Zollinger, S. A., & Suthers, R. A. (2004). Motor mechanisms of a vocal mimic: Implications for birdsong production. *Proceedings of the Royal Society B: Biological Sciences*, 271(1538), 483–491.