CHARACTERIZING THE STRUCTURE OF INFANTS' EVERYDAY MUSICAL INPUT

by

JENNIFER K. MENDOZA

A DISSERTATION

Presented to the Department of Psychology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2018

DISSERTATION APPROVAL PAGE

Student: Jennifer K. Mendoza

Title: Characterizing the Structure of Infants' Everyday Musical Input

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology by:

Dr. Caitlin Fausey	Co-Chairperson
Dr. Dare Baldwin	Co-Chairperson
Dr. Mike Wehr	Core Member
Dr. Melissa Baese-Berk	Institutional Representative

and

Dr. Sara Hodges	Interim Vice Provost & Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2018.

DISSERTATION ABSTRACT

Jennifer K. Mendoza

Doctor of Philosophy

Department of Psychology

June 2018

Title: Characterizing the Structure of Infants' Everyday Musical Input

Infants acculturate to their soundscape over the first year of life (e.g., Hannon & Trehub, 2005a; Werker & Tees, 1984). This perceptual tuning of early auditory skills requires integrating across experiences that repeat and vary in content and are distributed in time. Music is part of this soundscape, yet little is known about the real-world musical input available to infants as they begin learning sounds, melodies, rhythms, and words. In this dissertation, we collected and analyzed a first-of-its-kind corpus of music identified in day-long audio recordings of 6- to 12-month-old infants and their caregivers in their natural, at-home environments. We characterized the structure of this input in terms of key distributional and temporal properties that shape learning in many domains (e.g., Oakes & Spalding, 1997; Roy et al., 2015; Vlach et al., 2008; Weisleder & Fernald, 2013). This everyday sensory input serves as the data available for infants to aggregate in order to build knowledge about music. We discovered that infants encountered nearly an hour of cumulative music per day distributed across multiple instances. Infants encountered many different tunes and voices in their daily music. Within this diverse range, infants encountered consistency, such that some tunes and voices were more available than others in infants' everyday musical input. The proportion of music

produced by live voices varied widely across infants. As infants progressed in time through their days, they encountered many music instances close together in time as well as some music instances separated by much longer lulls. This bursty temporal pattern also characterized how infants encountered instances of their top tune and their top voice – the specific tune and specific voice that occurred for the longest cumulative duration in each infant's day. Finally, infants encountered many pairs of consecutive music bouts with repeated content – the same unique tune or the same unique voice. Taken together, we discovered that infants' everyday musical input was more consistent than random in both content and time across infants' days at home. These findings have potential to inform theory and future research examining how the nature of early music experience shapes infants' early learning.

CURRICULUM VITAE

NAME OF AUTHOR: Jennifer K. Mendoza

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
The Colorado College, Colorado Springs, CO
American University Center of Provence, Aix-en-Provence, France
Grinnell College, Grinnell, IA

DEGREES AWARDED:

Doctor of Philosophy, Psychology, 2018, University of Oregon
Master of Science, Psychology, 2012, University of Oregon
Bachelor of Arts, Psychology, 2007, The Colorado College

AREAS OF SPECIAL INTEREST:

Infant development
Cognitive development
Music perception and cognition

PROFESSIONAL EXPERIENCE:

Graduate Employee, UO Learning Lab, U. of Oregon, 1 year

Dixon Fellow, Co-op Family Center and Parenting Now!,
U. of Oregon, 1 year

ECE Class Teacher, Eugene Suzuki Music Academy, 5 years

Graduate Research Fellow, Team Duckling, Department of Psychology,
U. of Oregon, 5 years

Graduate Teaching Fellow, Department of Psychology, U. of Oregon, 2 years

Graduate Research Fellow, Acquiring Minds Lab, University of Oregon, 1 year

Research Assistant, Intramural Research Training Award (IRTA) Fellowship,
National Institutes of Health, 2 years

GRANTS, AWARDS, AND HONORS:

Scientific Research Grant, The quantity, quality, and stability of everyday music in infancy, GRAMMY Foundation, 2017

GEC Travel Award, Dept. of Psychology, U. of Oregon, 2017

Julie and Rocky Dixon Student Innovation Award, Graduate School, U. of Oregon, 2016

Junior Scholar Travel Award, International Congress for Infant Studies, 2016

GEC Travel Award, Dept. of Psychology, U. of Oregon, 2016

Travel Award, Women in Graduate Sciences, 2015

Marthe E. Smith Memorial Science Scholarship, College of Arts and Sciences, U. of Oregon, 2013

GEC Research Award, Dept. of Psychology, U. of Oregon, 2013

Travel Award, Women in Graduate Sciences, 2012

GEC Travel Award, Dept. of Psychology, U. of Oregon, 2012

Promising Scholar Award, University of Oregon Graduate School, 2011

Honorable Mention, NSF Graduate Research Fellowship Program, 2010

PUBLICATIONS:

Mendoza, J. K. & Baldwin, D. (2014). Shining light on infants' discovery of structure. In J. B. Benson (Ed.) *Advances in Child Development and Behavior, 46,* (pp. 113-148). UK: Academic Press.

Rich, B. A., Carver, F. W., Holroyd, T., Rosen, H. R., **Mendoza, J. K.,** Cornwell, B. R., … Leibenluft, E. (2011). Different neural pathways to negative affect in youth with pediatric bipolar disorder and severe mood dysregulation. *Journal of Psychiatric Research, 45,* 1283-1294.

Rich, B. A., Holroyd, T., Carver, F. W., Onelio, L. M., **Mendoza, J. K.,** Cornwell, B. R., … Leibenluft, E. (2010). A preliminary study of the neural mechanisms of frustration in pediatric bipolar disorder using magnetoencephalography. *Depression and Anxiety, 27(3),* 276-286.

ACKNOWLEDGMENTS

I would like to thank the members of my dissertation committee, Professors Caitlin Fausey, Dare Baldwin, Melissa Baese-Berk, and Mike Wehr, for their help in conducting this research and in preparing this manuscript. I am especially grateful to Professor Caitlin Fausey and Professor Dare Baldwin for their constant support and excellent mentorship throughout my graduate education. I would also like to thank Professor Sandra Trehub for the many insightful conversations about infants and music.

I am grateful to everyone in the University of Oregon Learning Lab. Special thanks go to the many, many research assistants who helped to collect and to code the data. I am also grateful to my colleagues and friends Naomi Aguiar, Jessica Kosie, Caitlin Mahy, Rose Hartman, Alison Shawber Sachet, and Amanda Van Scoyoc for offering their advice, viewpoints, and encouragement over the years.

Most of all, I would like to express immense appreciation to all of the infants and their families who participated in this research.

This research was supported by a Scientific Research Grant from the GRAMMY Museum and by Professor Fausey's University of Oregon faculty funds.

To Dr. Charlotte Mendoza, Dr. Alicia Mendoza, and Dr. Rhea Gendzier
for paving the way.

TABLE OF CONTENTS

Chapter Page

I. A BROAD OVERVIEW 1

The Natural Environments of Infants 1

How Do Infants Build Knowledge in the Domain of Music? 3

What is the Nature of Infants' Everyday Musical Input? 6

Capturing Infants' Everyday Musical Input 7

Characterizing the Content of Infants' Everyday Musical Input 8

Characterizing the Temporal Dynamics of Infants' Everyday Musical Input Across a Day 9

II. A CORPUS OF RAW AUDITORY MUSICAL INPUT AVAILABLE IN INFANTS' EVERYDAY ENVIRONMENTS 11

Introduction 11

Surveys of Infant Caregivers Provided Initial Insight 11

Small-Scale Studies Provided a Glimpse of Infant-Available Auditory Musical Input 13

The Potential to Capture Everyday Musical Input in Day-Long Audio Recordings 14

Collecting a Corpus of Infants' Everyday Musical Input 16

Current Research 17

Method 19

Ethics Statement 20

Open Science 20

Participants 20

Materials 22

Chapter | Page

Procedure 24

Data Processing 25

Identifying Music: Coding, Training, and Processing Procedures 26

Coding Music 25

Training Procedure for Identifying Music 27

Processing Music Data 30

Identifying Features, Voices, and Tunes: Coding, Training, and Processing Procedures 31

Coding Features, Voices, and Tunes 31

Training Procedure for Coding Features, Voices, and Tunes 34

Processing Features, Voices, and Tunes Data 37

Results 37

We Collected Day-Long Recordings from Each of 35 Infants 37

Human Coders Reliability Identified Music in Day-Long Audio Recordings 38

Human Coders Reliability Identified Features, Voices, and Tunes 39

A Corpus of Everyday, Infant-Available Music 42

An Overview of the Key Details about the Music in the Corpus 43

Goof Bouts 44

Features 44

Two Subsets of the Music Corpus: Tunes Music Corpus and Voices Music Corpus 45

Music Bouts with Exactly One Tune or Exactly One Voice 45

Live-Vocal Tunes and Live Voices 48

Chapter Page

Preliminary Checks for Covariance among Measures 49

Discussion 50

III. CONSISTENCY, DIVERSITY, AND SOCIAL QUALITY OF TUNES AND VOICES IN INFANTS' EVERYDAY MUSIC 52

Introduction 52

Consistency and Diversity Shape Learning in the Laboratory 54

Consistency and Diversity Shape Real-World Language Learning 55

Social Quality Shapes Infants' Language Learning 56

Current Research 57

Consistency and Diversity of Tunes and Voices 59

Social Quality 63

Method 67

Data Analysis 67

Tunes Music Corpus and Voices Music Corpus 67

Frequency Distributions 67

Quantifying Diversity 69

Quantifying Consistency 70

Chi-Square Goodness-of-Fit Tests 70

Rank-One Proportion 71

Balanced-Null Proportion 71

Rank-One Consistency Bias 72

Other-Top-Ranked Proportion 72

Other-Top-Ranked Consistency Bias 73

Chapter Page

Proportion of Live-Vocal Musical Input 74

Consistency and Diversity of Live-Vocal Tunes and Live Voices 74

Live-Vocal Music in the Mix of All Musical Input 74

Results 75

Tunes Music Corpus 75

How Many Tune Types Occurred in How Many Seconds of Music? 76

Did Infants Encounter Balanced Frequency Distributions of Tune Types? 77

To What Extent was One Specific Tune More Available Relative to a Balanced Null? 78

To What Extent were Some Tune Types More Available than Other Tune Types? 80

Voices Music Corpus 81

How Many Voice Types Occurred in How Many Seconds of Music? 81

Did Infants Encounter Balanced Frequency Distributions of Voice Types? 82

To What Extent was One Specific Voice More Available Relative to a Balanced Null? 83

To What Extent were Some Voice Types More Available than Other Voice Types? 85

Live-Vocal Tunes 86

What Proportion of Infants' Everyday Music was Live-Vocal? 87

How Many Live-Vocal Tune Types Occurred in How Many Seconds of Music? 87

Chapter Page

To What Extent were Some Live-Vocal Tune Types More Available than Others? 88

How Did Live-Vocal Tune Types Fit in the Mix of All Tune Types? 92

Live Voices 93

What Proportion of Infants' Everyday Music was Live? 93

How Many Live Voice Types Occurred in How Many Seconds of Music? 94

To What Extent were Some Live Voice Types More Available than Others? 95

How Did Live Voice Types Fit in the Mix of All Voice Types? 100

Discussion 101

What Does It Mean to Encounter Consistency Biases in Everyday Musical Input? 102

Outstanding Questions and Future Directions 103

Challenges and Limitations 106

IV. THE TEMPORAL DYNAMICS OF INFANTS' EVERYDAY MUSICAL INPUT ACROSS A DAY 110

Introduction 110

Bursty Temporal Dynamics of Natural Human Behaviors 112

Attention and Memory Matter for Learning 113

Temporal Spacing of To-Be-Learned Content Shapes Learning 114

Sequential Order of To-Be-Learned Content Shapes Learning 118

Current Research 121

Temporal Structure of Music Bouts 122

Chapter Page

Temporal Structure of Top Tunes and Top Voices 127

Top Tune and Top Voice Instance Pairs 129

Sequential Order of Specific Tunes and Specific Voices 129

Method 131

Data Analysis 131

Music Bouts in Two Corpora 131

Burstiness and Inter-Onset Intervals 131

Simulated Bounds of the "Random" Category for Burstiness 133

Gap Durations 134

Durations of Music Bouts, Top-Tune Instances, and Top-Voice Instances 135

Rank-One Temporal Pairs 136

Switch Rate 137

Shuffled Streams 138

Results 139

Tunes Music Corpus 139

Did Music Bouts Occur in Regular, Random, or Bursty Temporal Patterns across Infants' Days? 140

How Long were the Gaps between Music Bouts in the Tunes Music Corpus? 142

How Long were the Individual Music Bouts in the Tunes Music Corpus? 144

A Day of Bursty Music Bouts in the Life of One Baby 145

Did Infants' Top Tunes Occur in Regular, Random, or Bursty Temporal Patterns? 148

Chapter Page

How Long were the Gaps between Instances of Infants' Top Tunes? 150

How Often Did the Top Tune Occur Paired in Sequential Order with Other Tunes? 152

How Often Did Tune Types Switch Over the Course of Infants' Days? 157

Voices Music Corpus 160

Did Music Bouts Occur in Regular, Random, or Bursty Temporal Patterns across Infants' Days? 160

How Long were the Gaps between Music Bouts in the Voices Music Corpus? 162

How Long were the Individual Music Bouts in the Voices Music Corpus? 165

Did Infants' Top Voices Occur in Regular, Random, or Bursty Temporal Patterns across Infants' Days? 166

How Long were the Gaps between Instances of Infants' Top Voices? 168

How Often Did the Top Voice Occur Paired in Sequential Order with Other Voices? 170

How Often Did Voice Types Switch Over the Course of Infants' Days? 174

Discussion 176

Integration Demands in Infants' Everyday Musical Input 177

Unresolved Plausibility of the Perceptual Anchor Hypothesis 179

Outstanding Questions and Future Directions 180

What Do Different Burstiness Values Mean? 180

What is the Nature of Rank-One Repetitions? 182

Chapter Page

V. GENERAL DISCUSSION 185

A First-of-its-Kind Corpus of Infant-Available Music 186

What is the Pathway from Frequent to Familiar? 189

A Multi-Dimensional View of Music in Infants' Everyday Environments 193

Many Paths from Input to Learning in the Music Domain 197

Limits on Generalization 199

Conclusion 200

REFERENCES CITED 201

LIST OF FIGURES

Figure Page

2.1. Infants comfortably wore the LENA vests at home, and LENA recorder was easy to use 23

2.2. Inter-rater reliability for identifying music 39

2.3. Inter-rater reliability for identifying voices 41

2.4. All infants encountered a non-zero duration of music across a non-zero number of separate music bouts 42

2.5. The proportion of coded seconds that was identified as music 43

2.6. Half of music bouts and half of cumulative seconds of music contained exactly one tune 47

2.7. Number of music bouts and the cumulative duration of music that contained exactly one voice 48

3.1. A hypothetical frequency distribution of tune types 69

3.2. All infants encountered some diversity of tune types 76

3.3. The cumulative duration of music with exactly one tune 77

3.4. No infant encountered a balanced distribution of tune types 78

3.5. Consistency bias in the tunes infants encountered 79

3.6. Main measures of consistency were sensitive to the number of unique tunes 80

3.7. Almost all infants encountered more than one voice 82

3.8. The cumulative duration of music with exactly one voice 82

3.9. No infant encountered a balanced distribution of voice types 83

3.10. Compared to balanced null, infants' single top-most voice and other voices in the top of their frequency distributions were much more available 84

3.11. Main measures of consistency were each sensitive to the number of unique voices 85

Figure Page

3.12. Each infant encountered a non-zero proportion of music with live-vocal tunes 87

3.13. All infants encountered some diversity of live-vocal tune types 88

3.14. Cumulative duration of music with exactly one live-vocal tune 88

3.15. No infant encountered a balanced distribution of live-vocal tune types 89

3.16. Consistency bias in the live-vocal tunes infants encountered 90

3.17. Rank-one proportion, balanced-null proportion, and rank-one consistency bias were each sensitive to the number of unique live-vocal tune types 91

3.18. Some infants encountered live-vocal tunes in the top of their frequency distributions of all tune types 93

3.19. Every infant encountered a non-zero proportion of music containing a live voice 94

3.20. All infants encountered some diversity of live voice 95

3.21. The cumulative duration of music with exactly one live voice 95

3.22. Many infants encountered biased distributions of live voice types 96

3.23. Consistency bias in the live voices infants encountered 98

3.24. Rank-one proportion, balanced-null proportion, and rank-one consistency bias were each sensitive to the cumulative number of unique live voices 98

3.25. Live voices were commonly at the top of the frequency distributions of all voice types 100

3.26. Musical notation for the first two phrases of Twinkle Twinkle Little Star and The Alphabet Song 108

4.1. A schematic depicting 3 possible temporal structures of how music bouts may have occurred in time 124

4.2. A schematic depicting one example sequence of tune instances 137

Figure Page

4.3. Infants encountered music bouts in the tunes music corpus in a bursty temporal pattern 141

4.4. Most gaps between music bouts were very short in duration, while a some were considerably longer 142

4.5. Individual infants each encountered music-bout gaps with a mix of durations 144

4.6. Individual infants encountered music bouts of mainly short durations 145

4.7. Music bouts in one example infant’s day 147

4.8. Infants encountered instances of their top tune in a bursty temporal pattern 149

4.9. Most gaps between rank-one tune instances were very short in duration, while a some were considerably longer 150

4.10. Individual infants each encountered rank-one tune gaps with a mix of durations 152

4.11. Infants mostly encountered pairs of tune instances that did not involve their rank-one tune 153

4.12. Tune instances as they occurred in order in each infants’ day 155

4.13. More repetition of tune types in real tune data compared to shuffled streams 158

4.14. Infants encountered music bouts in the voices music corpus in a bursty temporal pattern 161

4.15. Many gaps between voices music bouts were short in duration, while some were much longer 163

4.16. Individual infants each encountered voices-music-bout gaps with a mix of durations 165

4.17. Infants encountered music bouts of short durations 166

4.18. Infants encountered instances of their top voices in a bursty temporal pattern 167

Figure Page

4.19. Many gaps between instances of infants' top voice were short in duration, and many were considerably longer 168

4.20. Individual infants each encountered rank-one voice gaps with a mix of durations 170

4.21. Infants mostly encountered their rank-one voice paired in sequential order with itself 171

4.22. Voice instances in order in each infant's day 173

4.23. Compared to the shuffled streams, most infants encountered a lower switch rate of voices 175

5.1. Infants encountered consistency in their musical input in multiple dimensions that tracked together 192

5.2. Cumulative duration of music, number of music bouts, number of voices, and switch rate of voices 195

5.3. Rate at which voice types switched from one music bout to the next clearly changed across time in one infant's day 197

LIST OF TABLES

Table Page

2.1. Infants ranging in age from 6 to 12 months each recorded one full day 22

2.2. The data in our corpus of infant-available everyday music contained the necessary forms of music, voices, and tunes for us to answer our main theoretical questions 44

4.1. A higher proportion of music-bout gap durations were short in infants' real data relative to the shuffled data 143

4.2. A higher proportion of rank-one tune gap durations were short in infants' real data relative to the shuffled data 151

4.3. Proportions of rank-one pair types in the tune instance pairs of infants' real data and in the shuffled data 156

4.4. Infants encountered a higher proportion of short durations of gaps between voices music bouts compared to the gaps that occurred in the shuffled data 164

4.5. A higher proportion of rank-one voice gap durations were short in infants' real data than in the shuffled data 169

4.6. Proportion of rank-one pair types that occurred in infants' real voices data and in the shuffled streams 174

CHAPTER I

A BROAD OVERVIEW

Infants acculturate to their soundscape over the first year of life (e.g., Hannon & Trehub, 2005a, 2005b; Werker & Tees, 1984). This perceptual tuning of early auditory skills requires integrating across experiences that vary in content and are distributed in time. Music is part of infants' early soundscape, yet little is known about the real-world musical input available to infants as they begin learning sounds, melodies, rhythms, and words. This raw sensory input serves as the data that infants aggregate to build knowledge about music. In this thesis, we addressed the broad question: What is the nature of the musical input that infants encounter in their real-world auditory environments? We first captured infants' natural ("everyday") musical input, and then we characterized the structure of that input in terms of key distributional and temporal properties that shape early learning in other domains (e.g., Oakes & Spalding, 1997; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Roy, Frank, DeCamp, & Roy, 2015; Vlach, Sandhofer, & Kornell, 2008; Weisleder & Fernald, 2013). In this chapter, we provide a broad overview to motivate the main research questions of this thesis. We examine these questions and report all results in the following chapters.

The natural environments of infants.

The study of infants' natural environments and behaviors is not new to developmental psychology (e.g., Bronfenbrenner, 1977; Gessel, 1940; Gessel, Ilg, Learned, & Ames, 1943; Hart & Risley, 1995; Rozin, 2001), but there has recently been renewed interest and endorsement for this approach in developmental science (e.g., Fausey, Jayaraman, & Smith, 2016; Lee, Cole, Golenia, & Adolph, 2017; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017; Weisleder & Fernald, 2013; for relevant review on infants' natural visual input, see Smith, Jayaraman, Clerkin, & Yu, 2018). For example, in one recent study of infants' motor development, researchers compared infants' walking behaviors during a structured task where infants walked in one direction along a straight path and during a free-play session where infants walked spontaneously around a room with naturalistically varying elevations (Lee, Cole, Golenia, & Adolph, 2017). Compared to the straight-path task, infants took fewer steps at once, took curved paths, and took steps in all directions during the free-play session. In a comparable study in the language domain, researchers demonstrated that the linguistic input available to infants during a 45-minute free-play session in their natural environments contained strikingly more silence and less speech relative to the linguistic input available to infants in a 5-minute structured laboratory setting (Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Likewise, in infants' naturalistic visual environments, a small number of objects dominated visual scenes from the infants' perspective, compared to more visually cluttered visual scenes available to adults (e.g., Smith, Yu & Pereria, 2011). Furthermore, the particular content of the distributions of visual input changed over the course of development (Fausey, Jayaraman, & Smith, 2016). Together, these studies highlight how studying infants in natural and naturalistic environments yielded richer data about the available sensory input and about infants' behaviors compared to traditional structured laboratory-based tasks.

The researchers conducting the above work have warned against drawing conclusions about infant development solely from laboratory-based research that oversimplifies complex, natural phenomenon (Lee, Cole, Golenia, & Adolph, 2017; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017; Weisleder & Fernald, 2013; for relevant review on infants' natural visual input, see Smith, Jayaraman, Clerkin, & Yu, 2018). The highly structured nature of laboratory-based tasks presents infants with problems that do not necessarily occur in the real world, and in response, infants may generate solutions in laboratory settings that they would not necessarily use "in the wild". By studying infants' natural environments, researchers can discover what challenges infants actually face as they integrate across their encountered sensory input to build knowledge across domains. In the current research, we examined the structure of music that occurred in infants' natural environments, to gain insight into the problems that infants solve as they build knowledge in the domain of music.

How do infants build knowledge in the domain of music?

In the domain of music, infants learn to recognize and to preferentially process musical patterns – sequences of pitches and rhythms – that are common to their own culture's musical system (Corrigall & Trainor, 2014; Hannon & Trehub, 2005a, 2005b; Trainor & Trehub, 1992; for review, see Hannon & Trainor, 2007). For example, in one study, infants were presented with rhythmic patterns that were either common or rare in their native culture's music. Rare patterns were common to other cultures' musical systems. Six-month-old infants detected subtle changes to both types of rhythmic patterns (Hannon & Trehub, 2005a), but 12-month-old infants only detected subtle changes in the rhythmic patterns that matched their own culture's patterns (Hannon & Trehub, 2005b). These findings were consistent with perceptual tuning that occurs over the first year in life in other domains, such as language processing (e.g., Kuhl et al., 1992; Werker & Tees, 1984) and face perception (e.g., Pascalis, de Haan, & Nelson, 2002; Scott, Pascalis, & Nelson, 2007; for review on perceptual narrowing in language and face perception, see also Maurer & Werker, 2013).

What underlies this developmental change in how infants process musical sounds? Scholars have assumed that this process of music enculturation occurs as the result of infants tracking regularities in the musical sounds of their environments (e.g., Corrigall & Trainor, 2014; Tillmann, Bharucha, & Bigand, 2000; for review, see Hannon & Trainor, 2007), but this idea has not yet been directly examined. What evidence would confirm that infants' everyday exposure to music shapes their processing of musical sounds? First, infants would need to encounter music in their natural environments. Second, the structure of the music available to infants in their everyday environments would need to reflect the structure of the culture-specific musical system. Third, infants would need to detect regularities available in the structure of the musical sounds they encounter. Finally, infants would need to aggregate across separate encounters with musical sounds to build up knowledge about the music in their own culture.

Some evidence exists that provides initial support for this proposed set of processes. In surveys and interviews, most caregivers have reported that they sing and play music for their infants on a daily basis (Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003; Ilari, 2005; Rideout, Vandewater, & Wartella, 2003; Rideout, 2013; Trehub et al., 1997; Young, 2008; Young, Street, & Davies, 2007).

This means that musical sounds were available in infants' everyday environments. Second, ample laboratory-based research has demonstrated infants' ability to implicitly track regularities in auditory input and that doing so shapes their subsequent auditory processing (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; for recent review, see Aslin & Newport, 2014). There are two missing pieces to this puzzle: what is the specific structure in the available musical sounds that infants encounter in their everyday lives, and how could tracking regularities available in that input shape infants' musical processing in the context of real-world learning?

Limited research has directly examined infants' real-world music experience and its impact on infants' auditory processing. One recent study examined the relationship between auditory neural activity in 2- to 3-year-old children and their informal music experience. Children with higher amounts of informal music experience exhibited enhanced and more mature auditory neural responses during an auditory change detection task compared to children with lower overall informal music experience (Putkinen, Tervaniemi, Huotilainen, 2013). A longitudinal study discovered that the frequency of experiencing musical activities at home at ages 2 and 3 years old was associated with children's academic, prosocial, and emotion skills assessed 2 years later (Williams, Barrett, Welch, Abad, & Broughton, 2015). However, each of these two studies had the same major limitation: the only measure of children's informal music experience was derived from a caregiver-report questionnaire. Because these studies did not directly examine the musical sounds available in young children's everyday environments, they provided only limited insight into how young children's musical input impacted their extra-musical skills. While these studies each took an important step towards better understanding the role of informal music experience on young children's development, they also highlighted the need for more information about the structure of the raw auditory musical input that is available to infants and young children in their everyday lives. Recently, researchers have designed a new parent-report questionnaire to capture information about a wider range of young infants' everyday musical activities compared to the information obtained by prior surveys (Politimou, Stewart, Müllensiefen, & Franco, 2018). This measure is still based on retrospective caregiver-report, which may not be effective for capturing fine-grained data about the quality, frequency, and timing of infants' everyday experiences as they occur at day-long timescales (e.g., Csikszentmihalyi & Lefevre, 1989). Future research could directly examine this by comparing the findings as captured via this new measure with recordings of the music available in infants' everyday environments. This new measure was not yet available at the time we collected data for the current study, so we were not able to address this in the current research.

What is the nature of infants' everyday musical input?

The music that infants encounter in their natural environments serves as the raw data over which they presumably track regularities and build knowledge in the domain of music. Yet, we know very little about the sensory input of music available in infants' everyday experiences. In the current research, we harnessed recent advances in wearable technologies (e.g., Ford, Baer, Xu, Yapanel, & Gray, 2008) to collect audio recordings of infants and caregivers at home in their natural environments, and we identified music that occurred during infants' recorded days. We then analyzed the structure of this input in terms of key distributional and temporal properties that shape learning in other domains (e.g., Oakes & Spalding, 1997; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Roy, Frank, DeCamp, & Roy, 2015; Vlach, Sandhofer, & Kornell, 2008; Weisleder & Fernald, 2013). While we were not the first to express a need for greater detail regarding the everyday experiences of music of infants and young children (Barrett, 2009; Cohen, 2008; Hawes, 1974; Koops, 2014; Lamont, 2008; Moog, 1976; Trehub et al., 1997; Young & Gillen, 2007), we were the first to capture and characterize a sizeable corpus of musical input available in day-long audio recordings of infants in their natural environments. In the subsequent chapters of this thesis, we addressed 3 main research questions in the current thesis: (1) In one full day, how much music did infants encounter and how often did they encounter it?, (2) In what ways did infants encounter consistency, diversity, and social quality in the tunes and voices that occurred in their everyday musical input?, and (3) How was music, as well as the tunes and voices that occurred in music, distributed in time across infants' days?

Capturing infants' everyday musical input. In Chapter 1, we report the details of collecting a corpus of day-long audio recordings of infants and their families at home in their natural environments. We also present information about how we identified music, tunes, and voices that occurred in the audio recordings of infants' days. Finally, we provide an overview of the cumulative duration of music and the number of times per day infants encountered music. These properties constrained the possible contents and temporal patterns of infants' musical input. For example, if infants encountered only a couple of minutes of music once in a day, then that music might have consisted of one voice and one tune that occurred in one instance – perhaps mom sang *Twinkle Twinkle*

Little Star to the infant before a nap. On the other hand, if infants encountered several hours of music distributed across multiple music instances, then this music might have consisted of many voices and many tunes that were available for different durations over the course of infants' days. Thus, differences in how much and how frequently infants encountered music could potentially result in different possibilities for how musical input is organized, ultimately providing different learning opportunities.

Characterizing the content of infants' everyday musical input. In Chapter 2 of this dissertation, we focus on the content of infants' everyday musical input. Specifically, we analyzed the consistency, diversity, and social quality of tunes and voices that occurred in infants' daily music. Existing research about infants' music experiences has suggested that infants encounter caregivers singing playsongs and lullabies (Trehub & Schellenberg, 1995; Trehub & Trainor, 1998). Therefore, we expected tunes and voices to be prominent components of the content of infants' everyday music. Did infants encounter the same tune each time music occurred in a day? Or did they encounter a different unique tune during each music instance in their day? Perhaps one tune occurred for a longer cumulative duration that other tunes. We addressed the same set of questions for both tunes and voices. We were particularly interested in the extent to which one tune (or one voice) was more available than others tunes (voices) within infants' encountered daily music. If infants predominantly encountered one unique tune or one unique voice in a day (i.e., high consistency), then that tune or voice might serve as an anchor to guide infants' processing of other, less frequent, input (e.g., Bortfeld et al., 2005; Kurumada, Meylan, & Frank, 2013; Smith & Slone, 2018; Valian & Coulson, 1988). We also determined the proportion of infants' everyday music that involved a live human voice, using this as a marker of high-social-quality musical input. One possibility was that infants mainly encountered caregivers singing (i.e., high-social-quality musical input). Another possibility was that caregivers relied mostly on technological devices to play recorded music for their infants (i.e., lower-social-quality music). These two possibilities represent endpoints on a spectrum. To characterize where on this spectrum infants' everyday musical input fell, we determined what proportion of infants' musical input was live and vocal. This was important to assess, because high-social-quality input engages infants' attention more so than low-quality input (e.g., Kuhl, Tsao, & Liu, 2003; Nakata & Trehub, 2004).

Characterizing the temporal dynamics of infants' everyday musical input across a day. In Chapter 3, we examined the temporal patterns of how music, tunes, and voices occurred across time in infants' days. For example, an infant could have encountered all of their daily music within a short period of time – maybe dad sang a few tunes right before naptime, for example. Alternatively, an infant could have encountered music spaced out across their days – perhaps, for example, the radio was on in the morning, then dad sang around naptime, then in the afternoon, the infant played with a toy that made music, and in the evening the infant's sister sang and hummed several tunes. It was important to determine how music occurred in time over a day, because different temporal dynamics have different consequences for infants' attention and memory, thus differentially shaping infants' learning (e.g., Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Roy, Frank, DeCamp, & Roy, 2015; Vlach, Sandhofer, & Kornell, 2008).

Taken together, by answering these questions, our research provides an altogether new window on the structure of musical input available in infants' everyday environments. By discovering the detailed nature of infants' everyday musical input, we gain insight into the actual challenges infants faced in their real-world learning as they build knowledge about music.

In Chapter 4, we provide a general discussion of the main findings of this research, of how infants' attention and memory may be engaged by music to shape and support their learning in the domain of music, and of how music might ultimately be used to bolster infants' development.

CHAPTER II

A CORPUS OF RAW AUDITORY MUSICAL INPUT AVAILABLE IN INFANTS' EVERYDAY ENVIRONMENTS

INTRODUCTION

The music that infants encounter in their natural environments serves as the raw data over which they presumably track regularities and build knowledge in the domain of music. Yet, we know very little about the sensory input of music available in infants' everyday experiences. In this section of the dissertation, we addressed the broad question: What musical input is available in infants' natural environments? Specifically, we asked: In one day, how much music did infants encounter and how often did they encounter it?

Surveys of infant caregivers provided initial insight.

Although a small number of studies have directly sampled infants' music experience via diary logs or audio and/or video recordings (e.g., Addessi, 2009; Bergeson & Trehub, 1999, 2002; Eckerdal & Merker; 2009; Koops, 2014; Trehub et al., 1997), the bulk of existing knowledge about infants' musical worlds has emerged from research that relied solely on caregiver report in the form of surveys and/or interviews (e.g., Custodero & Johnson-Green, 2003; Ilari, 2005; Young, 2008). While these studies have provided a glimpse into the everyday music lives of young infants, they were limited in several ways. For one, retrospective report is not optimal for capturing fine-grained data about the quality, frequency, and timing of infants' everyday experiences as they occur at day-long timescales (e.g., Csikszentmihalyi & Lefevre, 1989). For another, the questionnaires and interviews about infants' musical experiences were limited in the range and depth of information they solicited. For example, many of these studies asked how often caregivers sing and/or play recorded music for their infants (e.g., Custodero & Johnson-Green, 2003; de Vries, 2007; Rideout, 2013), and some studies asked caregivers about the musical genres they typically select to sing and/or play for their infants, such as classical, rock, pop, children's, country, jazz and so forth (Custodero & Johnson-Green, 2003; Ilari, 2005; Trehub et al., 1997). Yet, none asked caregivers to report any additional details about the specific musical sounds available to their infants. Which particular tunes do infants experience in a day? How many times do they encounter each tune per day? How similar is any one rendition of a tune to the next rendition of that tune? These questions all probe what structure is available to infants' in their everyday musical input. Ample research in other domains, such as language and category learning, has demonstrated that the structural properties of the available input shape what and how infants, children, and adults learn in those domains (e.g., Carvalho & Goldstone, 2013, 2017; Elio & Anderson, 1984; Goodman, Dale, & Li, 2008; Hart & Risley, 1995; Hirsh-Pasek et al., 2015; Hoff & Naigles, 2002; Mather & Plunkett, 2011; Mathy & Feldman, 2009; Ramírez-Esparza, Garíca-Sierra, & Kuhl, 2014; Roy, Frank, DeCamp, Miller, & Roy, 2015; Weisleder & Fernald, 2013; Zeithamova & Maddox, 2009; for recent reviews, see Ambridge, Kidd, Rowland, & Theakston, 2015; Carvalho & Goldstone, 2015). Therefore, it is important to answer questions about the structure of infants' everyday musical input – this structure likely shapes infants' auditory processing and guides their acquisition of musical knowledge. Such questions are not possible to answer based on currently available caregiver-report data. Recording the actual musical sounds infants encounter in their everyday lives would provide rich data with which to address these more detailed questions.

Small-scale studies provided a glimpse of infant-available auditory musical input.

A small handful of studies has attempted to capture the everyday musical input available to infants (Addessi, 2009; Bergeson & Trehub, 1999, 2002; Eckerdal & Merker; 2009; Koops, 2014). In two laboratory-based studies (Bergeson & Trehub, 1999, 2002), mothers were recorded singing to their infants during two sessions. Mothers displayed striking stability in their song performances, using the same absolute pitches and tempos across these two recording sessions. The extent to which mothers' singing exhibits such consistency outside of the laboratory context, amidst the hustle and bustle of everyday life, remains unknown. Three studies directly examined infants' everyday musical input by capturing videos of families at home (Addessi, 2009; Eckerdal & Merker; 2009; Koops, 2014). These studies showed that everyday music involved singing, listening, and dancing (Koops, 2014), which helped to support infant-caregiver dyad tuning (Addessi, 2009), to promote infants' vocalizations (Addessi 2009), and to engage infants (Eckerdal & Merker, 2009). While these studies were a tremendous advance over research relying solely on caregiver-report, they were still limited in several ways. Each of these studies recorded a small number of families for only a couple of hours total; further, each examined music in targeted settings and activities, such as during diaper changes (Addessi, 2009), in free play (Addessi, 2009; Eckerdal & Merker, 2009), or in the car (Koops, 2014). In one study (Eckerdal & Merker, 2009), caregivers were even informed that the researchers were interested in learning specifically about music. Caregivers could have altered their behavior to include more music and/or different kinds of music than they typically would have (this is less of a concern during longer duration studies in which it is more difficult for caregivers to maintain any special or altered behavior).

Taken together, these studies (Addessi, 2009; Bergeson & Trehub, 1999, 2002; Eckerdal & Merker; 2009; Koops, 2014) offer some tantalizing hints regarding the richness of infants' everyday music experiences. However, given the many limitations of each, the extent to which any of these studies captured naturally occurring music is questionable. Further, these studies failed to provide an extensive, systematic account of the music they captured. Questions about how often infants encounter music and about the detailed nature of those music encounters remain, as yet, unanswered. If we had a corpus of the raw auditory musical input that is available to infants in their everyday lives, then we could gain fundamental insight in to the structure of the available musical sounds – properties that likely matter for infants' auditory learning.

The potential to capture everyday musical input in day-long audio recordings.

Recent advances in wearable technologies have opened new avenues for the study of infants' early soundscapes (e.g., Ford, Baer, Xu, Yapanel, & Gray, 2008; Weisleder & Fernald, 2013; for recent review, see Ganek & Eriks-Brophy, 2017). Researchers have begun to use this technology to capture the everyday musical input available to infants and young children (e.g., Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Dean, 2017). Two separate case studies have provided initial insight into the everyday musical worlds of young infants (Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Costa-Giomi & Sun, 2016). In these studies, researchers recorded the musical sounds available in the everyday environments of one 15-month-old infant (Costa-Giomi & Benetti, 2017) and of 11-month-old twin infants (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016). Each infant wore a portable auditory recording device to capture their everyday auditory environments. The 15-month-old infant recorded a total of about 22 hours on two separate days (Costa-Giomi & Benetti, 2017), while the 11-month-old infants each recorded about 11 hours on the same, single day (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016). The musical experiences of these two families were similar in some ways and quite different in others. For example, in both families, infants encountered singing from a caregiver in less than 30 cumulative minutes during their recorded days (Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Costa-Giomi & Sun, 2016). Recorded music (i.e., music produced by radio, TV, and toys) occurred for over 8 hours in the recorded day of the 11-month-old twins (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016), while the 15-month old infant encountered no recorded music in their recorded days (Costa-Giomi & Benetti, 2017).

Taken together, these studies offered new insights into the nature of music available to infants in their everyday environments. They demonstrated that it is possible to identify and analyze infant-available music in day-long audio recordings of infants' natural home environments. The main limitation of these studies was that they each focused on only one family. Differences between the results of these two studies raise questions about potential individual differences in the musical input available to infants in their everyday lives. Such questions could begin to be addressed if this type of in-depth data about music in infants' everyday environments existed for a larger sample of infants.

Collecting a corpus of infants' everyday musical input.

A corpus of the raw auditory musical input data available to infants in their everyday lives does not yet exist. Is it possible to create such a corpus? What are the challenges? For one, such a corpus requires a set of caregivers willing and able to create audio recordings of themselves with their infants in their natural, at-home environments. Recent advances in lightweight, infant-friendly, easy-to-use recording technology (i.e., the LENA system; Ford, Baer, Xu, Yapanel, & Gray, 2008) can help alleviate this potential challenge. Several recent studies have recruited families who have successfully used the LENA recorder to capture the auditory input available to infants and young children in their everyday lives (e.g., Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Weisleder & Fernald, 2013; for recent review, see Ganek & Eriks-Brophy, 2017). Thus, this technology could be used by multiple families to capture their natural, at-home auditory environment for the purpose of studying the music that is available to infants.

Unlike the language learning studies that rely primarily on the LENA automatic algorithms to generate estimates of important properties of infants' linguistic input (e.g., Gilkerson, Richards, & Topping, 2015; Ramírez-Esparza, Garíca-Sierra, & Kuhl, 2014), to our knowledge, there is not yet an automatic way to accurately identify music in noisy recordings of infants' natural environments. Therefore, generating a corpus of raw auditory musical input would require human coding to identify when music occurs in the recorded data. Likewise, coding the music for specific factors of interest (e.g., live versus recorded music) would also require human coding. This is potentially a major roadblock to creating a corpus of everyday musical input, as human coding is a large investment in time and resources.

Finally, although most caregivers report using music daily with their infants (Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003; Ilari, 2005; Rideout, 2013; Rideout, Vandewater, & Wartella, 2003; Trehub et al., 1997), we do not know how much music might occur on any given day – do infants encounter one 10-second tune or do they encounter a thousand tunes that sum to hours of musical input? While this is one of the many details we hope to discover, it makes it difficult to estimate the amount of music we would likely capture in a day-long recording of infants' auditory environments. These challenges, though substantial, were not insurmountable.

Current Research

In the current research, we created a corpus of the raw auditory musical input available to infants in their everyday lives. The potential of such a corpus for advancing our understanding of how infants build music knowledge far outweighs the potential costs associated with the challenges articulated above. We were encouraged by the many recent studies successfully using the LENA system (e.g., Bergelson & Aslin, 2017; Gilkerson, Richards, & Topping, 2015; Ramírez-Esparza, Garíca-Sierra, & Kuhl, 2014; Soderstrom & Wittebolle, 2013; Weisleder & Fernald, 2013;). We also took inspiration from recent work on infants' everyday visual environment (Fausey, Jayaraman, Smith, 2016; Jayaraman, Fausey, & Smith, 2015; for methodological review, see Smith, Yu, Yoshida, & Fausey, 2015) about how to collect and characterize infant-available input.

In this dissertation, we collected day-long audio recordings of 35 infants and their caregivers while they were at home, in their natural, everyday environments. We identified music in these recordings via manual human coding, and we examined the distributional patterns of the contents of music (i.e., tunes and voices that occurred in the music) and the temporal patterns of how this music and its contents occurred over the course of the day in infants' everyday lives. In this chapter, we addressed the following questions: (1) Is it possible to collect day-long recordings of infants' everyday auditory environments? (2) Is it possible to reliably identify musical sounds in day-long recordings of infants' auditory environments? (3) What is the nature of the resulting corpus of musical sounds that occurred in infants' everyday environments? We first address how we successfully overcame the challenges articulated previously in order to collect and to identify music in the captured data. Then, we answer the question of how much music occurred across all infants' recorded days – the cumulative amount of music and the number of times music occurred in each infant's day. These details were important to discover, because they constrain which subsequent questions are possible to answer about infants' everyday musical input. For example, if most infants encounter quite a bit of music across many music instances during their day, then it will be possible for us to answer questions about the structure of that music, such as: How many unique tunes do infants encounter over the course of a day? How many times did any one specific tune repeat throughout the day? In addition, discovering that most infants encounter a non-zero amount of music would confirm that infants' real-world musical experiences were consistent with caregivers' reports that most infants encounter music on a daily basis (Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003; Ilari, 2005; Rideout, Vandewater, & Wartella, 2003; Rideout, 2013; Trehub et al., 1997; Young, 2008; Young, Street, & Davies, 2007). At the end of this chapter, we provide an overview of the key details of the corpus of the raw auditory musical input available in infants' everyday environments.

METHOD

In this dissertation, we analyzed one day-long recording from each of 35 different families. These data were collected as part of a larger study where the goal was to collect three day-long recordings from 35 different families at home (Trio Corpus; Fausey & Mendoza, in prep). Each recording in the Trio Corpus contained at least 10 hours total of recorded data, of which at least four hours must have been continuously recorded (Xu, Yapanel, & Gray, 2009). These sampling inclusion criteria yielded a dataset in which each recording contained the required amount of recorded data (Xu, Yapanel, & Gray, 2009) for the LENA software to provide an automatic analysis of several key properties of the linguistic input captured (e.g., the number of words spoken by adults in the infants' environments). These automatic analyses were important for the goals of the larger study but are not reported in this thesis. To be included in the Trio Corpus, families had to have three recordings that each met these sampling inclusion criteria. Sometimes families met this three-day criterion, and sometimes they did not. We randomly selected 35 recordings that met the daily quantity sampling inclusion criteria, with the following constraints: a roughly even distribution of days of the week recorded, a roughly even distribution of infants' age in weeks, and a roughly equal distribution of infants' sex (see Table 2.1).

Ethics Statement

The University of Oregon Institutional Review Board approved this research protocol. Caregivers provided informed consent for their infant's participation and for any other children in the home to be recorded.

Open Science

We support and practice open science. We are in the process of making all key components of this research publicly available. The audio recordings (the subset reported on here plus the larger Trio Corpus) will be shared to HomeBank (VanDam et al., 2016), with caregivers' consent. Other key elements of this research will be available on this project's Open Science Framework page (Mendoza & Fausey, 2017), including information given to participating families, questionnaires, instructions and manuals for collecting, pre-processing, coding, and analyzing the data, and the associated R code used for these steps. In this dissertation, we reference supplemental materials that can also be found on this project's OSF page. Upon publication, we will make this project's OSF page publicly available.

Participants

Thirty-five infants ranging in age from 6 to 12 months (20 females; Mean = 38.78 weeks, SD = 6.66 weeks; see Table 2.1) and their caregivers participated in this study. We recruited infants in this age range, because this is a developmental period during which everyday musical experiences appear to be influencing how infants build musical knowledge. Caregivers report that infants in this age range encounter music on a daily basis (e.g., Custodero & Johnson-Green, 2003), yet it is not until about 12 months of age that infants show processing advantages for pitch (Trainor & Trehub, 1992) and rhythmic patterns (Hannon & Trehub, 2005a, 2005b; for review, see Hannon & Trainor, 2007) of their own native musical system. Recording the auditory environments of infants ages 6- to 12-months old will capture the musical input available to infants that presumably shapes this process of perceptual tuning in the domain of music.

Families from the local community were recruited to participate through the University of Oregon Psychology department's developmental database and study-specific advertisements (e.g., flyers, Craigslist). To thank them for their participation, families received a $50 gift certificate to Target™ and a children's book for the infant. Critically, families were told that the study was about the mix of sounds infants hear in their everyday lives (e.g., people talking, radios playing, dogs barking, refrigerators running, cars driving by, etc.). Families were intentionally not informed that this study was about musical sounds nor were families selected with respect to musical expertise.

Each family provided information about their household. Most families (n=32) reported data for two main caregivers; three families provided information for one caregiver. In total, caregivers (n=67) reported their races as American Indian or Alaska Native (n=1), Asian (n=3), Black or African American (n=3), White (n=56) and multiracial (n=4). Across all 35 families, household annual income ranged from less than $5000 (n=1) to over $100,000 (n=7); however, this distribution was skewed such that close to half of the families (n=14) reported an annual household income of $75,000 or more. For education, caregivers (n=67) ranged from having completed less than high school (n=1) to having completed a Ph.D., M.D., or J.D. (n=8). Caregivers reported having completed 1.5 to 26 years of formal education (M = 17.15 years, SD = 4.11 years). The distributions of race, income, and education for the participating families were comparable to those of the local community, according to the U.S. Census (U.S. Census Bureau, 2016). Families were neither recruited nor selected with regard to informal or formal music training. Roughly half (n=32) of all caregivers (n=67) reported having some music experience (i.e., informal experience and/or formal training on voice and/or an instrument).

Table 2.1
Infants (n=35) ranging in age from 6 to 12 months each recorded one full day, ranging from 10-16 hours. The sample was roughly balanced for infants' age and sex and the day of the week recorded.

Baby	Age (weeks)	Sex	Day of the Week	Total Recorded (hours)	Duration Coded (hours)	Baby	Age (weeks)	Sex	Day of the Week	Total Recorded (hours)	Duration Coded (hours)
Baby01	28	M	Monday	16	9.01	Baby19	39	M	Sunday	11.72	8.92
Baby02	30	M	Wednesday	13.13	9.59	Baby20	39	M	Wednesday	13.25	10.26
Baby03	30	F	Sunday	16	2.76	Baby21	39	F	Saturday	16	9.18
Baby04	30	F	Sunday	11.88	11.33	Baby22	40	M	Monday	13.07	6.2
Baby05	31	F	Saturday	11.26	7.7	Baby23	41	M	Thursday	13.19	4.07
Baby06	31	F	Saturday	10.51	7.44	Baby24	41	M	Wednesday	16	5.59
Baby07	32	M	Thursday	11.2	9.27	Baby25	44	F	Friday	11.81	8.55
Baby08	33	M	Wednesday	16	8.73	Baby26	44	M	Thursday	16	11.22
Baby09	33	F	Tuesday	11.75	8.83	Baby27	45	F	Friday	12.21	7.8
Baby10	33	F	Friday	16	5.83	Baby28	45	M	Sunday	13.97	6.32
Baby11	35	F	Monday	12.56	8.07	Baby29	45	F	Thursday	10.75	10.26
Baby12	35	F	Monday	14.07	7.26	Baby30	46	F	Tuesday	11.87	8.05
Baby13	35	M	Saturday	16	4.59	Baby31	48	M	Sunday	13.92	9.86
Baby14	36	F	Thursday	16	8.68	Baby32	48	F	Saturday	11.63	5.79
Baby15	37	F	Thursday	16	6.99	Baby33	48	M	Friday	10.97	10.1
Baby16	37	F	Wednesday	15.9	2.39	Baby34	49	F	Thursday	11.39	5.69
Baby17	38	F	Monday	13.84	6.2	Baby35	53	M	Saturday	10.07	9.07
Baby18	39	F	Tuesday	10.84	8.33						

Materials

To capture infants' everyday auditory environment, we gave each family three lightweight recording devices (digital language processors; DLPs) and three vests from the Language Environment Analysis system (LENA; Ford, Baer, Xu, Yapanel, & Gray, 2008; Gilkerson & Richards, 2008; see Figure 2.1). Each DLP records up to 16 hours of audio data. In addition, caregivers received an information packet containing instructions and a daily dairy log (see Supplemental Materials). In the diary log, caregivers noted where the infant was (e.g., home, car), who was around the infant (e.g., mom, dad, sibling), and whether it was typical day for their family. Caregivers also indicated when they turned the DLP on or off and any times the infant was not directly wearing the DLP (e.g., during naps or baths).

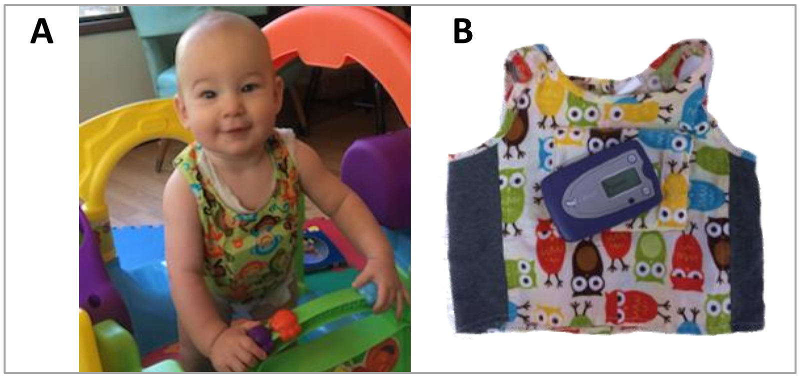

Figure 2.1. (A) Infants comfortably wore the LENA vests at home to record their auditory environments. (B) The LENA recorder was easy to use and fit securely into a pocket on the front of the vest.

We designed three in-house questionnaires to collect information about caregivers' race, income, education and music training and experiences, about infants' race, language exposure, motor development, and music experiences, and about infants' favorite songs, games, toys, videos, TV shows, and books. We also collected a standard measure of infants' receptive and expressive vocabulary (i.e., the MacArthur-Bates Communicative Development Inventory: Words and Gestures; Fenson et al., 1993). We collected this information as part of the larger Trio Corpus study. In the current thesis, we did not make or assess predictions about how differences in these factors might relate to difference in infants' everyday musical input. We have reported basic demographic data related to the families in the *Participants* section above to demonstrate that the current sample of families was representative of the local population.

Procedure

To collect the data, one primary researcher (Mendoza) met with caregivers individually, either in the lab or at the family's home, to provide the study materials (i.e., DLPs, vests, information packet, diary log) and to explain how to use the DLPs. The researcher also reviewed the consent form and provided an opportunity for caregivers to ask questions. The same researcher met with all families included in this dissertation. At this time, caregivers provided informed consent for their infant's participation and for any other children in the home to be recorded.

Caregivers were instructed to have their infant wear the DLP in the vest and to have the DLP on and recording during all hours in the day starting when their infants first woke up in the morning until their infant went to sleep for the night. Caregivers were free to turn the DLP on and off as they saw fit, but they were encouraged to "set it and forget it" (LENA Research Foundation Team, 2014), leaving the DLP on and recording continuously. Caregivers were also instructed to record only inside their own homes. This was not a scientifically motivated decision. Rather, it was the result of a restriction given by the University of Oregon's Institutional Review Board (based on an Oregon State Law) that stated anyone being audio recorded had to be informed of the recording. In order to capture families' natural activities and behaviors, we could not ask caregivers to constantly monitor their environment and alert people they interacted with that their infant was wearing an audio recording device – we wanted caregivers to forget about the recorder and to go about their days as usual. Thus, we asked them only to record at home in order to comply with the restriction while still recording infants' natural environment. When families left their home, they were encouraged to remove the vest but to leave the

DLP on and recording at home. Leaving the DLP on made it easier for families when they returned, because they could just put the vest back on their infant and not worry about whether the device was still recording. Further, caregivers were instructed to remove the vest but to leave the DLP on and recording nearby while infants were napping or bathing – this protected the safety of the infants (the vest and DLP were not designed to be slept in nor submerged in water) while still recording the auditory environment during these activities. Caregivers were asked to indicate in their diary log the times when the infant was not directly wearing the DLP.

A researcher (Mendoza) called the families on each recording day during the week they participated, offering a chance for caregivers to ask questions and for the researcher to provide feedback. At the end of the week, a researcher met with the family to collect the study materials and to administer the questionnaires. At this time, the researcher also provided a more detailed explanation of the nature of the study. If caregivers asked, then we explained that our initial research questions were about infants' musical experiences. Caregivers also had the opportunity to ask questions at this time.

Data pre-processing

All recordings were processed prior to being coded for music. Sounds occurring during times that families indicated as private (e.g., "do not listen") or as being outside of the home while the DLP was still recording were automatically replaced with silence. If a recording had less than 4 hours of data remaining after removing private and outside-home sections, then it was not included for further analysis. We set this criterion based the pre-processing of roughly 100 recordings that were part of the larger Trio Corpus

(Fausey & Mendoza, in prep). We calculated the mean and standard deviation of the duration recorded minus the cumulative duration replaced with silence due to privacy or being outside of the home – 4 hours was approximately 2 standard deviations below this mean. During any period of time when the DLP had been turned off, silence was automatically inserted for the duration the DLP was off, thus preserving the real-time information within the file. Periods of extended silence (e.g., naptimes) were automatically detected and labeled as not for coding (-22dB relative to the full-scale signal for that recording for 3 minutes or longer). Brief sounds (under 3 minutes) that interrupted two otherwise adjacent periods of silence as well as extended periods (at least 10 minutes) of highly regular sound (e.g., white noise machine) were identified by hand and labeled as not for coding. For this pre-processing, we used multiple custom scripts (see Supplemental Materials) to identify and edit portions of the recordings. These processing steps are consistent with prior research (e.g., Weisleder & Fernald, 2013; Bergelson & Aslin, 2017).

Identifying music: Coding, training, and processing procedures

Coding music. Trained coders listened continuously to the codable segments of the audio recordings that were identified during the pre-processing steps, listening for "music". Sounds that were consistent with a "music" judgment included those that involved live singing and/or instrument playing (by caregiver, sibling, etc.), recorded singing and/or instrument playing (from radio and/or other electronic sources, like toys, phones, etc.), and vocally produced pitched, rhythmic, repetitive patterns (e.g., humming, whistling, "vocal play"). Sounds that were consistent with a "no music" judgment included those that were produced by the target infant (as determined by contextual cues), speech (including infant-directed speech and routinized speech, like book reading), infant babbling and/or an imitation of infant babbling, sound effects (e.g., “beep beep”), the sounds of a non-music household object (e.g., computer keyboard) or appliance (e.g., microwave, blender), and sounds produced by a non-human animal (e.g., birdsong).

Upon hearing a musical sound, coders were instructed to mark each instance of a music bout. We defined a bout as the uninterrupted, continuous presence of music. Music bouts were determined independently from the musical content present. For example, a bout could include one, two, or more songs; likewise, a single song could be sung across multiple (interrupted) bouts. When coders heard a music bout begin, they marked the file to indicate the bout’s onset. Then, they created a second marker in the file to indicate the bout’s offset. The end of the music bout could be signaled in one of two ways: 1) the source of the music actually stopped producing musical sounds, or 2) the musical sounds became too faint or too obscured to be perceived. Coders were instructed to ignore musical sounds that were mostly or completely obscured by other sounds or that were too faint to be clearly perceived. Coders were provided with a coding manual that they reviewed in its entirety at the start of every coding session. This manual contained multiple examples from pilot data audio recordings and rationale explaining what should and should not count as “music” as well as what should count as one single music bout versus separate music bouts. These examples served as clear anchors to guide coders’ judgments (see Supplemental Materials). Two independent coders identified music in each of the 35 recordings. We assessed inter-rater reliability using the same procedure described below to assess whether coders passed the training file.

Training procedure for identifying music. To overcome the challenge of manually coding these recordings for music, we developed rigorous coding schemes and training procedures for identifying music and also for subsequent coding of features, voices, and tunes within the music that occurred (described below). All coders first coded at least one training file before coding any actual data. We selected three day-long recordings from pilot data to serve as training files, and an expert coder coded music bouts for all three to create one set of standard codes per file. Coders used ELAN Linguistic Annotator (Version 4.9.4; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) to listen continuously to the training file and to code the start and end of music bouts (see "Coding music" section for details).

We exported the lists of coders' start and end times (in seconds) for each music bout they identified, and then we aggregated these data to examine the number of seconds per minute coded as music for each minute that was coded. We compared this list with that of the standard coder, calculating a Pearson correlation. To be considered a "trained" coder, the correlation coefficient of this analysis had to be at least r = .90.

To the best of our knowledge, no prior study has attempted to assess inter-rater reliability between two human coders' coding across the full length of day-long data. In prior research, many studies (e.g., Belardi et al., 2017; Chang, de Barbaro, & Déak, 2017; Cole, Robinson, & Adolph, 2015; Gilkerson & Richards, 2008; Goldstein, Schwade, & Bornstein, 2009; Hirsh-Pasek et al., 2015; Karasik, Tamis-LeMonda, & Adolph, 2016; Konishi, Stahl, & Golinkoff, 2016; Kretch, & Adolph, 2016; Kretch, Franchak, & Adolph, 2013; Luo & Tamis-LeMonda, 2015, Oller, Buder, Ramsdell, Warlaumont, Chorna, & Bakeman, 2013; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Suskind et al., 2016; Weisleder & Fernald, 2013; Yu & Smith, 2016) have reported inter-rater reliability between two human coders for coding discrete phenomenon (e.g., number of words, speaker identity) in short segments (i.e., 3-30 minutes) for a subset of the data (e.g., 25% of each infants' data). While these studies used different measures of coder agreement depending on the nature of the data (e.g., percent agreement, Cohen's kappa, correlations), generally, these studies found values ranging from the equivalent of about 80% to 99% agreement. As one example from prior research using LENA recordings, human coders made a categorical decision about the identity of a speaker (i.e., adult or not an adult) in three 10-minutes segments from each of 2 day-long recordings (i.e., 1 total hour of data), and the percent agreement between 2 coders was 88%; Gilkerson & Richards, 2008). Other studies using LENA recordings have compared the LENA automatic analyses to that of human coding (e.g., Canault et al., 2015; Caskey, Stephens, Tucker, & Vohr, 2014; Ko, Seidl, Cristia, & Reimche, 2016; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Soderstrom & Wittebolle, 2013; VanDam & Silbert, 2016; Xu, Yapanel, & Gray, 2009; Weisleder & Fernald, 2013). In these studies, researchers also compared subsets of the recorded data (e.g., six 10-minute segments per recording). These studies found values ranging from about 70% to 95% agreement. Therefore, the extent to which we should expect coder agreement in the current research to be comparable to that of prior research was unknown. Nevertheless, our criterion of r = .90 was squarely within the range of inter-rater reliability reported in prior research, and it should therefore indicate data that has been coded reliably.

In the current research, if coders met the r = .90 training criterion, then they were moved into coding the real data. If coders did not meet this training criterion, then they received additional feedback, practice, and further training. Then they coded a second training file. Coders were given a maximum of three training files to code. If they failed to reach the training criterion after the third training file, then they were removed from coding music bouts for this project. In all, 20 unique coders completed the training procedure for coding music bouts. Fourteen coders passed after completing the first training file, three coders passed after completing two training files, and one coder passed after completing all three training files. Two coders failed to reach the training criterion after the third training file. An additional three coders began the training procedure but decided to stop working on this project prior to completing the training.

Processing music data. In order to analyze the patterns in the identified music, we first aligned coders' ELAN coding with time of day. To do this, we expanded codes from ELAN (onset and offset times of musical bouts listed in seconds and milliseconds) into a running list of seconds that started at 0 (midnight) and continued for 129,600 seconds (i.e., a 36-hour period). This list contained a '1' in every second during the day that music was identified. Because the timescale of the ELAN coding was finer grained than seconds, we rounded the ELAN onset and offset times inclusively (i.e., start time down, end time up) to ensure that we fully captured all of the music that was identified by the coders. In this process, if two consecutive music bouts that were separated by less than 1 second as coded in ELAN, then they were merged into one music bout in the "rounded-seconds" list. This happened for 689 (.13) of the original ELAN music bouts across all 35 recordings. As reported informally by coders, the possible sources of very short sounds that interrupted music and resulted in coding two separate music bouts included vocalizations from the infant, rustling of the vest as infants moved, and spoken words interleaved between musical sounds (which seemed especially common in the sounds produced by infants' toys). As a result of this conversion process, the shortest possible duration for a music bout (and also for a pause between music bouts) was 1 second. All subsequent analyses were performed using these rounded-seconds music bout data.

Identifying features, voices, and tunes: Coding, training, and processing procedures

Coding features, voices, and tunes. Trained coders listened to the segments of ELAN files that were previously identified as "music bouts". Some music bouts contained multiple voices and/or multiple tunes. Coders were instructed to listen to the full length of each music bout, coding each voice and/or tune separately. Coders were encouraged to begin listening prior to the start of the music bout to gain contextual information about the voices and tunes.

Coders completed the coding of voices and tunes in multiple passes, entering their coding into a separate Excel file per coding pass. They were provided with a detailed coding manual including multiple examples from pilot auditory recording and rationale explaining how these example music bouts were coded for each coding pass (see Supplemental Materials). For each music bout, coders first judged whether the music in the bout was live and/or recorded. "Live" music bouts contained any musical sound produced by a human who was clearly present in-person in the infant's environment (i.e., human voice or live instrument being played). "Recorded" music bouts included any musical sound clearly produced by any electronic source (e.g., TV, radio, toy, phone, Pandora). Then, in a separate coding pass, coders judged whether each bout contained vocal and/or instrumental music. Any musical sound that was produced by a live or recorded voice – any adult, non-target child, or recorded character (e.g., singing, humming, whistling, vocal play) – was judged as “vocal”. Any musical sound produced by a live or recorded music instrument (e.g., piano, guitar, drum, etc.) was judged as “instrumental”, which also included non-vocal musical sounds from toys and phones.

Next, for the music bouts judged to be “vocal”, coders then identified the specific voice(s) that produced music in the bout. For live voices, coders listed the voice’s relation to the infant (e.g., mom, dad, grandma). For recorded voices, coders listed the specific artist, character, or band (e.g., Taylor Swift, Daniel Tiger, Maroon 5). If coders did not know the specific voice, they could use external resources (e.g., the Internet) to help them identify the voice. However, coders were not allowed to use any song-identifying software (e.g., Shazam) that would require the software to directly access or “listen to” the raw audio recordings (i.e., confidential data). They were also not allowed to consult any human resource – if they were to discuss the data with other coders, then this would have violated the independence of their coding. If they were to discuss the data with anyone outside of the research team, then this would have violated the confidentiality of the families. If after searching, the coders could not determine the specific identity, then they made up a distinctive identity (e.g., female voice 1, squeaky cartoon voice 2). For each voice, coders judged whether the current voice was the same or different as all previously coded unique voices. If it was the same, then coders listed the same specific identity (known or made-up) as when the voice occurred previously (e.g., mom and mom). If it was different, then coders listed (or made up) a unique specific identity (e.g., mom and female voice #3). Critically, if coders encountered repeated instances of the same voice across music bouts, then they listed exactly the same identity for each instance of the same voice.

Finally, for all music bouts, coders identified the title of the tune(s) that occurred. Every music bout was considered to have at least one tune. For standard tunes, coders used the known title (e.g., *Itsy Bitsy Spider*, *Shake It Off*). As with voices, if coders did not know the title, then they used external resources to help identify titles for standard tunes. If coders still could not identify a standard tune's title, then they made up a title for the tune. For invented tunes and adapted tunes (e.g., tune with a standard melody paired with invented lyrics), coders made up short, descriptive titles (e.g., "Everybody loves potatoes", "Short Whistle 4"). As with voices, coders judged whether the current tune was the same or different as all previously coded unique tunes. If it was the same, then coders listed the same specific title (known or made-up) as when the tune occurred previously (e.g., *Itsy Bitsy Spider* and *Itsy Bitsy Spider*). If it was different, then coders listed (or made up) a unique specific identity (e.g., *Itsy Bitsy Spider* and *fast pop song #3*). Critically, if coders encountered repeated instances of the same tune across music bouts, then they listed exactly the same title for each instance of the same tune.

The four passes of this coding scheme enabled us to count the number of times unique voices and unique tunes occurred in repeated instances during infants' days as well as to calculate the cumulative duration of each unique voice and each unique tune that occurred, in order to determine the extent to which some unique voices and some unique tunes were more available than others in infants' musical input. Two independent coders identified features, voices, and tunes in each of the 35 recordings. We assessed inter-rater reliability using the same procedure described below to assess whether coders passed the training file.

Training procedure for coding features, voices, and tunes. From listening to pilot data, we anticipated that many music voices and tunes would come from media sources (e.g., TV shows, children's toys, etc.). Since not all coders had the same knowledge of sounds from media sources likely to be present in the data, coders first increased their own familiarity with common musical sounds likely to occur in everyday environments before they began coding voices and tunes in the data. To do so, coders listened to sound clips taken from currently popular children's TV shows, children' music, and children' toys. Then coders listened to sound clips from currently popular TV shows, various music genres, and common electronic devices targeted for adults (see Supplemental Materials). Coders were clearly informed that the purpose of this familiarization step was to highlight the wide range of sounds that are likely to be common in infants' everyday environments, rather than to provide specific examples of particular voices and tunes that they should listen for as they coded the data. This familiarization also served to remind coders that they would likely hear musical sounds in the data from sources they have not personally encountered before (e.g., children's TV shows they have never seen, music from genres they do not listen to, etc.), and that they should still strive to identify the specific voices and specific tunes therein.

All coders were then instructed to review the coding manual for the original coding of music bouts, reading the information and listening to each of the examples. Some coders had participated in this original coding of music bouts, so this step was a refresher for them. For coders who had not completed any coding of music bouts, this review gave them a better understanding of the kinds of sounds that were and were not coded as music in these data (see Music Coding section for details).

Finally, all coders completed at least one training file. We selected three recordings from pilot data to serve as training files, and an expert coder coded features, voices, and tunes for all three to create one set of standard codes per file. Coders used ELAN to listen to the music bouts in the training file and Microsoft Excel to record their judgments about voices and tunes. Coders were provided with a coding manual that detailed the instructions for each coding pass and included audio examples to guide coders' coding (see Music Coding section for details). The music in each training file contained a broad range of voices and tunes presented in all combinations of features, including many music bouts with multiple voices and/or multiple tunes.

We compared each coder's coding to that of the standard coder, evaluating each coding pass separately (see Music Coding section for full details about coding passes). Prior to this assessment, coders' files were cleaned to remove punctuation, spaces, capital letters, and to fix spelling mistakes. Coders first judged each music bout as "live", "recorded", "vocal", and/or "instrumental", and we examined the proportion of agreement between the standard coder and the coder trainee separately for each of these four features. To be considered a "trained" coder, proportion agreement had to be at least .90 for each feature. Coders next identified specific voices and then specific tunes. We selected Tschuprow's *T* as an index of agreement between the standard coder and the coder trainee, because it assesses contingencies between two sets of nominal data. We needed a measure of contingency, because it is possible for coders to invent their own labels for voice identities and tune titles. For example, imagine a day when an infant heard the song *Hey Jude* by The Beatles six times. Every time this tune occurred, the standard coder labeled the voice identify as "The Beatles" but the coder trainee labeled it "Unknown Band 1". If we were to use proportion agreement or kappa, then this would look like 0% agreement (i.e., the two coders never labeled these tune instances with the same label as each other). However, a contingency table reveals the systematicity in these coders' coding – while they disagreed with each other on the specific identity, they did agree that all six of those bouts should receive the same voice identity label as one another. Tschuprow's T (unlike Cramer's V, another common measure of contingency) allows for differences in the number of unique items in each set (i.e., it is possible to input a rectangular 2-way frequency table). Values of T range from 0 to 1, and T will only ever equal 1 (i.e., complete agreement between two sets of nominal data) if both datasets have the same number of unique items (i.e., a square 2-way frequency table). To be considered a "trained" coder on judging voice identities and tune titles, the Tschuprow's T value had to be at least .90. These criteria of *proportion agreement* = .90 for features coding and T = .90 for voices and tunes coding were consistent with the criterion of r = .90 for identifying music, and all of these criteria were informed by the inter-rater reliability values reported in prior research that we reviewed earlier in this chapter (in the "*Training procedure for identifying music"* section). If coders met these training criteria, then they were moved into coding the real data. If coders did not meet these training criteria, then they received additional feedback, practice, and further training. Then they coded a second training file. Coders were given a maximum of three training files to code. If they failed to reach the training criteria after the third training file, then they were removed from coding voices and tunes for this project. In all, 16 unique coders completed the training process for coding features, voices, and tunes – 13 coders (.81) passed after completing the first training file, one coder (.06) passed after completing two training files, and 2 coders (.13) failed to meet the training criteria after completing three training files. An additional one coder began the training process but decided to leave this project prior to completing it.

Processing features, voices, and tunes data. As we did for the music bouts, we expanded and aligned coders' coding of features, voices, and tunes into the running list of seconds from midnight. This expansion resulted in multiple columns to capture all of the coded data. There was one column per feature (i.e., Live, Recorded, Vocal, Instrumental) and each contained a '1' for each second that contained music that had been coded as that feature. In each case, all of the seconds associated with the entire duration of each music bout were filled in with a '1'. Then the content of the voices and tunes coding (from coders' cleaned excel files) was aligned to the rounded seconds list just as it was for features – the entire content of each bout was the number of seconds that represented the duration of that bout. All subsequent analyses were performed using these data.

RESULTS

We collected day-long recordings from each of 35 infants.

Our first question was whether families could successfully record full days of their infants' at-home environments. The answer was yes, they did. Across all 35 days, caregivers recorded a total of 1,680,351 seconds (467.01 hours) of audio data. Per day, caregivers recorded 47,284 seconds (Median=13.13 hours; SD=2.06 hours; i.e., time when the LENA DLP was turned on). Most caregivers (.77) left the recorder on continuously. After pre-processing, the final dataset had a total of 970,873 seconds (269.69 hours) of codable data. No periods of privacy (e.g., "do not listen") were indicated for this set of recordings. Caregivers indicated being outside of the home in 14 recordings, so we replaced a combined total of 233,123 seconds (64.75 hours; per file: Median=0 hours, SD=3.10 hours) with silence. Filtering low-level sounds removed 428,671 seconds (119.7 hours; per file: Median=2.85 hours, SD=2.66 hours). Manual editing to remove brief sounds and extended periods of highly regular sound resulted in removing an additional 47,684 seconds (13.24 hours; per file: Median=.07 hours, SD=1.31 hours). The drop from total recorded hours was expected due to the normal duration of sleep during infancy (Galland, Taylor, Elder, & Berbison, 2012) and the mix of other activities that occur for infants in this age range. Each day contained 29,052 seconds (Median = 8.07 hours; SD=2.20 hours) of codable data, which is comparable to prior results using LENA recordings (e.g., Weisleder & Fernald, 2013). Table 2.1 summarizes key information about each infants' recorded day. Critically, these results indicate the corpus was not specially constructed. In other words, caregivers did not choose particular moments of their days to record. Caregivers had the recorder on all day, and the resulting set of recordings consisted of infants' natural auditory environment.

Human coders reliably identified music in day-long audio recordings.

Our next question was whether human coders could reliably identify music that occurred throughout day-long audio recordings of infants' at-home environments. We found that they could. Two coders independently coded music for each of the 35 days. To assess inter-rater reliability, we aggregated the sections of coded sound per file into one-minute bins. This yielded the number of seconds per each minute that each coder judged as ‘music’ throughout the coded sections of each day-long recording. We calculated Pearson correlations with these data for the two coders per day. Inter-rater reliability was remarkably high (Median = .93, SD = .08; see Supplemental Materials for further details). Figure 2.2 shows inter-rater reliability for one infant’s recorded day – for this day, the recorder was on for 16 continuous hours and 9.01 hours were coded (after pre-processing). The two coders for this day coded music highly reliably (Pearson correlation, $r = .986$, $p < .001$). Because inter-rater reliability was high across all 35 coded days, we randomly selected one coder’s codes per recording to serve as the set of music that was then further coded for features, voices, and tunes. All subsequent analyses are reported based on this set of data.

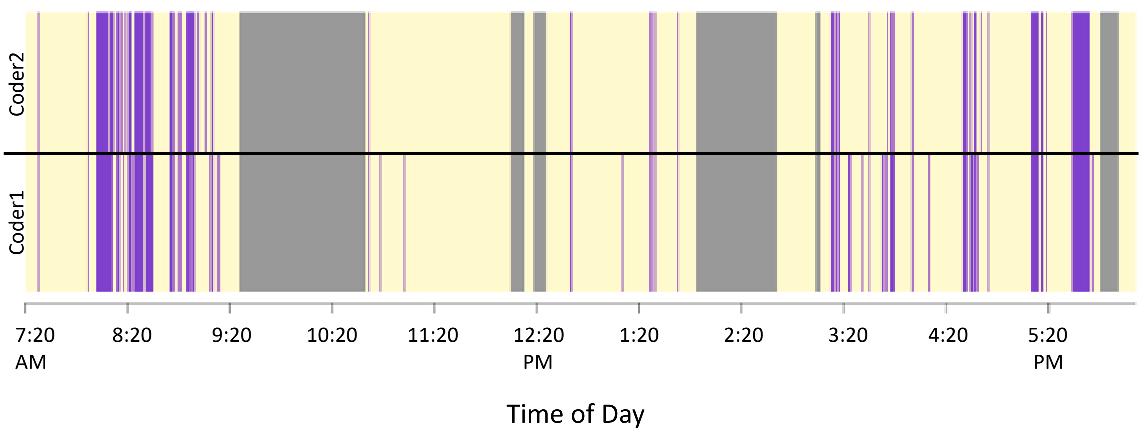

Figure 2.2. Inter-rater reliability for identifying music was remarkably high. These two time series plots depict which seconds Coder 1 (bottom row) and Coder 2 (top row) identified as music (purple) in the same infant’s day (~11 hours recorded from 7:20am until 5:20pm). The vertical alignment of the purple bars shows the high agreement ($r = .986$) between the two coders. Gray sections indicate the recorder was turned off and yellow (plus purple) shows the portions that were coded.

Human coders reliably identified features, voices, and tunes.

Next, we asked whether human coders could reliably code features, voices, and tunes across music in day-long recordings. We found that they could. Each of the 35 days was coded in full by two coders. We assessed inter-rater reliability for these recordings using the same metrics and criteria as we used to determine if coders met the training criteria (i.e., at least .90 proportion agreement for coding features; Tschuprow's *T* of at least .90 for coding voice identities and tune titles). For judging features of music bouts, inter-rater reliability was high across the board: For coding "live", the median proportion agreement was .98 (SD = .05), for coding "recorded", the median proportion agreement was .99 (SD = .04), for coding "vocal", the median proportion agreement was .98 (SD = .03), and for coding "instrumental", the median proportion agreement was .99 (SD =.05). For identifying specific voices and specific tunes, inter-rater reliability was also high. The median Tschuprow's *T* value for voices was .94 (SD =.07) and the median Tschuprow's *T* value for tunes was .90 (SD =.06). Figure 2.3 shows an example for one infant's day where the inter-rater reliability for coding voices (T = .95) was comparable to the median value across all infants. In this plot, the voices identified by Coder 1 are on the x-axis, and the voices identified by Coder 2 are on the y-axis. The rectangles are colored by frequency counts, with brighter blue indicating a higher number. Rectangles along the diagonal indicate agreement between the coders. In this example, both coders identified *Mom* as having occurred in the greatest number of music bouts, as indicated by the bright blue square in the bottom-left corner of the plot. The fact that most of the blue squares appear on the diagonal indicates the high agreement between the two coders. This plot also makes clear where coders disagreed. For example, in one instance where Coder 1 identified the voice as *Mom*, Coder 2 identified the voice as *Grandma* (as indicated by the top-most, left-most colored square). The full set of plots depicting coder agreement for tunes and voices in each infant's day are available in the Supplemental Materials.

Because inter-rater reliability for coding features, voices, and tunes was remarkably high, we selected the data from one randomly selected coder per infant to be used in all subsequent analyses.

In sum, we successfully collected day-long, at-home recordings of 35 infants and their caregivers in their natural, at-home environments. Further, we reliably identified music, including the features, voices, and tunes of the music. Critically, we captured the phenomenon of interest: music during infants' everyday lives. We have created a corpus of infant-available everyday musical input. We now address our first main question: how much musical input do infants encounter in their natural environments in a day?

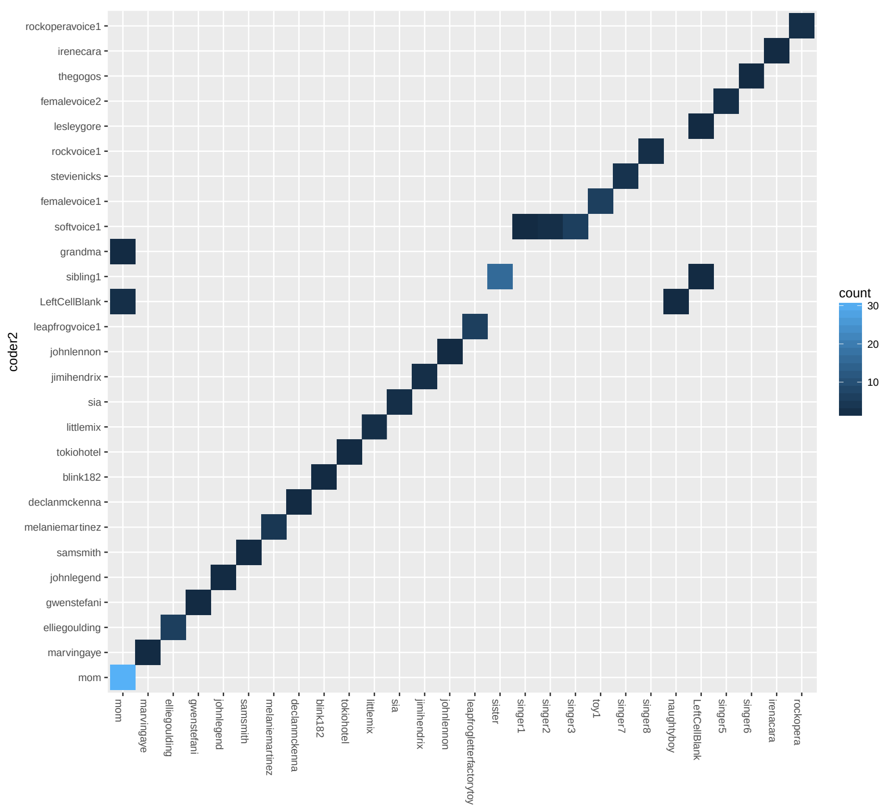

Figure 2.3. Inter-rater reliability for identifying voices in the music that occurred was high. This heatmap plot shows the voices identified by Coder 1 on the x-axis, and the voices identified by Coder 2 on the y-axis for one infant's day. The rectangles are colored by frequency counts, with brighter blue indicating a higher number of instances of that combination of voices identified by the coders. Rectangles along the diagonal indicate agreement between the coders. Inter-rater reliability was high between these two coders (T = .95), and this value reflected the median value of coder agreement across all infants.

A corpus of everyday, infant-available music.

Across all 35 infants' recorded days, we identified a total of 151,390 seconds (42.05 hours) of music that occurred in a total of 4,816 separate music events. Every infant encountered music during their day. The cumulative duration of music per day ranged from 459 seconds to 15,626 seconds (~8 minutes to ~4 hours) (Median = 3,351 seconds; SD = 3,960.39 seconds). Furthermore, individual infants encountered music from 24 to 435 times per day (Median = 127 music bouts, SD = 90.30 music bouts). Figure 2.4 shows how the cumulative duration of music and the number of music bouts varied across infants.

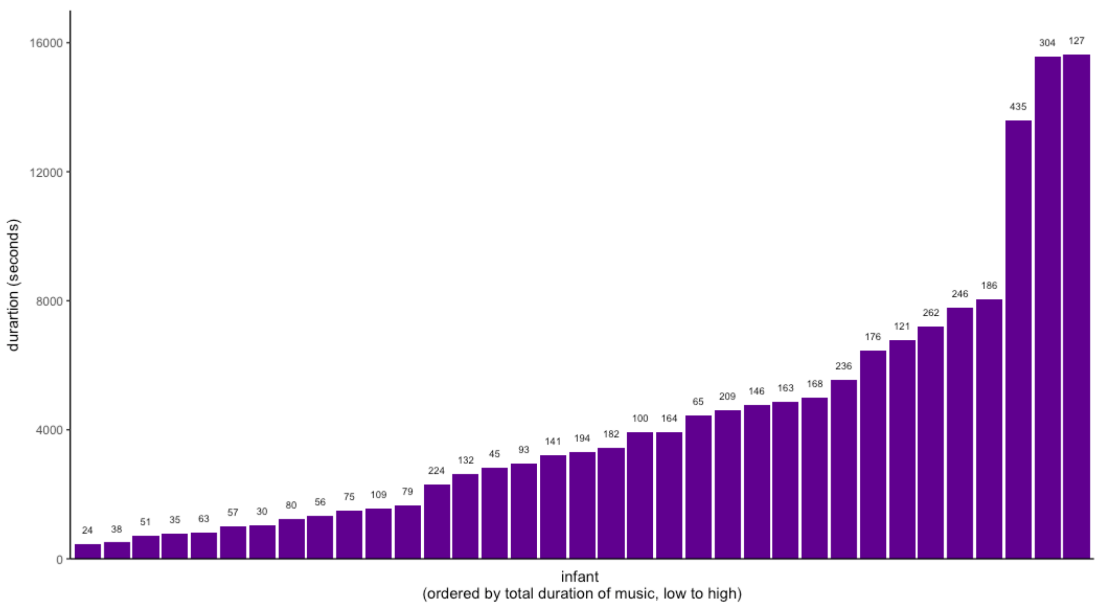

Figure 2.4. All infants encountered a non-zero duration of music (height of each bar) across a non-zero number of separate music bouts (number printed above each bar). The cumulative duration of music and the cumulative number of music bouts varied across infants. Neither the cumulative duration nor the number of music bouts correlated with any of the other reported measures.

Because the total duration of coded data differed across infants, it is difficult to interpret differences in the total amount and number of instances of music. Therefore, we calculated the proportion of coded seconds that was identified to be music. For individual infants, music accounted for between .02 and .54 (Median = .12, SD = .13) of the cumulative duration of coded seconds (see Figure 2.5). Taken together, these findings revealed that infants encountered music across multiple music events during their days.

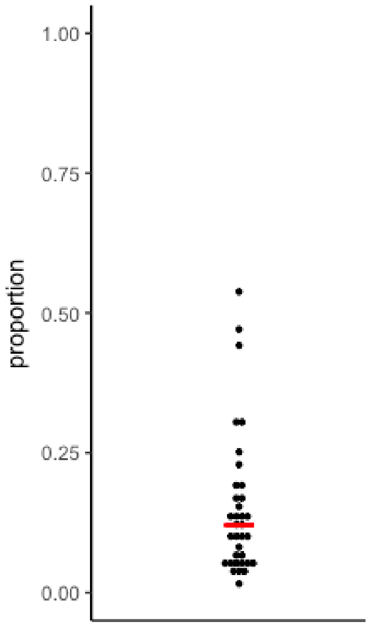

Figure 2.5. The proportion of coded seconds that was identified as music varied from .02 to .54 across individual infants' days. The median proportion of coded seconds that was identified as music across infants (.12) is shown by the red line.

An overview of the key details about the music in the corpus.

In subsequent chapters of this dissertation, we report in detail on the coding and analyses of features (i.e., live, recorded, vocal, and/or instrumental music), voices, and tunes identified in the infant-available music of this corpus. In Table 2.2, we provide an overview of the key details of music in this corpus – the grand totals in cumulative duration (in seconds) and in number of instances (i.e., onsets) of music, features, voices, and tunes across all 35 days.

Table 2.2.
The data we have collected in our corpus of infant-available everyday music contained the necessary forms of music (i.e., live, recorded, vocal, and instrumental), voices, and tunes for us to answer our main theoretical questions of interest about the structure of infants' everyday musical input.

CORPUS SUMMARY		Duration (seconds)	Number of Instances
RECORDING	Recorder On	1,680,351	47
	Coded Data	970,873	215
MUSIC	All Music Bouts	151,390	4,816
	Non-Goof Music Bouts	151,221	4,798
	Live Music	58,067	1,639
	Recorded Music	125,040	3,504
	Vocal Music	99,542	2,399
	Instrumental Music	127,561	3,571
TUNES MUSIC CORPUS (music bouts with exactly one tune)	Tunes Music	98,960	4,023
	Live-Vocal Tunes	33,743	1,289
VOICES MUSIC CORPUS (music bouts with exactly one voice)	Voices Music	63,490	2,149
	Live Voices	31,957	1,323

Goof bouts. For each day, features, voices, and tunes were coded by two different coders than the two who originally coded music bouts for that recording. Occasionally, the coders coding features, voices, and tunes came across a section that had been previous coded as "music" in which they could not discern a musical sound. These were considered "goof" bouts, and they were not coded for features, voices, or tunes. Goof bouts were rare (n=18 across the whole corpus), and they accounted for a very small proportion (.001) of the total number of music bouts in the corpus. Thus, 151,221 seconds of music that occurred in 4,798 non-goof music bouts were coded and analyzed for features, voices, and tunes.

Features. Feature were coded at the level of music bouts, and the features categories (i.e., live, recorded, vocal, and instrumental) were not mutually exclusive. Almost three quarters of music bouts contained recorded music (.73) and/or instrumental music (.74; see Table 2.2). This contrasted with existing literature (e.g., Trehub & Schellenberg, 1995; Trehub & Trainor, 1998) that describes music with infants as caregivers singing (i.e., live-vocal music). However, given the increasing use of technology for music listening (e.g., Greasley & Lamont, 2011; Krause & North, 2015), scholars have recently questioned the prevalence of recorded music in infants' everyday lives (Trainor & Hannon, 2013). Discovering a larger proportion of recorded music was consistent with evidence reported in a recent case study of one family's daily musical input (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016).

Two subsets of the music corpus: Tunes Music Corpus and Voices Music Corpus.

Music bouts with exactly one tune or exactly one voice. It was possible for a music bout to contain more than one unique tune and/or more than one unique voice. For example, if mom were singing along to Katy Perry's *Roar*, then that music bout would contain two voices: *Mom* and *Katy Perry*, and one tune: *Roar*. In our current coding scheme, coders did not identify precise timing of individual voices and/or tunes that occurred within the same music bout, nor did they code features separately for each voice instance and each tune instance per bout. In the given example, it could have been that *Mom* and *Katy Perry* were both available throughout the entire duration of the music bout. Or it could have been that mom sang along for only part of the duration of the music bout, and then Katy Perry's voice continued alone. Or it could have been that the music bout included an instrumental interlude of *Roar*, such that only mom's voice was available for part of the music bout and not Katy Perry's voice. In on-going work, we are further coding these music bouts with multiple voices and/or multiple tunes to obtain precise timing and features coding for each individual instance per bout.

For the research reported in this thesis, we focused our analyses on the clearest subset of music bouts – music bouts that contained exactly one unique tune (“tunes music corpus”) and music bouts that contained exactly one unique voice (“voices music corpus”). We excluded any music bouts that contained multiple tunes or multiple voices. All subsequent analyses are based on these two corpora – for analyses related to tunes, we analyzed the tunes music corpus; likewise, for analyses related to voices, we analyzed the voices music corpus. These two corpora were not mutually exclusive. Music bouts that contained exactly one tune *and* exactly one voice were included in both corpora. As an example, if a music bout consisted of *Dad* singing *Happy Birthday* and no other musical sounds, then that music bout would be included in both the tunes music corpus and the voices music corpus. On the other hand, the example given above of *Mom* singing along to Katy Perry’s *Roar* would have been excluded from the voices music corpus, because it had two unique voices (*Mom* and *Katy Perry*), but it would have been included in the tunes music corpus, because it only had one unique tune (*Roar*).

Across all 35 days, this resulted in keeping a total of 4,023 music bouts with exactly one tune (.84 of all music bouts) and of 2,149 music bouts with exactly one voice (.86 of all vocal music bouts). In terms of duration, this yielded 98,960 seconds of music with exactly one tune (.65 of the total duration of all music) and 63,490 seconds of music with exactly one voice (.64 of the total duration of all vocal music). Of these values, 1,842 music bouts (total duration = 46,338 seconds) had both exactly one tune and exactly one voice and were thus included in both corpora. In individual infants’ days, the proportion of music bouts that had exactly one tune ranged from .64 to .98 (Median = .85, SD = .08; see Figure 2.6A), and the proportion of cumulative seconds of music with exactly one tune ranged from .38 to .96 (Median = .72, SD = .16; see Figure 2.6B). In other words, the bulk of infants' music bouts (and of their cumulative duration of music) contained exactly one tune. For music with exactly one voice, individual infants exhibited considerable variation – individual infants retained from .12 to .90 (Median = .46, SD = .22; see Figure 2.7A) of their cumulative music bouts and from .07 to .87 (Median = .41, SD = .20; see Figure 2.7B) of the cumulative seconds of music.

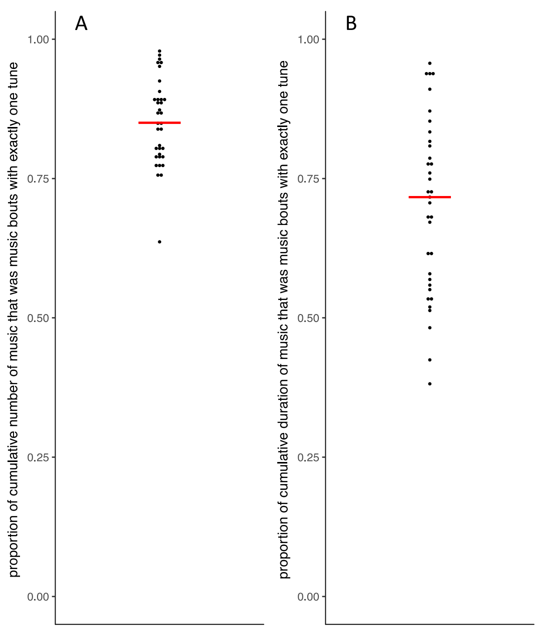

Figure 2.6. For most infants, at least half of their music bouts (A) and half of their cumulative seconds of music (B) contained exactly one tune. The red lines depict the median values across infants.

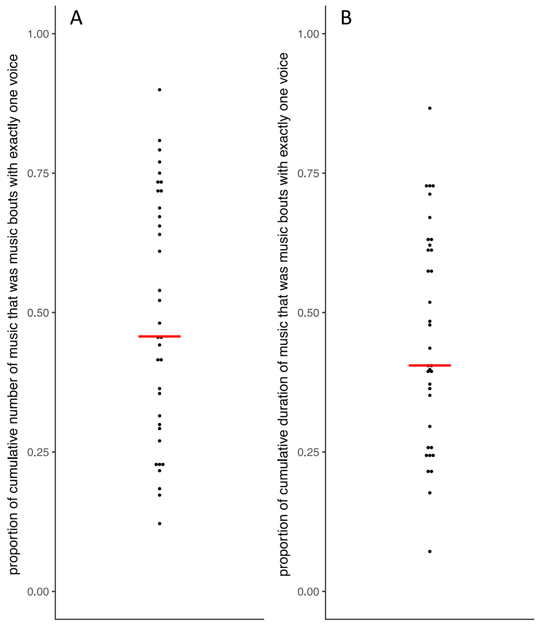

Figure 2.7. The number of music bouts (A) and the cumulative duration of music (B) that contained exactly one voice varied widely across infants. The red lines represent the median values across infants.

Live-vocal tunes and live voices. In subsequent chapters of this thesis, we will analyze the subset of music that was coded as both 'live' and 'vocal', because this combination is likely to yield high-quality musical input. Because the features categories were not mutually exclusive, there were two possible sets of live-vocal music: Inclusively live-vocal music was coded as 'live' and 'vocal' and could have also been coded as 'recorded' and/or 'instrumental'. Exclusively live-vocal music was coded as 'live' and 'vocal' but could not have also been coded as 'recorded' or 'instrumental'. For example, if *Dad* played the guitar while singing *Dancing in the Dark*, then this music bout would be coded as 'live', 'vocal', and 'instrumental'. This example would be counted as inclusively live-vocal music; but, it would be excluded from the set of exclusively live-vocal music. We analyzed the inclusively live-vocal set of music in order to broadly characterize the structure of high-social-quality input.

In sum, the corpus of infant-available music we have collected contains the necessary data – various forms of music (i.e., live, recorded, vocal, and instrumental), voices, and tunes – for us to answer our main theoretical questions about the structure of infants' everyday musical input.

Preliminary checks for covariance among measures.

In this research, we examined many variables: demographic variables about the infants and their families (e.g., infants' age in weeks), sampling variables about the recordings (e.g., cumulative duration recorded) and structural variables about the music identified in the recordings. We conducted preliminary checks for covariance among these measures by examining Spearman rank correlations with a Bonferroni correction for multiple comparisons (see Supplemental Materials for further details). The two most important variables to examine were the cumulative duration recorded and the cumulative duration coded. If either of these variables were correlated with any of our dependent measures, then this would raise potential concern about whether any discoveries about the structure of infants' everyday music depended on how much data was recorded or how much sound was coded in the infants' day. Because we included only recordings with at least 10 cumulative hours recorded and the maximum possible recorded duration was 16 hours, there was a limited range in the cumulative duration recorded across the recordings in the corpus. The cumulative duration recorded did not correlate with any other measures reported in this thesis. The amount of coded sound, which varied from 8,595 seconds to 40,775 seconds, also did not correlate with any other measures reported in this thesis. These findings were critical – the lack of significant correlations between the recording variables and our main measures of interest enabled us to make claims about how infant-available music differed across families, rather than how the sampling of infants' days differed.

DISCUSSION

We have collected a first-of-its kind corpus of the raw auditory musical input available to young infants in their everyday lives. The rigorous coding schemes and training procedures we developed led to reliable coding of music in these auditory recordings. This newly collected corpus contains over 42 hours of music captured in audio recordings of infants and caregivers, at home in their natural environments.

We highlight two important aspects of the music in this corpus: First, no recorded days lacked music altogether. Second, no infants encountered music only once per day. Or stated differently, every infant encountered music that occurred across multiple music events in a day. Given that caregivers commonly report singing and playing music for their infants daily (Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003; Ilari, 2005; Rideout, 2013; Rideout, Vandewater, & Wartella, 2003; Trehub et al., 1997), we expected to capture music in many of the recorded days. But, it did not have to be the case that every infant in our sample encountered music, nor that every infant encountered multiple music instances in their day. Further, since all infants in the sample encountered music, the patterns we discover and report will represent the musical input available across multiple infants' recorded days – an important advance over prior research which has reported on patterns available in one family's daily musical input (Costa-Giomi, 2017; Costa-Giomi & Sun, 2016).

Our predictions about how consistency, quality, and temporal dynamics of infants' everyday musical input shape their real-world music learning depended on infants encountering a non-zero amount of music across multiple music events in their days. Our research provides the first empirical evidence on the cumulative duration and frequency of occurrence of music in infants' everyday lives. Both frequency of occurrence and cumulative duration influence attention and memory (e.g., Hintzman, 1970) and therefore have important implications for shaping how infants learn about music. We will further address these implications in subsequent chapters of this thesis. The results reported here set the stage for us to answer questions about the structure of infants' everyday musical input in subsequent chapters. We asked and answered the following two broad questions: (1) In what ways did infants encounter consistency, diversity, and social quality in the tunes and voices that occurred in their everyday musical input?, and (2) How were individual instances of music and its contents distributed over time across infants' days? In Chapter 2, we use the corpus of music available to infants in their everyday lives to examine the consistency, diversity, and social quality of the voices that produced music and of the tunes that occurred. In Chapter 3, we will examine the temporal dynamics of music, tunes, and voices as they occurred over the course of a day. By capturing and characterizing music in day-long recordings of infants in their natural, at-home environments, we provided an altogether new window on the structure of musical input available to infants in their everyday lives. In turn, a more detailed understanding of infants' everyday musical input has potential to yield fundamental insights into how infants build knowledge in the domain of music.

CHAPTER III

CONSISTENCY, DIVERSITY, AND SOCIAL QUALITY OF TUNES AND VOICES IN INFANTS' EVERYDAY MUSIC

INTRODUCTION

Research in many domains with infants, children, and adults points to three key dimensions of encounters with items that shape what people learn: consistency, diversity, and social quality of the input. It is clear how these properties matter for learning in laboratory-based experiments (e.g., Baldwin, Markman, Bill, Desjardins, Irwin, & Tidball, 1996; Carvalho & Goldstone, 2013; Elio & Anderson, 1984; Horst, Parsons, & Bryan, 2011; Kuhl, Tsao, & Liu, 2003; Kurumada, Meylan, & Frank, 2013; Oakes & Spalding, 1997; Perry, Samuelson, Malloy, & Schiffer, 2010; Rost & McMurray, 2009; Twomey, Ranson, & Horst, 2013; Valian & Coulson, 1988). Yet, little is known about how these properties could matter for learners facing real-world input (Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, & Trueswell, 2013; Clerkin, Hart, Rehg, Yu, & Smith, 2017; Ramírez-Esparza, García-Sierra, & Kuhl, 2014, 2017; Roy, Frank, DeCamp, & Roy, 2015). This is because researchers are just beginning to detail the degree of consistency, diversity, and social quality of to-be-learned items in the context and scale of natural everyday experience (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Cole, Robinson, & Adolph, 2015; Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2015; Lee, Cole, Golenia, & Adolph, 2017; Roy, Frank, DeCamp, & Roy, 2015). Here, we tackle these foundational dimension descriptions in the domain of music.

Suppose you want to learn how to recognize the tune *Twinkle Twinkle Little Star*. How would you do it? You would need to encounter it (e.g., hear a person singing *Twinkle Twinkle Little Star*). But how often would you need to encounter it to successfully encode it, recognize it, and discriminate it from other tunes? Do you need to hear *Twinkle Twinkle Little Star* many, many, many times or would just once be enough? When you hear *Twinkle Twinkle Little Star*, would you need to hear it sung by a live voice (e.g., your mom) or by a recorded voice (e.g., from a toy)? Each time you hear music, would you need to always encounter *Twinkle Twinkle Little Star*, or would you also need to experience other tunes (e.g., *Itsy Bitsy Spider, Rock-a-bye Baby*)? If you hear multiple tunes, then would you need to hear each one equally often, or would you need to hear *Twinkle Twinkle Little Star* more than the others?

If you frequently encounter a live voice (e.g., mom) singing *Twinkle Twinkle Little Star*, then you could pay attention to and form a robust memory of mom singing *Twinkle Twinkle Little Star.* You could then use this strong memory as an anchor, comparing it to other renditions of *Twinkle Twinkle Little Star* and notice how they overlap. In other words, you could aggregate all the instances of *Twinkle Twinkle Little Star* (sung by mom, sister, grandma, etc.) and keep them separate from instances of other tunes (e.g., all the instances of *Itsy Bitsy Spider*). This could prevent you from being overwhelmed by variation, while also helping you to move beyond one specific item. This example illustrates how consistency, diversity, and social quality might shape music learning, given what we know about how these key properties shape learning in other domains, both in the lab and in natural settings.

Consistency and diversity shape learning in the laboratory.

Many laboratory-based studies on language learning and category learning have demonstrated that consistency and diversity of presented information matter for learning and memory. Broadly, consistency refers to repetition of experience. Learners of all ages build knowledge by aggregating across repeated instances; consistency helps learners integrate across those experiences (Goldstein et al., 2010). Diversity refers to variability – anything that is not an exact repetition is diversity. Thus, diversity can take many forms. For example, variability could refer to the number of items encountered and/or to within-category differences across a set of items. Repetition and variability of the input both shape what and how infants, children, and adults learn in laboratory settings (e.g., Braithwaite & Goldstone, 2015; Deutsch, 1975; Elio & Anderson, 1984; Gómez, 2002; Hintzman & Block, 1971; Hintzman, Grandy, & Gold, 1981; Horst, Parsons, & Bryan, 2011; Houston & Jusczyk, 2000; Nosofsky, 1988; Oakes & Spalding, 1997; Perry, Samuelson, Malloy, & Schiffer, 2010; Rost & McMurray, 2009; Singh, 2008; Schwab & Lew-Williams, 2016; Twomey, Ranson, & Horst, 2013; Thiessen, 2011). Consistency and diversity have largely been studied separately, but how they work together also matters for learning (e.g., Bortfeld et al., 2005; Cameron-Faulkner, Lieven, & Tomasello, 2003; Casenhiser & Goldberg, 2005; Frost et al., 2016; Kurumada, Meylan, & Frank, 2013; Shi et al., 2006; Valian & Coulson, 1988). For example, in one study, infants recognized novel words better when the novel word followed a familiar, high-frequency word (e.g., the infant's own name or 'mommy') than if it followed an unfamiliar word (Bortfeld et al., 2005). In these studies, learning is shaped by encountering high-frequency 'anchor' items (i.e., consistency; Valian & Coulson, 1988) that bootstrap subsequent learning of other low-frequency items (i.e., diversity; e.g., Kurumada, Meylan, & Frank, 2013; Oaks & Spalding, 1997).

Consistency and diversity shape real-world language learning.

It is unclear how the findings from laboratory-based studies map to real-world learning. Researchers have operationalized consistency and diversity in the lab in a variety of ways, depending on the particular study, rather than basing the input on what learners actually encounter in the real world. The best available evidence about consistency and diversity of real-world input comes from research on the natural linguistic input available to infants. First, by one estimate, infants encounter between 2,000 and 29,000 words in a day (Weisleder & Fernald, 2013). Greater total amounts of linguistic input have been linked to faster, more robust vocabulary growth (Hart & Risley, 1995; Hoff, 2003; Hoff, & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; Pan, Rowe, Singer & Snow, 2005; Rowe, 2012; Weisleder & Fernald, 2013). Second, words that occur more frequently in natural speech input (i.e., more consistent) are acquired earlier (e.g., Goodman, Dale, & Li, 2008; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; for recent review, see Ambridge, Kidd, Rowland, & Theakston, 2015), and young children more often produce the nouns they have more frequently encountered than nouns they have less frequently encountered in their available linguistic input (e.g., Naigles & Hoff-Ginsberg, 1998; Roy, Frank, DeCamp, Miller, & Roy, 2015). Third, natural linguistic input to infants and young children contains a higher proportion of repeated utterances than speech to older children and adults, suggesting that degrees of consistency in natural input may change over time (e.g., Fernald & Morikawa, 1993; Hills, 2012; Hoff-Ginsberg, 1985; Kaye, 1980; Kaye & Charney, 1981; Snow, 1972). Finally, in natural language, the distribution of word frequencies is "biased" (i.e., non-uniform) such that a few words occur at high frequencies making up the bulk of the tokens, while a large number of words occur infrequently; this type of biased distribution closely reflects a particular mathematical form known as *Zipf's law* (Zipf, 1945; for review, see Piantadosi, 2014). To the extent that a frequency distribution of words is biased, then there are some words that are consistent amidst a range of diversity (for comparable evidence about distributional properties in natural, infant-available visual information, see Clerkin, Hart, Rehg, Yu, & Smith, 2017; Smith & Slone, 2017). How infants link these properties of natural consistency and diversity to real-world learning has only recently been articulated as a question and the answers remain unknown.

Social quality shapes infants' language learning.

In most domains, there is evidence that some learning opportunities are higher quality than others. In language learning, one marker of quality is social context – signified by a live social agent (e.g., Kuhl, Tsao, & Liu, 2003) interacting in an infant-directed (e.g., Kuhl, 2007) and contingent manner (e.g., Goldstein, King, & West, 2003). Infants more readily learn from a wide range of linguistic input, including statistical regularities (e.g., Thiessen, Hill, & Saffran, 2005), phonetic contrasts (e.g., Kuhl, Tsao, & Liu, 2003), and object-label mappings (e.g., Baldwin, Markman, Bill, Desjardins, Irwin, & Tidball, 1996), when that input is presented in a social context (for reviews on social context and language learning, see Hoff, 2006; Baldwin & Meyer, 2007). In one striking example, 9- to 10-month-old infants learned to discriminate a foreign phonetic contrast from Mandarin Chinese only after engaging in live, social interaction with a native Mandarin speaker; infants who experienced video recordings of these same interactions failed to discriminate the foreign contrast (Kuhl, Tsao, & Liu, 2003). Several possible mechanisms have been proposed to underlie such 'social gating' effects (Kuhl, 2007): social context may engage infants' attention and arousal (Samuelson & Smith, 1998; Kuhl, 2007), boost infants' memory (Samuelson & Smith, 1998), provide specific content about referential information (Kuhl, 2007), and/or engage infants' social understanding of agents' goals and intentions (Baldwin & Moses, 2001; Tomasello, 2010).

Current Research

In the current research, we evaluated distributional properties of infant-available musical input. Infants encountered about one hour of music in their everyday auditory environments (see Chapter 1). Based on caregiver report and laboratory-based studies (Custodero & Johnson-Green, 2003; Ilari, 2005; Longhi, 2009; Mendoza & Fausey, 2015; Trehub et al., 1997), we expect that infant-available music will contain tunes (i.e., melodic sequences), some of which are produced by voices. Specifically, we examined consistency, diversity, and social quality of the tunes and voices that occurred in infant-available music.

We focused on tunes, because during their first year of life infants learn to process, encode, and remember tunes. Laboratory-based studies have documented that 5- to 11-month-old infants detect changes in the pitch frequency of a single tones in a musical sequence (Chang & Trehub 1977a, 1977b; Trehub, Bull, Thorpe, 1984; Trehub, Thorpe, Morrongiello, 1987). Infants recognize melodic sequences as ‘the same’ across transpositions (i.e., sequences that retain the relative pattern of pitch intervals but that are comprised of different sets of absolute pitch frequencies; Chang & Trehub, 1977a; Plantinga & Trainor, 2005; Trehub, Bull, & Thorpe, 1984). Further, 6- to 8-month-old infants show long-term memory for melodies based on relative pitch (e.g., Ilari & Polka, 2006; Plantinga & Trainor, 2005; Trainor, Wu, Tsang, 2004; for reviews, see Trainor & Hannon, 2013; Trehub, 1989, 2006). In these studies, infants were exposed to musical stimuli and their subsequent learning and memory were evaluated. This provides evidence that the musical input infants encounter matters for their learning and memory and raises questions about the musical input infants encounter in their everyday lives. We focused on vocal music, because voices are especially salient acoustic signals (e.g., Belin, Zatorre, & Ahad, 2002; Levy, Granot, & Bentin, 2001; Nakata & Trehub, 2004; for review, see Belin, Fecteau, & Bédard, 2004). Children and adults also show better memory for vocal melodies compared to instrumental ones (e.g., Weiss, Schellenberg, & Trehub, 2017; Weiss, Trehub, & Schellenberg, 2012; Weiss, Schellenberg, Trehub, & Dawber, 2015; Weiss, Trehub, Schellenberg, & Habashi, 2016). While this has not been directly tested in infants, it is plausible that vocal music could have a comparable impact on infants’ learning and memory. Thus, in the current research, we examined the input that presumably drives infants’ musical learning in order to discover the structure of infant-available musical input by determining the distributional properties of tunes and voices that occur in infants’ everyday music. We addressed the broad question: In what ways did infants encounter consistency, diversity, and social quality in the tunes and voices that occurred in their everyday musical input?

Consistency and diversity of tunes and voices. To characterize consistency and diversity of the tunes and of the voices that infants encountered in their everyday music, we answered several questions. First, we analyzed tune types in the set of music bouts that contained exactly one tune (tunes music corpus) and we analyzed voice types in the set of music bouts that contained exactly one voice (voices music corpus) We refer to unique, specific tunes or voices as "types" (e.g., tune types: *Twinkle Twinkle Little Star, Itsy Bitsy Spider, Wheels on the Bus*; voice types: *Mom, Dad, Lady Gaga, Daniel Tiger*), and we refer to each separate occurrence of a tune or a voice as an "instance" or "token". To assess whether infants encountered diversity in the tunes and/or the voices of their daily musical input, we asked: (1) how many tune types occurred in how many seconds of tunes music and how many voice types occurred in how many seconds of voices music? If infants encountered only one tune (voice) type in all of the cumulative music seconds of their day, then this would present maximal consistency, and no diversity. For example, an infant could encounter their mom singing *Itsy Bitsy Spider* in all of their daily music. If infants encountered more than one tune (voice) type, then that would introduce some diversity. For example, an infant might encounter their mom singing *Itsy Bitsy Spider,* a voice from a toy singing *Wheels on the Bus,* their dad humming *Wheels on the Bus,* and the radio playing Lady Gaga's *Born This Way.* It is plausible that encountering a greater cumulative duration of music could result in greater diversity of tune (voice) types – that an infant would have more opportunity to encounter more, different specific tunes (voices). While this possibility might seem intuitive, it does not have to be the case. It is also possible that encountering longer cumulative durations of music would yield greater consistency of tune (voice) types – if there is one tune (voice) type that dominates infants' musical input, then more music might mean more of that one specific tune (voice). In questionnaires and diary logs, caregivers have previously reported that multiple different voices (i.e., mothers, fathers, and siblings) sing multiple different tunes to their infants (e.g., Custodero & Johnson-Green, 2003; Ilari, 2005; Mendoza & Fausey, 2015; Trehub et al., 1997), so we expected that infants would encounter diversity in the tune types and the voice types of their daily music.

Because we expected that infants would not encounter maximal consistency of tune types and voice types, we aimed to quantify the consistency that infants did encounter in their everyday musical input. To do so, we separately examined how tune types and voice types were distributed within their ranges in infants' encountered seconds of music in a day. One possibility is that infants encountered each tune (voice) type for an equal cumulative duration – a "balanced" frequency distribution of voice types. Alternatively, tune types and voice types could have occurred for durations that deviated from being balanced – this would yield "biased" frequency distributions, in which some tune (voice) types would be more available than others. For our second question, we asked: (2) did infants encounter balanced frequency distributions of types or of voice types? As an example of a balanced frequency distribution**,** an infant's one hour of music could contain 6 voice types that each occurred for 10 minutes: mom singing, dad humming, big sister singing and whistling, Beyoncé singing on the radio, Daniel Tiger singing in YouTube videos, and a singing voice from a toy. In this way, the infant would encounter each specific voice equally often. In contrast, an infant's one hour of music might contain 6 tune types: 23 minutes of *Twinkle Twinkle Little Star*, 14 minutes of *Itsy Bitsy Spider*, 12 minutes of *Dancing in the Dark*, 5 minutes of *Hello My Old Heart*, and 4 minutes of *Eine kleine Nachtmusik*, and 2 minutes of *Beat It.* Thus, the infant would encounter some tune types (*Twinkle Twinkle Little Star, Itsy Bitsy Spider, Dancing in the Dark*) for longer durations and other tune types (*Hello My Old Heart, Eine kleine Nachtmusik*, and *Beat It*) for shorter durations relative to a balanced model. Given that natural input in other domains exhibits biased frequency distributions (e.g., Zipf, 1945; Clerkin, Hart, Rehg, Yu, & Smith, 2017; Smith & Slone, 2017), we expected infants to encounter biased frequency distributions of tune types or of voice types.

If infants were to encounter biased frequency distributions of tune types and voice types, then what would be the particular form of these frequency distributions? There are many different types of biased (or "non-uniform") frequency distributions (e.g., power-law, exponential distributions). Ideally, to determine the precise form of the frequency distributions of tune types and voice types that infants encountered in their everyday music, we would fit a curve to each infant's data to assess whether the infant's actual frequency distribution matched the shape of the particular curve. However, differentiating among many possible types of biased frequency distributions requires large amounts of data (e.g., Clauset, Shalizi, & Newman, 2009). Even though we have collected the largest existing corpus of infant-available music, we did not have enough music data from each infant to fit curves (see Supplemental Materials for further information about fitting a power law to these data). In lieu of fitting a curve to the infants' data, we evaluated consistency and diversity by quantifying two components of the frequency distributions: the top tune (voice) type and the rest of the tune (voice) types in the top of the frequency distribution. We first focused on the single tune (voice) type that occurred for the longest duration (i.e., the "top type"). We asked: (3) to what extent was the top tune type or top voice type more available relative to a balanced null distribution? It could be that the top tune (voice) type occurred only a little bit longer than would be expected if each tune (voice) type occurred for an equal duration. Or, the top tune (voice) type could have occurred for substantially longer relative to a balanced null distribution. A larger deviation from a balanced null distribution would mean that one specific tune (voice) type was particularly available in infants' everyday music – this is one index of consistency in tune (voice) types.

Next, we zoomed out from the top tune (voice) type to examine other components of the frequency distributions. We asked: (4) to what extent were some tune types or voice types more available than other tune types or voice types? As in the examples above, it could be that tune (voice) types occurred in a balanced frequency distribution such that each tune (voice) type was equally available. Or it could be that tune (voice) types occurred in a biased frequency distribution, such that some tune (voice) types were more available than others relative to a balanced null model. To assess this, we examined the degree to which other tune (voice) types besides the top tune (voice) type also occurred for longer durations that would be expected given a balanced null distribution. This measure served as another index of the consistency of tune (voice) types that occurred in infants' daily music.

Taken together, these measures yielded insight into the mix of consistency and diversity of tune types and voice types that infants encountered in their everyday musical input. These questions about consistency and diversity were important to answer because the answers revealed the structure of the input that is available to infants – structure that has potential to shape infants' music learning. In music research, empirical evidence from one in-laboratory study has demonstrated that the combination of consistency and diversity supports music learning (Loui & Wessel, 2008) – adults better recognized unfamiliar melodies generated by a novel, artificial music system when they were exposed to a small range and greater repetition of the tunes than when they experienced a larger range and less repetition. Therefore, it is important to characterize the consistency and diversity that infants encounter in the tunes and voices of their real-world musical input – it is possible that infants' everyday musical input presents a helpful mix of consistency and diversity that could ultimately support infants' music learning.

Social quality. To examine the social quality of infants' everyday music input, we focused on the subset of music that was coded as being both 'live' and 'vocal'. We considered this live and vocal music to be a proxy for high social quality, because infants are typically engaged more by live presentations than by recorded ones (e.g., Kuhl, Tsao, & Liu, 2003). Moreover, children and adults show better memory for vocal music compared to instrumental music, demonstrating a processing advantage for vocal music (e.g., Weiss, Schellenberg, & Trehub, 2017; Weiss, Trehub, & Schellenberg, 2012; Weiss, Schellenberg, Trehub, & Dawber, 2015; Weiss, Trehub, Schellenberg, & Habashi, 2016). Together, these findings suggest that live-vocal music may play a particularly strong role in shaping infants' processing. Thus, we answered three questions about the frequency distributions of live-vocal music that infants encountered.

First, we asked: (5) what proportion of infants' everyday music was high-social-quality by virtue of being coded as 'live' and 'vocal' (i.e., "live-vocal" music)? At one extreme, infants might only encounter recorded music from radios, TVs, and toys. At the other extreme, infants' musical input could consist solely of a live caregiver singing to them. Most likely, infants' everyday musical input fell in the middle of these extremes – a mix of music, some of which was live and vocal. To the best of our knowledge, the only estimate of the proportion of infants' everyday music that is live comes from an experience-sampling study of music available to 3-year-old children (Lamont, 2008). In this study, caregivers were called three times per day for 7 days. Of the resulting 437 sampled episodes, caregivers reported that music occurred during 353 (.81) of them, and only 55 of these music episodes (.16) included live singing. The remaining music episodes consisted of recorded music, mainly from CDs and TV programs. However, it is not clear to what extent the everyday musical experiences of 3-year olds should generalize to the everyday musical input available to young infants. For one, surveys have shown that young infants encounter a greater total amount of music (e.g., Rideout, 2013) than preschoolers do. For another, the quality of caregivers' singing differs depending on whether it is directed to infants versus preschoolers (Bergeson & Trehub, 1999), suggesting the function of music may be different for infants versus children. Thus, the proportion of infants' everyday musical input that is of high-social quality by virtue of being both live and vocal remains unknown.

By some accounts (e.g., Medina, Snedeker, Trueswell, Gleitman, 2011; Trueswell et al., 2013), it is *only* high-quality input that shapes the developing system. In fact, in the domain of language learning, the amount of child-directed speech (i.e., high-quality), and not the cumulative linguistic input, that young children encountered in their everyday lives strongly predicted their subsequent language processing skill and vocabulary growth (Weisleder & Fernald, 2013). If it is also true in the domain of music learning that only high-quality input impacts the developing system, then it is important to characterize the distributional properties of that input. Thus, for the subset of live-vocal music (i.e., high-social-quality music), we answered the same set of questions about the consistency and diversity of the frequency distributions of tune types and voice types as we did for all music. Broadly, we asked: (6) did infants encounter balanced frequency distributions of live-vocal tune types or of live voice types? We make the same predictions as with the frequency distributions of all voice types and of all tune types – infants might encounter equal amounts of each live-vocal tune and of each live voice. Or the frequency distributions of live-vocal tune types and of live voice types might exhibit a consistency bias, such that some live voices and some live-vocal tunes occur more than would be expected given balanced null distributions of live-vocal tunes and live voices.

Because there is also evidence that infants learn from all input, and not just from some portions of the input (Yu & Smith, 2007; Yurovsky, Fricker, Yu, & Smith, 2014; for relevant discussion, see also Smith, Suanda, & Yu, 2014), it was important to understand how high-social-quality input fits into the mix of all encountered input data. Therefore, we examined live-vocal music in the mix of all music; we asked: (7) how did live-vocal tunes and live voices occur in the mix of all tunes and all voices? One possibility is that live-vocal music was clustered at one end or the other of the frequency distributions of tune types and of voice types. It could be that only the specific tunes at the top of the frequency distributions – those that occurred for the longest cumulative durations – were both live and vocal and that only the specific voices that occurred for the longest cumulative durations were live. It could also be the opposite, such that only the specific tunes that occurred for the shortest cumulative durations were both live and vocal and that only the specific voices that occurred for the shortest cumulative durations were live. In this case, live-vocal music would be clustered at the bottom of the frequency distributions of tune types and of voice types. This possibility would most closely reflect the everyday music available to 3-year-old children (Lamont, 2008), but it remains unknown to what extent this also characterizes infant-available music. Finally, it could be that live-vocal music was spread out across the frequency distributions of tune types and of voice types, such that a mix of specific tunes that occurred for longer and shorter durations were both live and vocal and a mix of specific voices that occurred for longer and shorter durations were live.

Answering these three questions will yield initial estimates of the social quality of the musical input available to infants in their everyday environments and will provide insight into the distributional properties of the high-social-quality musical input that infants encounter. Prior research has shown that the social quality of infants' formal musical experiences relates to their music learning – in one study, infants who participated in a music education class designed to promote active, social engagement acquired knowledge of Western tonal structure at a younger age than their peers who participated in a class where music was simply played passively in the background (Gerry, Unrau, & Trainor, 2012). It is possible that infants encounter highly social musical input in their natural environments and that such input likewise shapes infants' everyday music learning.

In sum, using the tunes music corpus and the voices music corpus, we addressed multiple questions about the distributional properties of the contents of music that infants encounter in their everyday lives. Specifically, we addressed the following questions: (1) how many tune types occurred in how many seconds of tunes music and how many voice types occurred in how many seconds of voices music?, (2) did infants encounter balanced frequency distributions of tune types or of voice types?, (3) to what extent was the top tune type or top voice type more available relative to a balanced null distribution? and (4) to what extent were some tune types or voice types more available than other tune types or voice types? Then, we addressed three broad questions about the social quality of the tunes and of the voices infants encountered in their daily music distributions: (5) what proportion of infants' everyday music was both live and vocal?, (6) did infants encounter balanced frequency distributions of live-vocal tune types or of live voice types?, and (7) how did live-vocal tunes and live voices occur in the mix of all tunes and all voices?

METHOD

Details about the participants and the procedures for collecting, pre-processing, and coding music, features, voices, and tunes in the data are described in Chapter 1.

Data Analysis

Tunes music corpus and voices music corpus. It was possible for a music bout to contain more than one unique tune and/or more than one unique voice. Here, we excluded music bouts with more than one tune or more than one voice (see Chapter 1 for details). For analyses of tunes, we analyzed 98,960 seconds (27.49 hours) of music that contained exactly one tune (tunes music corpus). For analyses of voices, we analyzed 63,490 seconds (17.64 hours) of music with exactly one voice (voices music corpus).

Frequency distributions. To address our questions about the combination of consistency and diversity of tune types and voice types, we created frequency distributions based on the cumulative durations of tune types and separately of voice types that occurred in each infant's musical input data. We first added the duration of each instance of the same tune type (i.e., the cumulative number of seconds across every instance of *Twinkle Twinkle Little* Star). We then calculated the proportion of the cumulative duration of music that was accounted for by the cumulative duration of each specific tune (voice) type per infant. We then ordered the tune (voice) types by their rank frequency, such that the left-most tune (voice) type accounted for the greatest proportion of the cumulative duration of instances. The advantage of examining proportions (rather than raw durations) is that proportions take into account differences across infants in cumulative durations of music with exactly one tune or exactly one voice. We constructed separate frequency distributions for tune types and voice types in each infants' daily music. Because our research questions focused on how the cumulative duration of music was distributed across different tune (voice) types, we opted to present our frequency distributions with types on the x-axis and proportion on the y-axis. This orientation helped to make our main comparisons of interest visually clear.

Figure 3.1 displays a hypothetical frequency distribution of tune types, to illustrate how these distributions were constructed. This hypothetical frequency distribution depicts a cumulative 3,600 seconds (1 hour) of music distributed across 6 tune types (each bar). The height of the bar depicts the proportion of the cumulative duration of music that was accounted for by the cumulative duration of that tune type. The cumulative duration (in seconds) of each tune type is printed above each bar, and the number of instances per tune type is printed within each bar. The gray horizontal line depicts the proportion of the cumulative duration of music that would be accounted for by any one tune type if each tune type occurred for an equal duration (*balanced-null proportion*). The values for several key measures of consistency and diversity of tune types are printed – we continue to refer back to Figure 3.1 throughout this section as we explain the main measures that we used to index consistency and diversity of tune types and of voice types.

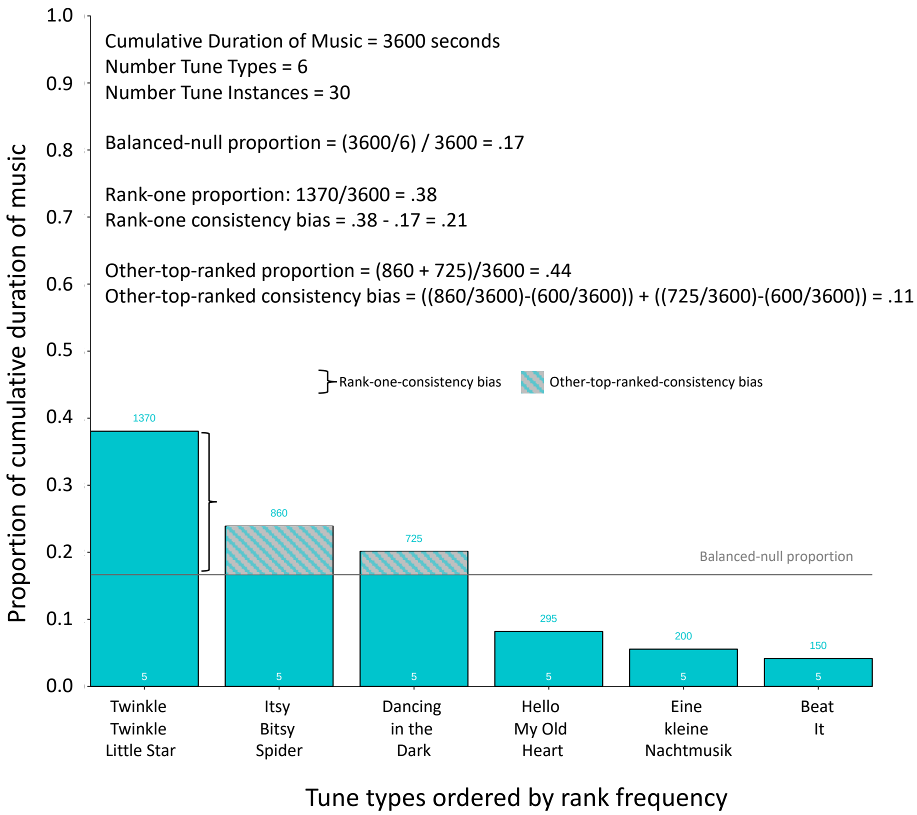

Figure 3.1. A hypothetical frequency distribution of tune types that an infant could have encountered in their daily music. The depicted key properties of this frequency distribution and main measures of consistency and diversity of tune types are fully explained in the main text.

Quantifying diversity. We reported the range of tune (voice) types as a measure of diversity. Another common method to measure diversity, especially in the domain of language, is type-token ratio (TTR; e.g., Richards, 1987). In language, this is calculated as the number of unique words divided by the total number of all words that occurred in a section of speech or text. TTR is a measure of diversity per unit time. In our research, we were interested in characterizing diversity at the day-long timescale. When the unit of time was the full day, the TTR of tunes and voices reduced to the range.

Quantifying consistency. To determine whether some tune types or some voice types were more available than others (i.e., exhibited consistency), we first visually inspected the frequency distributions to determine the proportion of frequency distributions that appeared to be balanced versus biased. We confirmed whether each infant's actual frequency distributions of tune types and of voice types reflected a balanced frequency distribution using chi-square goodness-of-fit tests. Then, we examined the components of the frequency distributions, calculating several additional measures to index consistency of tune types and of voice types in infants' everyday musical input.

Chi-square goodness-of-fit tests. As a coarse measure of consistency of tunes and voices in the music of infants' everyday environments, we examined each infant's actual frequency distributions of tune types and voice types relative to estimated balanced frequency distributions using a chi-square goodness-of-fit test. The infant's actual data served as the observed data. A balanced frequency distribution based on number of tune (voice) types in each infant's own music data served as the expected data. In each test, the null hypothesis was that the infants' actual data reflected a balanced frequency distribution; the alternative hypothesis was that the infant's actual data did not reflect a balanced frequency distribution. As an example, let's consider the frequency distribution in Figure 3.1. This frequency distribution contained 3,600 total seconds of music distributed across 6 tune types, so the expected durations of each tune type would be 600 (i.e. 3,600/6) for a balanced frequency distribution. Given 5 degrees of freedom and significance value of $p < .05$, the critical value of the chi-square distribution would be 11.07. The test statistic for the actual frequency distribution of tune types in Figure 3.1 would be 2,494.33, so we could reject the null hypothesis. This frequency distribution does not reflect a balanced frequency distribution. We conducted this test separately for each infant's frequency distribution of tune types and voice types. We summarized these tests by reporting the proportion of infants that had a significant chi-square goodness-of-fit test, indicating that their actual data did not reflect a balanced frequency distribution. In other words, these infants would have frequency distributions of tune types or voice types that were biased in favor of consistency.

Rank-one proportion. Next, we focused on the single tune (voice) type that occurred for the longest duration (i.e., the "rank-one type"). We calculated the proportion of the frequency distribution of tunes (voices) that was accounted for by the rank-one tune (voice) type – the *rank-one proportion*. This measure is an index of consistency, because it reveals how much of the distribution was the same specific tune (voice). A higher proportion indicates greater consistency of one specific tune (voice).

Balanced-null proportion. For each infant, we then calculated the *balanced-null proportion* – the proportion of the cumulative duration of music that would have been accounted for by one single tune (voice) type if the frequency distribution were balanced. In a balanced null distribution, all tune (voice) types would occur for an equal duration. To compute *balanced-null proportion*, we used the cumulative duration of music and the total number of tune (voice) types from infants' actual frequency distributions. We first divided the cumulative duration of music by the total number of tune (voice) types to determine the duration of any one tune (voice) type in the balanced null distribution.

Then, we divided this resulting duration by the cumulative duration of music to determine the proportion of the balanced null distribution accounted for by any one tune (voice) type. This value is sensitive to the total range of tune (voice) types – the greater the number of tune (voice) types, the lower the *balanced-null proportion*.

Rank-one consistency bias. We then compared infants' actual data with the balanced null distributions, subtracting the *balanced-null proportion* from the *rank-one proportion*. If an infant encountered each tune (voice) type equally in their everyday musical input, then this difference score would equal zero. If an infant encountered their rank-one tune (voice) type for a longer duration than other tune (voice) types, then this difference score would be greater than zero. Because larger difference scores indicate greater consistency of the top tune (voice) type, we refer to this difference as the *rank-one consistency bias*. As an example, let's consider Figure 3.1 – the rank-one tune type of the frequency distribution, *Twinkle Twinkle Little Star*, occurs for 1,370 cumulative seconds, accounting for over one-third (.38) of the cumulative duration of all music (depicted by the height of the left-most bar in Figure 3.1). Given 6 tune types, the *balanced-null proportion* is .17 (depicted by the gray horizontal line in Figure 3.1). The difference between these two values – the *rank-one consistency bias* – is .21.

Other-top-ranked proportion. Next, we zoomed out beyond the rank-one tune (voice) type to examine the other tune (voice) types that occurred that occurred for a longer duration than would be expected if the frequency distribution of tune (voice) types were balanced. Thus, these "other-top-ranked" tune (voice) types did not include the rank-one tune (voice) type. Examining the distribution of the other-top-ranked tune (voice) types provided a more global measure of consistency relative to examining only the rank-one tune (voice) type. For the other-top-ranked tune (voice) types, we calculated the same set of measures as we did for the rank-one tune (voice) type. First, we determined the *other-top-ranked proportion*: the sum of each top-ranked tune (voice) type's cumulative duration out of the cumulative duration of music. A higher proportion indicated greater consistency of tune (voice) types.

Other-top-ranked consistency bias. In addition, we compared each infant's actual frequency distribution of tune (voice) types to a balanced null distribution. We calculated the difference between the proportion of each other-top-ranked tune (voice) type and the *balanced-null proportion.* The sum of these differences was the *other-top-ranked consistency bias*. Larger difference scores indicated greater consistency of tune (voice) types, such that some tune (voice) types were more available than other tune (voice) types. Returning to Figure 3.1, the frequency distribution has 2 other-top-ranked tune types – *Itsy Bitsy Spider* and *Dancing in the Dark*. The sum of the cumulative duration of these 2 other-top-ranked tune types is 1,585 seconds. Therefore, the *other-top-ranked proportion* is .44 (i.e., 1,585/3,600). If this frequency distribution were balanced, then each tune type would account for .17 of the cumulative duration of music (the *balanced-null proportion*). The differences between the proportion of each of the other-top-ranked tune types and the balanced-null proportion are .072 and .035, respectively. The sum of these is .11 – the *other-top-ranked consistency bias*. This value indicates that some tune types occurred for longer durations than others.

By examining these multiple measures, we sought converging evidence about the combination of consistency and diversity of tune types and of voice types that infants encountered in their everyday music. To address our questions about the social quality of infants' daily musical input, we used the following measures.

Proportion of live-vocal musical input. We first calculated the proportion of music with exactly one tune that was coded as both 'live' and 'vocal', and the proportion of music with exactly one voice that was coded as 'live' in each infant's day. We defined "live-vocal" music inclusively, such that it could contain not only music that had live, unaccompanied voices singing to infants, such as mom singing *Twinkle Twinkle Little Star*, but also music that had live, voices plus recorded and/or instrumental music, such as mom singing along with a recording of Bruce Springsteen singing *Dancing in the Dark*.

Consistency and diversity of live-vocal tunes and live voices. To characterize consistency and diversity of specific live-vocal tunes and specific live voices that occurred during infants' musical input, we calculated the same set of measures that we calculated to characterize the frequency distributions of all tune types and all voice types: *chi-square goodness-of-fit test, rank-one proportion, balanced-null proportion, rank-one consistency bias, other-top-ranked proportion, balanced-types-proportion,* and *other-top-ranked consistency bias.*

Live-vocal music in the mix of all musical input. To assess how live-vocal music occurred in the mix of all music, we first calculated the proportion of the total number of tune (voice) types that occurred as live-vocal music for each infant. This value was based on the number of tune (voice) types, rather than the duration of music. Therefore, if any duration of the cumulative seconds of the tune (voice) type had been coded as both 'live' and 'vocal', then that tune (voice) type was counted as being live-vocal. In this way, this measure captured the range of tune (voice) types that were live-vocal – an index of diversity. This *live-vocal-type diversity* ranged from 0 to 1. A value of 1 would indicate maximal diversity, such that live-vocal music was distributed across all tune (voice) types. As this value decreased, it indicated that live-vocal music was distributed across fewer total tune (voice) types, reflecting less diversity of live-vocal music across tune (voice) types. A value of exactly 0 would indicate that no tune (voice) types were live-vocal. Next, for each infant, we zoomed in on the single top tune and single top voice. We determined whether the rank-one tune (voice) type occurred as live-vocal music for any duration in each infant's day. We then calculated the proportion of infants who encountered their rank-one tune (voice) type as live-vocal music.

Because several properties of the recordings varied across infants (i.e., duration recorded, duration coded, duration of music, etc.), we calculated and reported proportions for dependent variables when possible. For example, we reported the cumulative duration of voice instances out of the cumulative duration of music, rather than the raw cumulative duration of voice instances. (see Supplemental Materials for raw counts of the main dependent measures).

RESULTS

------------------------------------ TUNES MUSIC CORPUS ------------------------------------

In this section of the thesis, we examined consistency, diversity, and social quality in the tunes that occurred in their everyday musical input. For the remainder of this "Tunes Music Corpus" section, when we refer to "music", we mean the cumulative duration of the music bouts that had exactly one tune.

How many tune types occurred in how many seconds of music?

We discovered that individual infants encountered from 14 to 213 different tune types (Median = 51, SD = 41.74). All infants encountered more than one specific tune in the music of their day, meaning that all infants encountered at least some diversity of tune types (see Figure 3.2). These tune types occurred in cumulative durations of music ranging from 361 seconds to 9,309 seconds (Median = 2,404 seconds, SD = 2,326.61 seconds; see Figure 3.3).

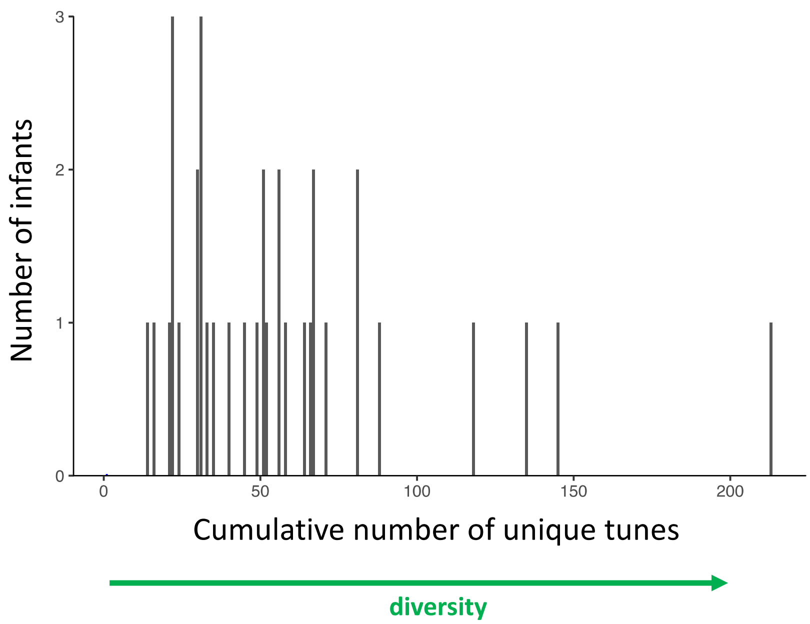

Figure 3.2. All infants encountered some diversity of tune types (i.e., more than one specific tune) in the music of their days.

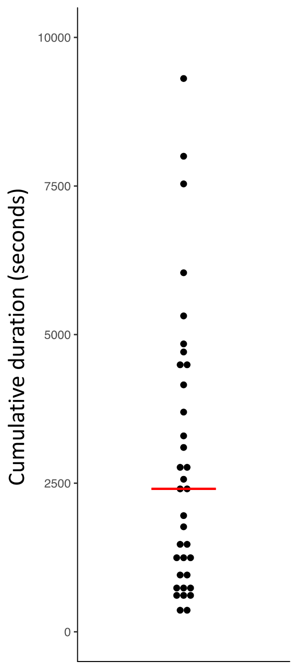

Figure 3.3. The cumulative duration of music with exactly one tune ranged from 361 seconds to 9,309 seconds across infants. The red line indicates the median duration (2,404 seconds). This value was not correlated with the cumulative duration recorded or the cumulative number of coded seconds.

Did infants encounter balanced frequency distributions of tune types?

First, we constructed individual frequency distributions of tune types for each infant (see Figure 3.4). From a simple visual inspection, none of these individual frequency distributions appeared to be balanced such that each tune type occurred for an equal duration. Next, we addressed this question quantitatively, using chi-square goodness-of-fit tests to compare each infant's actual frequency distribution of tune types to a balanced null distribution with the same number of tune types. These tests confirmed that none of the infants' actual frequency distributions reflected balanced frequency distributions (i.e., $p < .05$ for each infant's test). Because infants' frequency distributions of tune types were not balanced, we next calculated multiple measures to assess in what ways they were biased, indexing the combination of consistency and diversity of tune types that infants encountered in their daily musical input.

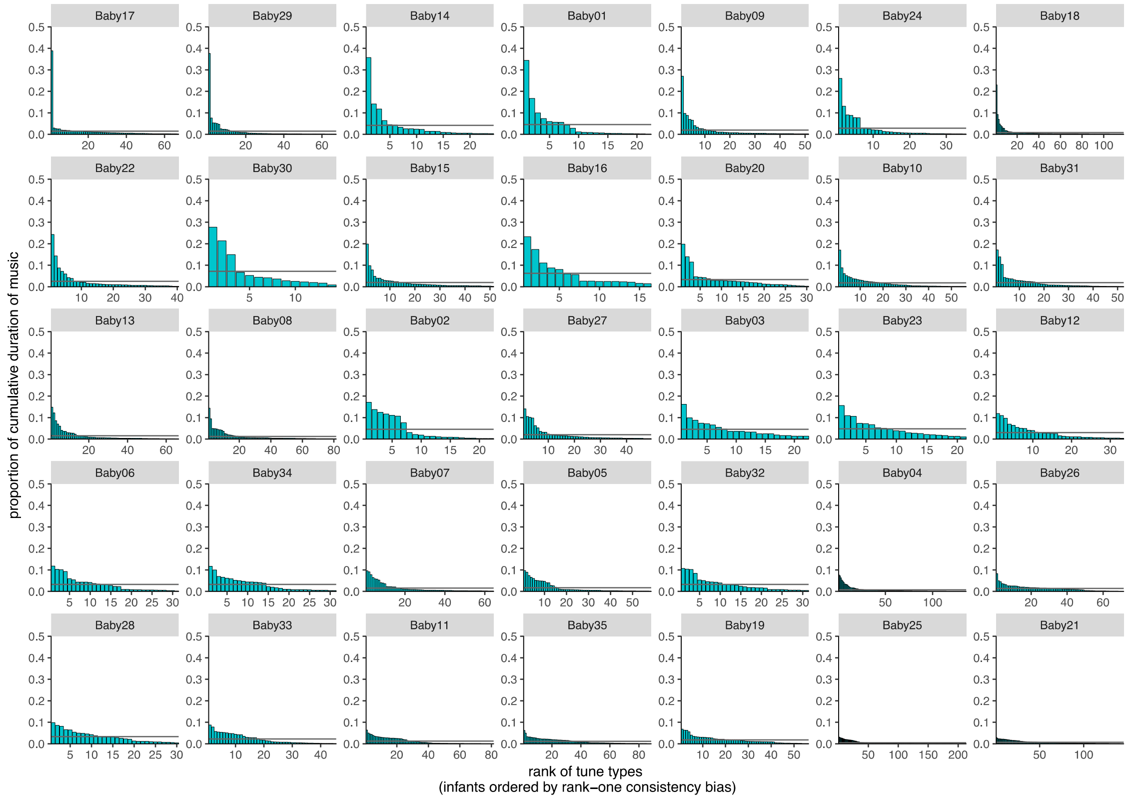

Figure 3.4. Based on visual inspection, no infant encountered a balanced distribution of tune types. Each plot depicts the frequency distribution of tunes for an individual infant. The unique tune types are on the x-axis ordered by their rank-frequency. The height of each bar represents the proportion of the cumulative duration of all tunes instances that was accounted for by each specific tune type. Note: the maximum value displayed on the y-axis is .50, but the maximum possible value is 1 (no tune type accounted for > .5 of any cumulative duration). The gray horizontal line shows the balanced-null proportion. The individual plots are ordered by the rank-one consistency bias (largest to smallest, top-left to bottom-right).

To what extent was one specific tune more available relative to a balanced null?

Here, we zoomed in to examine the single tune type that occurred for the longest duration. This "rank-one" tune type accounted for .15 (Median, SD = .10, range: .03-.39) of the cumulative duration of music (*rank-one proportion*). Thus, one specific tune accounted for 15% of the cumulative duration of all tunes. We next calculated, for each infant, the proportion of the cumulative duration of music that would be accounted for by one tune type if each tune type occurred for an equal duration (*balanced-null proportion,* Median = .02, SD = .02). Then, we examined this value relative to that of infants' real frequency distributions of tune types. This difference – the *rank-one-consistency-bias* for tunes – ranged from .02 to .37 across infants (Median = .12, SD = .09; see Figure 3.5A). A paired t-test confirmed that the actual proportion of the rank-one tune was significantly greater than the *balanced-null proportion* ($t(34) = 9.08$, $p < .001$). These findings revealed that one specific tune was more available to infants in their everyday music than would be expected given a balanced null – this reflected consistency, because a considerable proportion of infants' daily musical input contained the same specific tune.

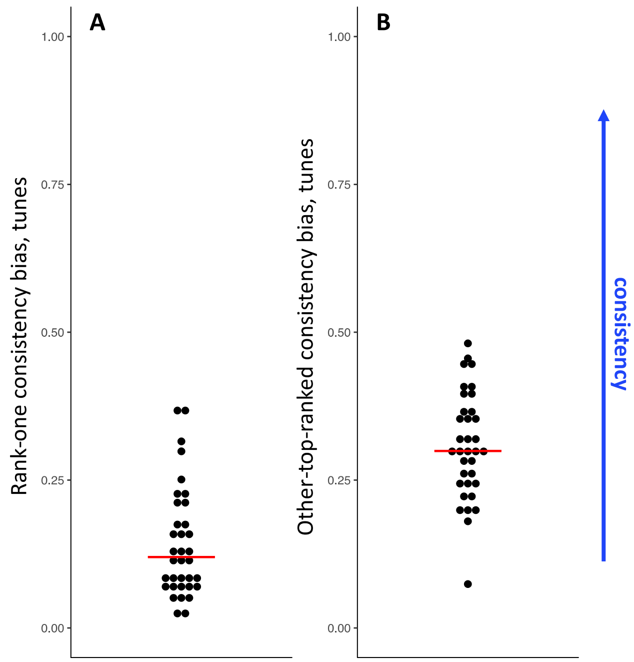

Figure 3.5. Consistency bias in the tunes infants encountered. Compared to a balanced null, (A) the single top tune and (B) the other top tunes were very much more available. The red lines show the median values.

The *rank-one proportion, balanced-null proportion,* and resulting *rank-one consistency bias* are each sensitive to the total number of types an infant encountered. In Figure 3.6, we show how the *balanced-null proportion* (gray line) and the *rank-one proportion* (points) for tune change with respect to the cumulative number of tune types in the music of each infant's day. The distance between each point and the balance-null proportion line depicts the *rank-one consistency bias* for each infant for the tunes that occurred in the music of their recorded days. By definition, the balanced null proportion had to decrease as the number of unique tunes increased. It did not have to be the case that the rank-one proportion or the rank-one consistency bias also decreased as the number of unique tunes increased. This is reasonable, given that we analyzed tunes and voice in music that occurred during a restricted duration (i.e., one full day).

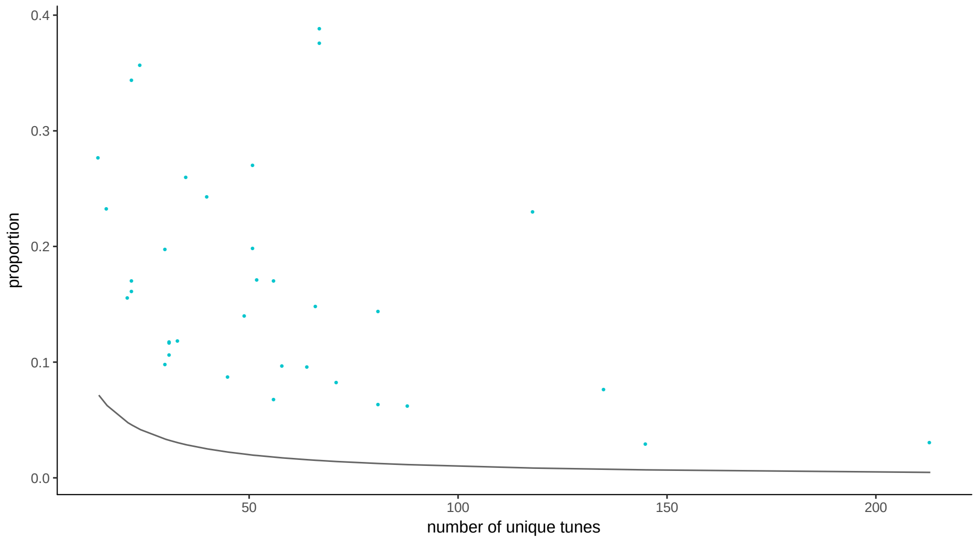

Figure 3.6. The main measures of consistency were each sensitive to the number of unique tunes. The rank-one proportion for tunes (teal points) and the balanced-null proportion are shown with respect to the cumulative number of unique tunes that occurred in each infant's day. The distance from each point to the balanced-null line represents that infant's consistency bias of the rank-one tune.

To what extent were some tune types more available than other tune types?

Finally, we zoomed out to examine the other tune types that accounted for a greater proportion of the cumulative duration of tune instances than would be expected if the frequency distributions of tune types were balanced (i.e., tune types in the top of the frequency distribution, not including the rank-one tune type). The cumulative duration of these "other-top-ranked" tune types accounted for over half of the cumulative duration of all tune instances (Median = .61, SD = .12, range: .31 - .79; *other-top-ranked proportion*). We examined the proportion of each other-top-ranked tune type relative to the balanced-null proportion. The resulting *other-top-ranked consistency bias* was .30

(Median, SD = .09; range: .07 - .48; see Figure 3.5B), which means the other top tunes occurred for a cumulative longer duration compared to a balanced null model. A paired t-test confirmed that the *other-top-ranked proportion* was significantly greater than would be expected given a balanced null distribution ($t(34) = 31.69$, $p < .001$). The *other-top-ranked consistency bias* was greater than zero for all infants, which means that some specific tunes were more available than other specific tunes in infants' everyday musical input. Thus, all infants encountered some degree of consistency of the tune types in their daily music. Figure 3.5 shows both measures of consistency bias for tunes in infants' everyday musical input.

---------------------------------- VOICES MUSIC CORPUS ------------------------------------

We now address the same set of questions about consistency, diversity, and social quality of voices that infants encountered in the music of their everyday lives. For the remainder of this "Voices Music Corpus" section, when we refer to "music", we mean the cumulative duration of the music bouts that had exactly one unique voice.

How many voice types occurred in how many seconds of music?

To answer this first question, individual infants encountered from 1 to 66 different voice types (Median = 8, SD = 13.16; see Figure 3.7). Most infants encountered more than one specific voice in the music of their day, meaning that most infants encountered at least some diversity of voice types. One infant encountered exactly one specific voice in their musical input, reflecting maximal consistency of voice types. The cumulative duration of music ranged from 108 seconds to 5,616 seconds (Median = 1,077 seconds, SD = 1,619.76 seconds; see Figure 3.8).

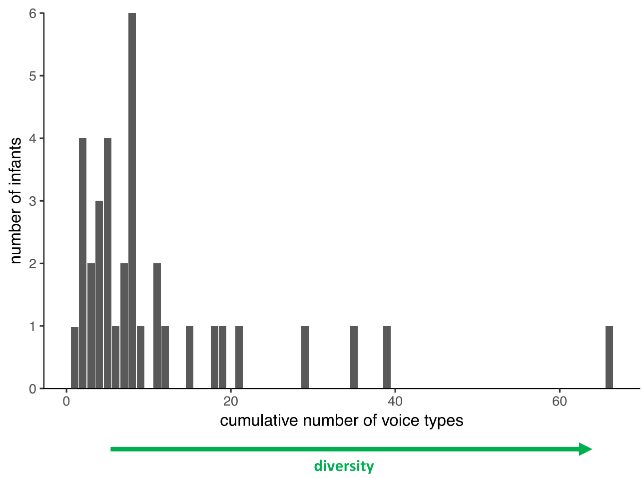

Figure 3.7. Almost all infants encountered more than one voice type in the music of their days.

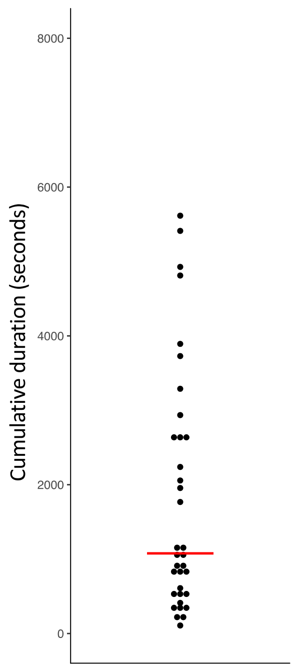

Figure 3.8. The cumulative duration of music with exactly one voice ranged from 108 seconds to 5,616 seconds across infants. The red line indicates the median (1,077 seconds). This value was not correlated with the cumulative duration recorded or the cumulative duration of coded seconds.

Did infants encounter balanced frequency distributions of voice types?

To answer this question, we calculated several measures to index consistency and diversity of voice types that infants encountered in their everyday music. First, we constructed individual frequency distributions of voice types for each infant (see Figure

3.9). Based on a visual inspection, none of these individual frequency distributions of voice types appeared to be balanced such that each voice type occurred for an equal duration. Chi-square goodness-of-fit tests comparing each infant's real frequency distribution of voice types to a balanced null distribution with the same number of voice types confirmed that none of infants' real frequency distributions were consistent with a balanced null model (i.e., $p < .05$ for each infant's test).

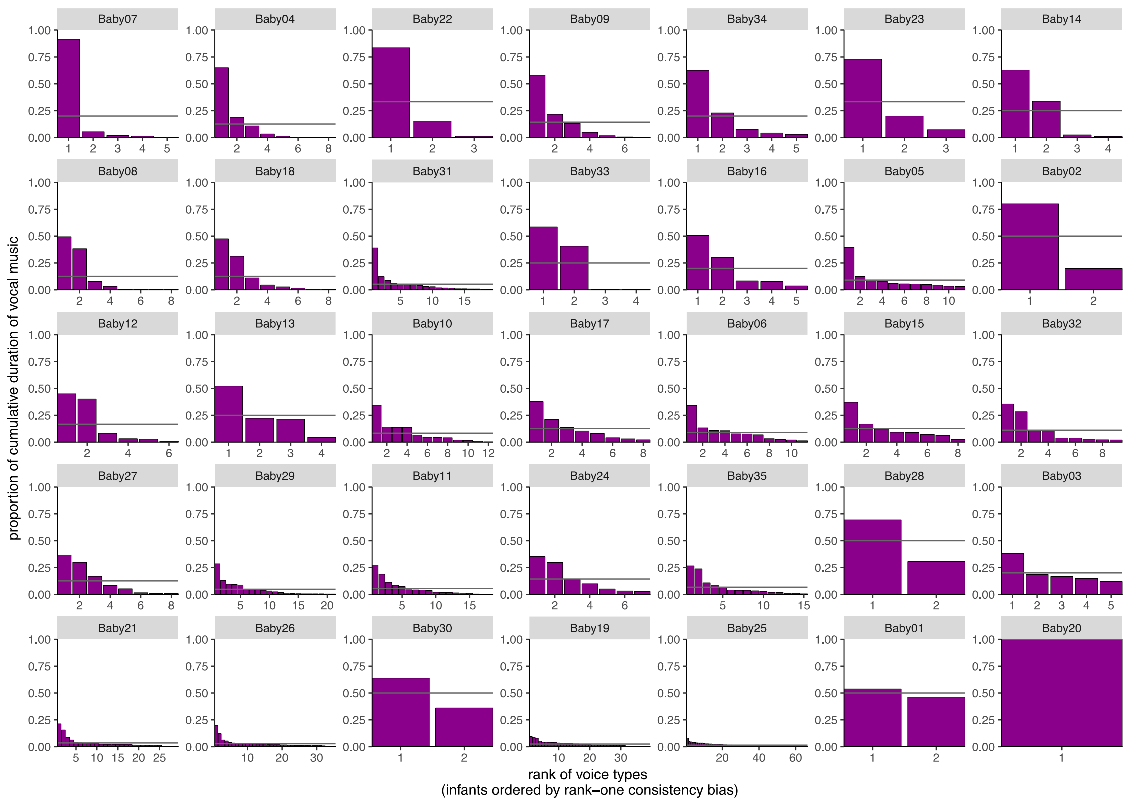

Figure 3.9. From a visual inspection, no infant encountered a balanced distribution of voice types. Each plot depicts the frequency distribution of voices for an individual infant. The unique voice types are on the x-axis ordered by their rank-frequency. The height of each bar represents the proportion of the cumulative duration of all voice instances that was accounted for by each specific voice type. The gray horizontal line shows the balanced-null proportion. The individual plots are ordered by the difference between the rank-one proportion and the balanced-null proportion (largest to smallest, top-left to bottom-right).

To what extent was one specific voice more available relative to a balanced null?

As another index of consistency and diversity of voice types, we computed the proportion of the cumulative duration of music that was accounted for by the single top voice type (*rank-one proportion*). For individual infants, the *rank-one proportion* of voices ranged from .08 to 1.0 (Median = .45, SD = .22). Thus, for most infants, one specific voice accounted for almost one-half of the cumulative duration of all voices – this signals consistency of voice types in individual infants' everyday musical input. For each individual infant, we next calculated the proportion of one voice type if each voice had occurred for an equal duration (*balanced-null proportion*, Median = .13, SD = .20), and we compared this value to that of the infant's real frequency distribution of voice types. The difference – the *rank-one-consistency-bias* for voices – ranged from 0 to .71 across infants (Median = .25, SD = .14; see Figure 3.10A). A paired t-test confirms that the actual proportion of the rank-one voice was significantly greater than the *rank-one balanced-null proportion* ($t(33) = 12.06$, $p < .001$).

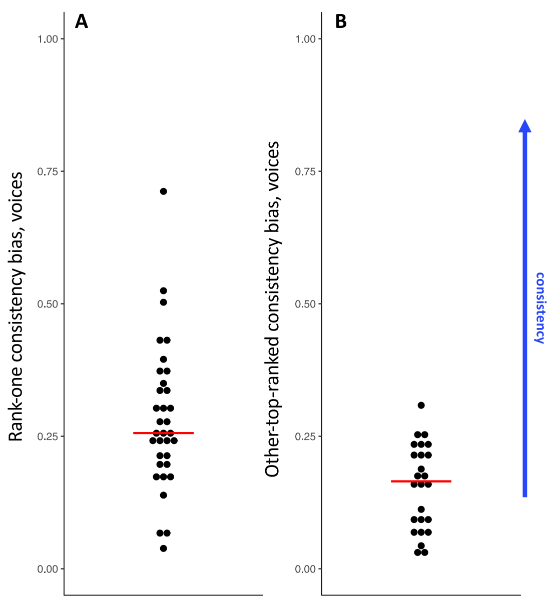

Figure 3.10. Compared to a balanced null, infants' single top-most voice (A) and the other voices that occurred in the top of their frequency distributions of voice types (B) were much more available. The red lines show the median values.

Because the main measures of consistency for voices were each sensitive to the range of voice types, we examined the relationships among the *rank-one proportion* for voices (magenta points), the *balanced-null proportion*, and the cumulative number of unique voices that occurred in each infant's day (see Figure 3.11). The distance from each point to the balanced-null line represents that infant's *rank-one consistency bias* for voices that occurred in the music of their days. We observed the same pattern as we did for tunes, that all three measures decrease as the number of unique voices increases.

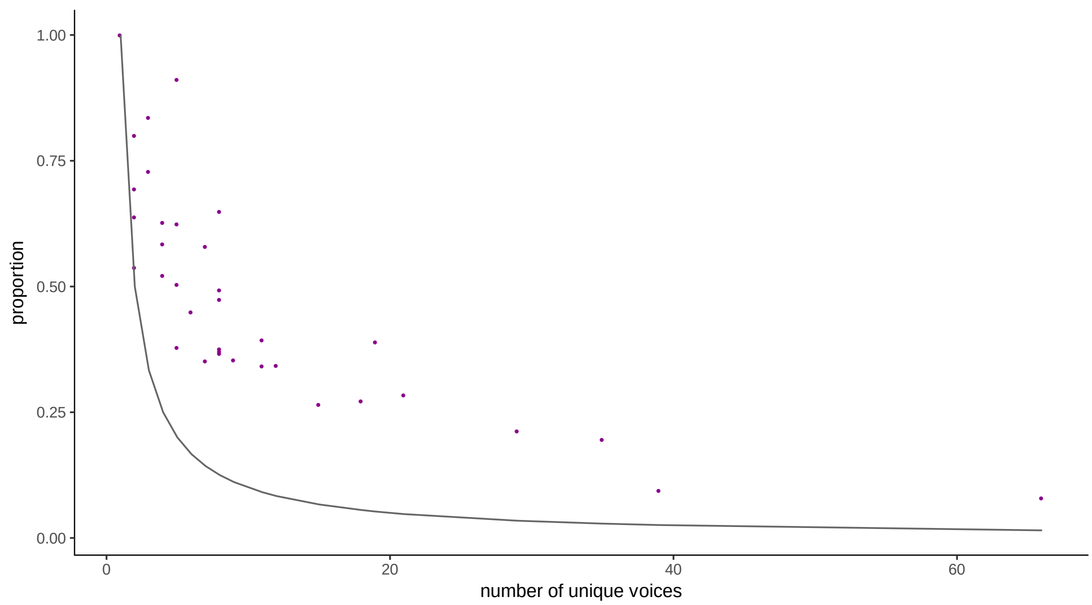

Figure 3.11. The main measures of consistency were each sensitive to the number of unique voices. The rank-one proportion for voices (magenta points) and the balanced-null proportion are shown with respect to the cumulative number of voice types that occurred in the music of each infant's day. The distance from each point to the balanced-null line is that infant's rank-one consistency bias for voices.

To what extent were some voice types more available than other voice types?

Next, we examined the other voice types, in addition to the single top voice, that accounted for a greater proportion of the cumulative duration of voice music than would be expected if each unique voice occurred for an equal duration. Several infants (.29) encountered only one unique voice for a longer cumulative duration than would be expected by the balanced null, and therefore had no additional voice types in the "other-top-ranked" category. These infants were excluded from this analysis. For the remaining infants, the cumulative duration of these “other-top-ranked” voice types accounted for over one-third of the cumulative duration of music (Median = .35, SD = .13, range: .12 - .67; *other-top-ranked proportion*). One-fifth of a balanced null distribution would be accounted for by the same number of voice types as the number of “top-ranked” voice types (Median = .20, SD = .08; range: .09 - .38; *balanced-type-proportion*). The *other-top-ranked consistency bias* – the difference between these two values – was .16 (Median, SD = .08; range: .03 - .31; see Figure 3.10B). A paired t-test confirms that the actual proportion of the other top-ranked voice types was significantly greater than the *balanced-null proportion* ($t(24) = 9.74$, $p < .001$). This confirmed that many infants encountered several specific voices that were more available than other specific voices – a sign of consistency in the voices that infants encountered in their everyday music.

-------------------------------------- LIVE-VOCAL TUNES --------------------------------------

So far, what we have reported about is all kinds of music that occurred in infants’ everyday environments. This music could have been live either with a voice, like mom singing, or without a voice, like dad playing the flute. It could have been recorded music either with voices, like The Beatles singing *Hey Jude* on the radio, or without voices, like Tchaikovsky’s 5th Symphony playing on the radio. It could also have been any combination of these. Now, we are going to focus on only the music that was both live and vocal. This subset of music produced by a live voice represented the high-social-quality musical input infants encountered in their everyday environments. We first determined what proportion of infants’ everyday musical input was coded as both ‘live’ and ‘vocal’. For these analyses, we continued to use the tunes music corpus and the voices music corpus, calculating the proportion of each that was coded as both ‘live’ and ‘vocal’ (“live-vocal”).

What proportion of infants’ everyday music was live-vocal?

The proportion of the cumulative duration of music that was coded as ‘live’ and ‘vocal’ ranged from .05 to .91 across infants (Median = .40, SD = .26; see Figure 3.12). Each infant encountered musical input that contained a live-vocal tune.

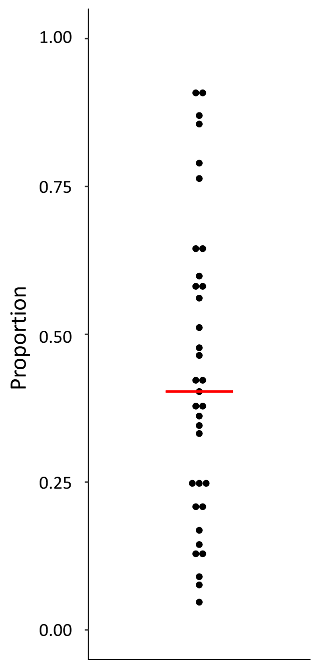

Figure 3.12. Each infant encountered a non-zero proportion of music with live-vocal tunes. The red line depicts the median proportion (.40) of the cumulative duration of music that contained a live-vocal tune.

How many live-vocal tunes occurred in how many seconds of music?

Within their daily music with exactly one live-vocal tune, infants encountered from 3 to 104 unique live-vocal tunes (Median = 16, SD = 20.20; see Figure 3.13). Thus, all infants encountered some diversity of live-vocal tune types (i.e., more than one specific live-vocal tune) in the music of their days. The cumulative duration of music with exactly one live-vocal tune varied from 77 to 2,368 seconds (Median = 623 seconds; SD = 740.94 seconds) across individual infants' recorded days (see Figure 3.14).

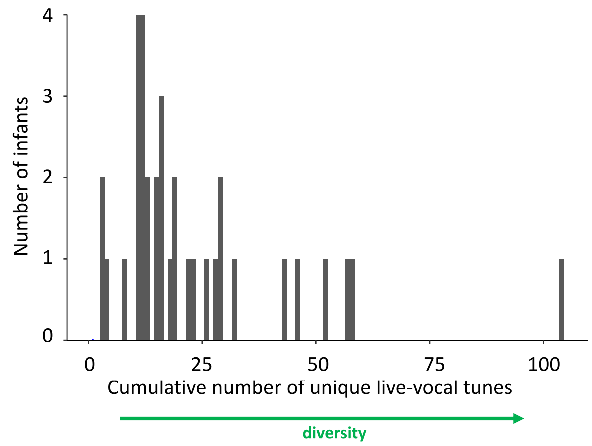

Figure 3.13. All infants encountered some diversity of live-vocal tune types (i.e., more than one specific live-vocal tune) in the music of their days.

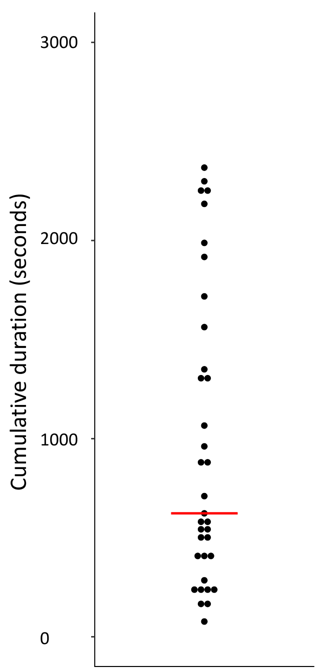

Figure 3.14. The cumulative duration of music with exactly one live-vocal tune ranged from 77 seconds to 2,368 seconds across infants. The red line indicates the median (623 seconds). This value was not correlated with the cumulative duration recorded or the cumulative duration of coded seconds.

To what extent were some live-vocal tune types more available than others?

To address whether infants encountered some live-vocal tune types more than others, we calculated the same set of measures that we did for examining consistency and diversity of all tune types. We first visually examined the frequency distributions of live-vocal tune types for each infant (see Figure 3.15), noting that none of the frequency distributions of live-vocal tune types appeared to be balanced. Chi-square goodness-of-fit tests comparing each infant's actual frequency distribution of live-vocal tune types to a balanced null distribution confirmed that none of the infants' actual frequency distributions of live-vocal tunes reflected balanced frequency distributions (i.e., $p < .05$ for each infant's test).

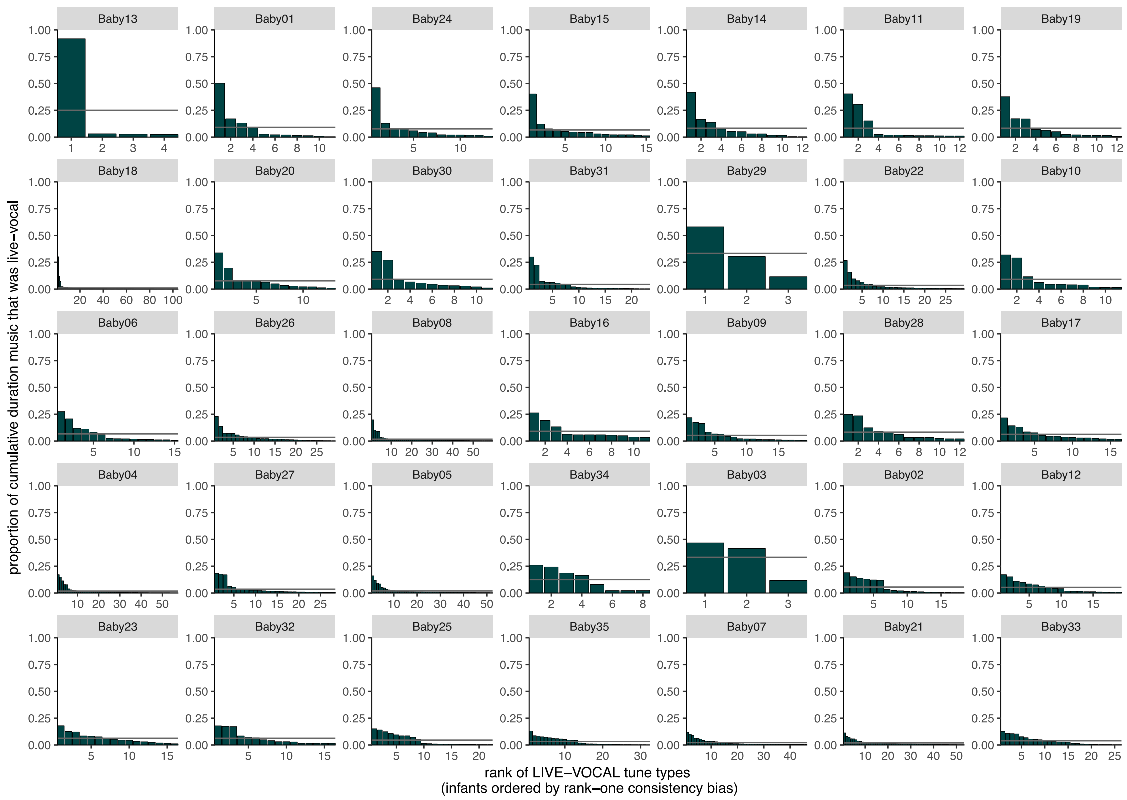

Figure 3.15. Based on a visual inspection, no infant encountered a balanced distribution of live-vocal tune types. Each plot depicts the frequency distribution of live-vocal tune types for an individual infant. The unique live-vocal tunes are on the x-axis ordered by their rank-frequency. The height of each bar represents the proportion of the cumulative duration of live-vocal music that was accounted for by each specific live-vocal tune type. The gray horizontal line shows the proportion each live-vocal type would be if the distribution were balanced (based on the number of voice types and cumulative duration of voice instances for each infant's voice dataset). The individual plots are ordered by the difference between the rank-one proportion and the balanced-null proportion (largest to smallest, top-left to bottom-right).

Next, to quantify the degree of consistency of the live-vocal tune types, we calculated the proportion of the cumulative duration of live-vocal music that was accounted for by the rank-one live-vocal tune type (*live-vocal rank-one proportion*), which ranged from .09 to .92 across infants (Median = .26, SD = .16). In other words, one specific live-vocal tune accounted for about a quarter of the cumulative duration of all live-vocal music. We then compared this value to the proportion of one live-vocal tune type if each live-vocal tune occurred for an equal duration ("balanced-null proportion", Median = .06, SD = .08). The *live-vocal rank-one consistency bias* ranged from .07 to .67 across infants (Median = .17, SD = .12). We confirmed with a paired t-test that the actual proportion of the rank-one live-vocal tune was significantly greater than the proportion of one live-vocal tune type given a balanced distribution ($t(34) = 10.52$, $p < .001$; see Figure 3.16A).

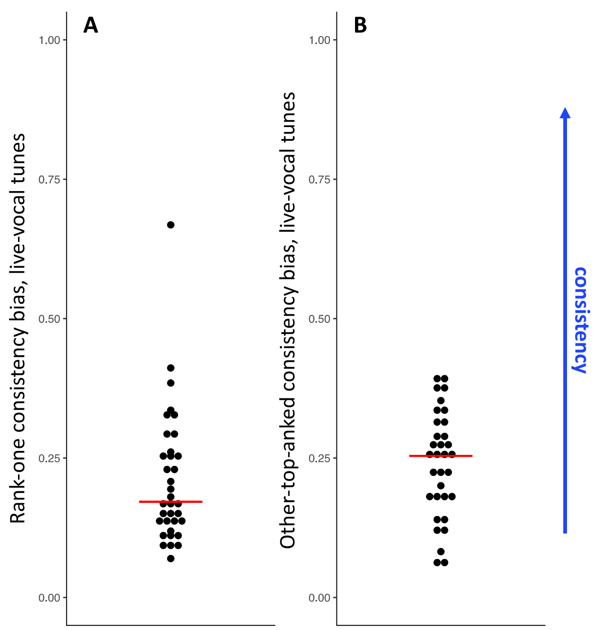

Figure 3.16. Consistency bias in the live-vocal tunes infants encountered. Compared to a balanced null, (A) the single top live-vocal tune and (B) the other top live-vocal tunes were very much more available. The red lines indicate the median values.

As we did for tunes and voices, we also examined the relationship among the *rank-one proportion*, *balanced-null proportion*, *rank-one consistency bias* for live-vocal tunes with respect to the cumulative number of live-vocal tune types that occurred in the music of infants' days (see Figure 3.17). This revealed that even as the number of unique live-vocal tunes increased, the rank-one live-vocal tune still accounted for about 20% of the live-vocal music, yielding relatively higher rank-one consistency biases.

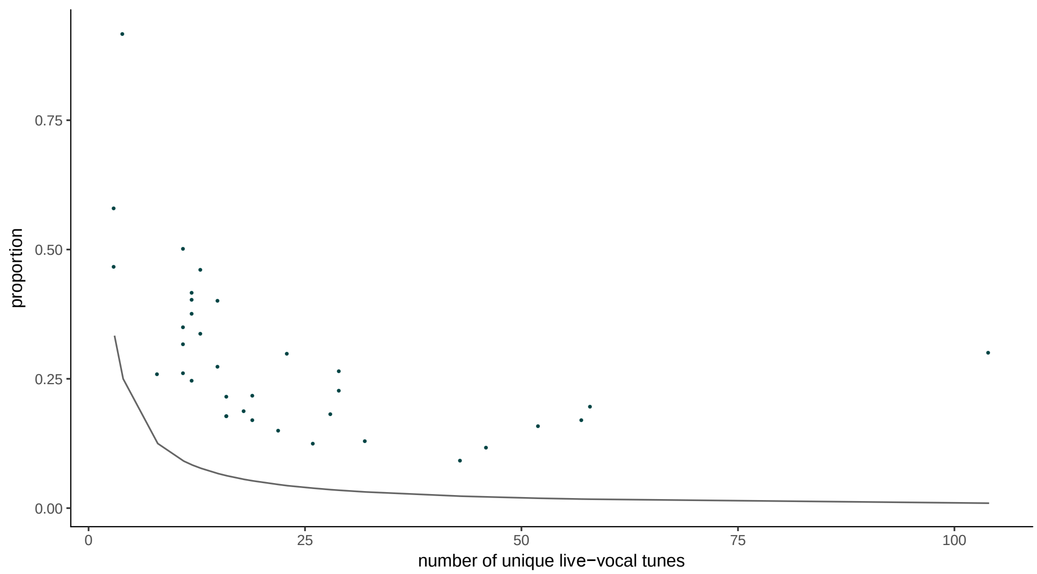

Figure 3.17. The rank-one proportion (dark teal points), the balanced-null proportion (gray line), and the rank-one consistency bias (distance between points and the balanced-null line) were each sensitive to the cumulative number of unique live-vocal tune types that occurred in the music of infants' recorded days.

Then, we zoomed out to examine the other live-vocal tune types that were in the top of the frequency distributions. These live-vocal tunes accounted for .48 (Median, SD = .15, range: range: .20 - .76; *live-vocal other-top-ranked proportion*) of the cumulative duration of live-vocal music. This means that nearly half of the cumulative duration of live-vocal music was accounted for by tune types that occurred for longer durations than would be expected given balanced null distributions. Compared to the balanced null proportion (Median = .24, SD = .08), the *live-vocal other-top-ranked consistency bias* was .25 (Median, SD = .10, range: .06 - .39; see Figure 3.16B). We confirmed this difference with a paired t-test ($t(32) = 14.54$, $p < .001$). These findings provide additional evidence that infants encountered at least one live-vocal tune type for a longer duration than would be expected if they encountered each live-vocal tune type for an equal duration. In other words, the frequency distributions of live-vocal tune types exhibited a consistency bias, some specific live-vocal tunes were more available than others in infants' everyday musical input.

How did live-vocal tune types fit in the mix of all tune types?

We next examined how live-vocal tunes occurred in the mix of all the tunes in infants' everyday musical input (see Figure 3.18). As the first index of how live-vocal music was distributed across tune types, we calculated the proportion of all tune types that were coded as containing both live and vocal music. This value ranged from .04 to .90 (Median = .43, SD = .24). In other words, nearly half of all tune types contained live-vocal music. We also discovered that the rank-one tune type in the frequency distributions of all tune types was live-vocal for 15 of the infants (.40). These results indicate that live-vocal tune types were not always at the top of the mix of all tune types. Only for some infants did their daily musical input exhibit a consistency bias towards one specific live-vocal tune type.

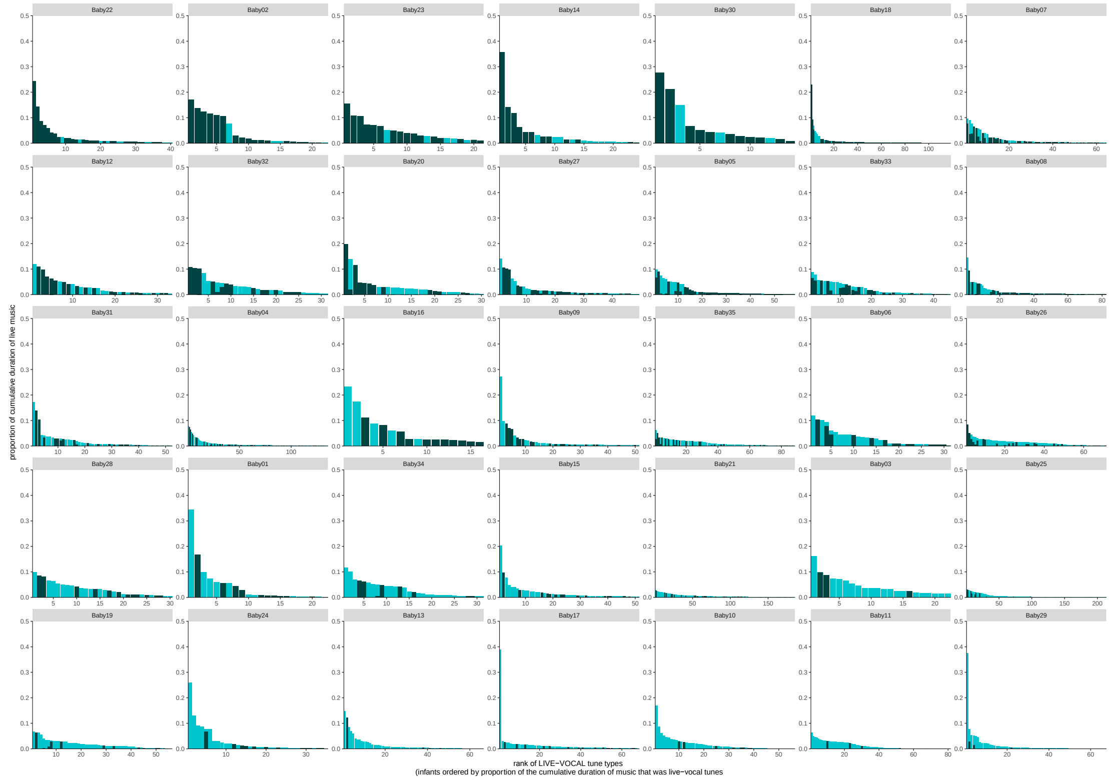

Figure 3.18. Some infants encountered live-vocal tunes in the top of their frequency distributions of all tune types. The rank-one tune type occurred as live and vocal for 14 infants. Each plot depicts the tunes types that occurred in the music of an individual infant's day. Each bar represents one unique tune, and the darker bars depict the unique tune that were live and vocal. The height of the bars indicates the proportion of the cumulative duration of music that was accounted for by each tune type.

-- LIVE VOICES --

We now answer the same set of questions about the live voices that produced music in infants' everyday lives.

What proportion of infants' everyday music was live?

The proportion of the cumulative duration of music that was coded as 'live' ranged from .07 to 1 across infants (Median = .58, SD = .29). As evident in Figure 3.19, every infant encountered a non-zero proportion of musical input produced by a live voice.

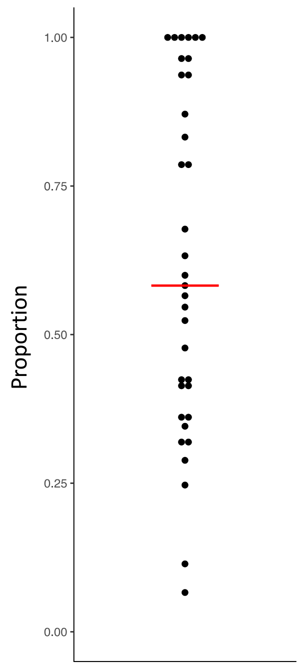

Figure 3.19. Every infant encountered a non-zero proportion of music containing a live voice. The median (.58) proportion of the cumulative duration of music that contained a live voice is shown by the red line.

How many live voice types occurred in how many seconds of music?

Within their daily music with exactly one live voice, infants encountered from 1 to 8 unique live voices (Median = 2, SD = 1.30; see Figure 3.20). Two infants encountered exactly 1 live voice type in their daily music. The remaining infants all encountered some diversity of unique live voices (i.e., more than one live voice type) producing the music that occurred in their recorded days. Across individual infants, the cumulative duration of music with exactly one live voice ranged from 59 seconds to 3,248 seconds (Median = 528 seconds; SD = 783.69 seconds; see Figure 3.21).

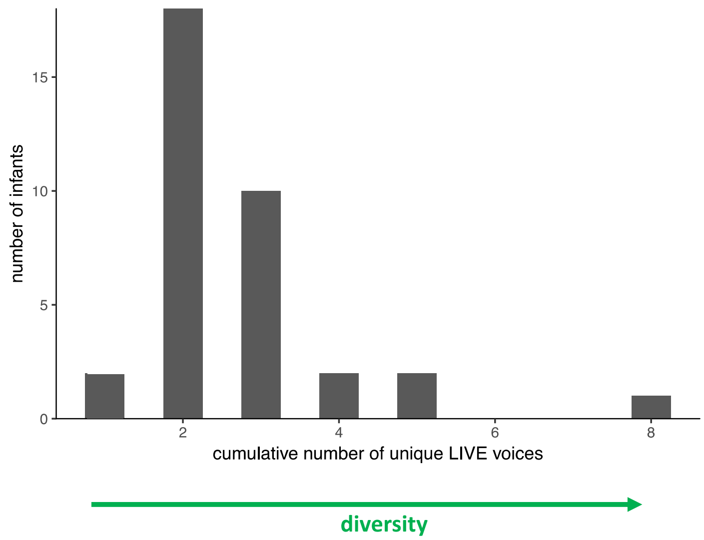

Figure 3.20. All infants encountered some diversity of live voice types (i.e., more than one specific live voice) in the music of their days.

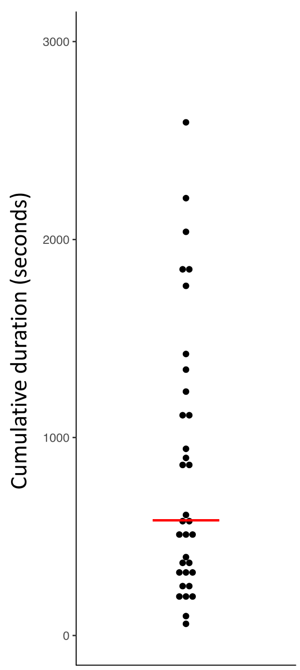

Figure 3.21. The cumulative duration of music with exactly one live voice ranged from 59 seconds to 3,248 seconds across infants. The red line indicates the median (582 seconds). This value was not correlated with the cumulative duration recorded or the cumulative duration of coded seconds.

To what extent were some live voice types more available than others?

To address this question, we computed the same set of measures as when we examined consistency and diversity of all voice types. First, we visually examined the frequency distributions of live voice types for each infant (see Figure 3.22), which revealed that many infants appeared to have one specific live voice that accounted for the bulk of the cumulative duration of music containing live voices. We quantitatively assessed whether infants' actual frequency distributions of live voice types were balanced using chi-square goodness-of-fit tests. Two infants encountered only one live voice type in the music of their everyday lives. These infants were not included in this analysis. For all of the remaining infants, the results of this analysis revealed that the infants' actual frequency distributions of live voice types did not reflect balanced frequency distributions (i.e., $p < .05$ for each infant's test).

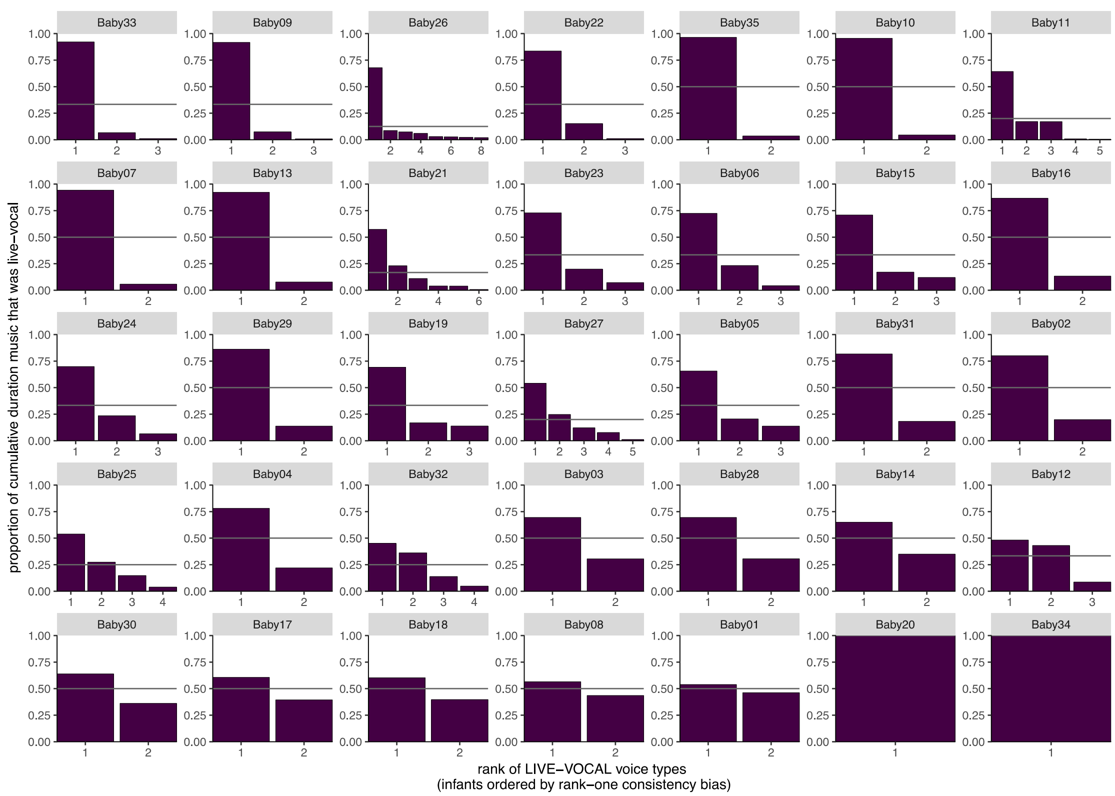

Figure 3.22. By visually inspecting the frequency distributions of live voice types in individual infants' everyday musical input, we observe that many infants encountered biased distributions of live voice types. The unique live voice types are on the x-axis ordered by their rank-frequency. The height of each bar represents the proportion of the cumulative duration of all live voice instances that was accounted for by each specific live voice type. The gray horizontal line shows the proportion each live voice type would be if the distribution were balanced (based on the number of live voice types and cumulative duration of live voice instances for each infant's voice dataset). Individual plots are ordered by the difference between the rank-one proportion and the balanced-null proportion (largest to smallest, top-left to bottom-right).

We further quantified the degree of consistency of the live voice types by calculating the proportion of the cumulative duration of music containing live voices that was accounted for by the rank-one live voice (*rank-one proportion*). For live voices, this value ranged from .45 to 1 across infants (Median = .70, SD = .16). This finding showed that infants encountered high consistency of live voice types in their everyday music, such that one live voice type accounted for the bulk of the total live voice musical input. For each infant, we also calculated the proportion of one live voice type if the frequency distribution of live voice types were balanced (*balanced-null proportion*, Median = .50, SD = .18). We compared this value to that of the infant's real frequency distribution of live voice types. We computed *the rank-one consistency bias* as the difference between the *rank-one proportion* and the *balanced-null proportion*. This *rank-one consistency bias* for live voices ranged from 0 to .59 across infants (Median = .34, SD = .16; see Figure 3.23A). A paired t-test confirmed that the actual proportion of the rank-one live voice was significantly greater than the proportion of one live voice type if the distribution were balanced ($t(32)=12.39$, $p < .001$). In other words, infants' everyday musical input exhibited consistency of live voice types. As we have done previously, we also examined how the *rank-one proportion, balanced-null proportion,* and *rank-one consistency bias* each changed in relation to the cumulative number of live voices (see Figure 3.24). We observed that the rank-one live voice type accounted for at least 50% of the live music for almost all infants, regardless of the cumulative number of live voice types that occurred in the infants' days. This yielded relatively higher rank-one consistency bias values across all cumulative numbers of live voice types.

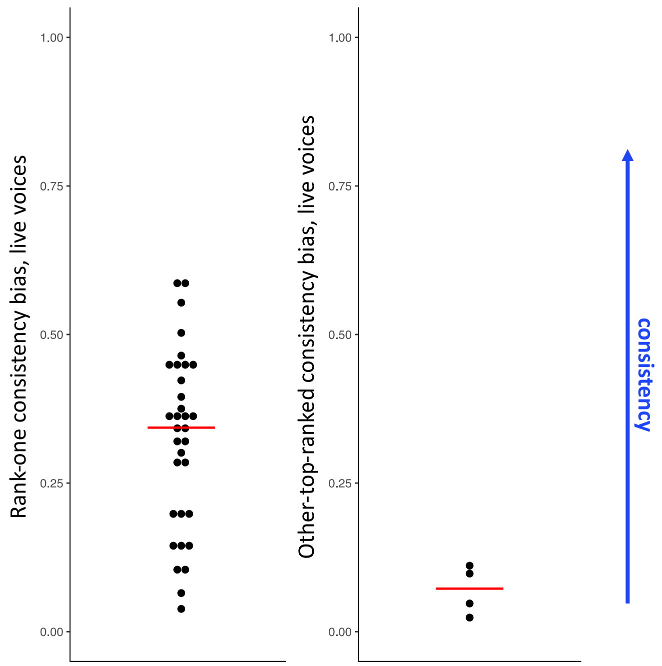

Figure 3.23. Consistency bias in the live voices infants encountered. Compared to a balanced null, (A) the single top live voice was much more available. Only a few infants encountered other unique voices in the top of their frequency distributions (B), and these voices were somewhat more available compared to a balanced null. The red lines represent the median values.

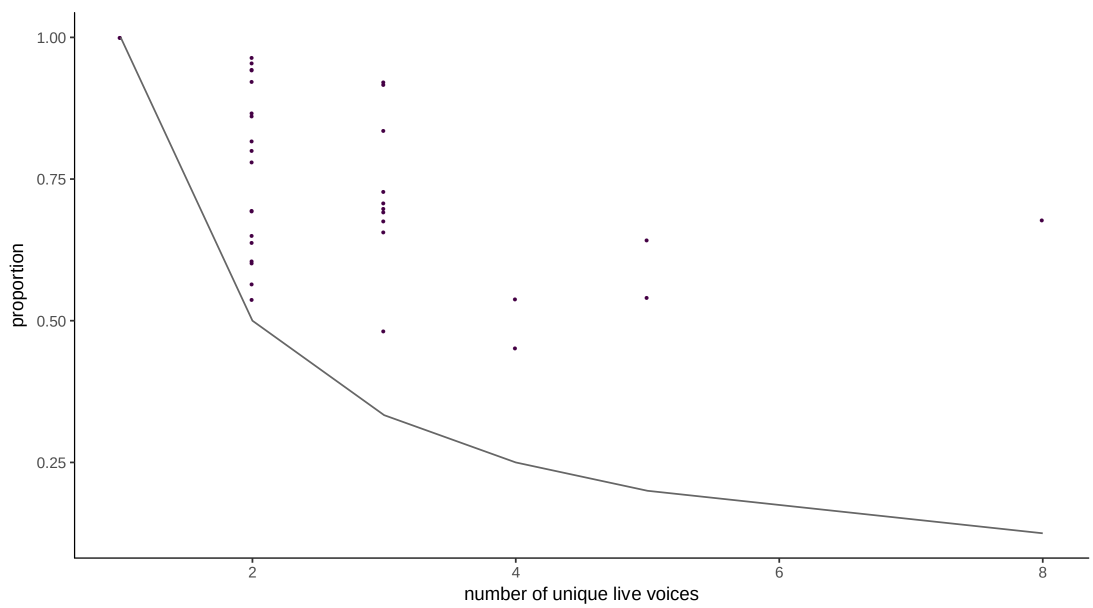

Figure 3.24. The rank-one proportion (dark magenta points), the balanced-null proportion (gray line), and the rank-one consistency bias (distance between points and the balanced-null line) were each sensitive to the cumulative number of unique live voices that occurred in the music of infants' recorded days.

Lastly, we zoomed out to examine the live voice types that were at the top of the frequency distributions of live voices besides the rank-one live voice type. Most infants (.89) only had one unique live voice type that occurred for a longer duration than would be expected if each live voice type occurred for an equal duration. Each of the remaining 4 infants encountered a total of 2 unique live voices in the top of their distribution. This means that for these few infants, in addition to their rank-one live voice type, only one additional voice type occurred for a longer duration than would be expected if each live voice type occurred for an equal duration. For these infants, the "other top" live voice types accounted for from .23 to .43 of the cumulative duration of live music (Median = .32, SD = .08; *live voice other-top-ranked proportion*), meaning that half of the live voice types occurred for a longer duration than would be expected if the frequency distributions of live voice types were balanced. We compared this value to the balanced null proportion (Median = .25, SD = .06), in order to determine the *live-voice other-top-ranked consistency bias*. For live voices, the *other-top-ranked consistency bias* was .07 (Median, SD = .04, range: .02 - .11; see Figure 3.23B).

These results provide further evidence that infants encountered one live voice type for a longer duration than would be expected if they encountered each live voice type for an equal duration. In other words, the frequency distributions of live voice types were biased towards consistency such that one or two live voice types accounted for the bulk of the cumulative duration of live voice instances.

How did live voice types fit into the mix of all voice types?

We next examined how live voices occurred in the broader context of all voices that occurred in music in infants' everyday environments (see Figure 3.25). We first calculated the proportion of all voice types that were coded as containing live music. This value ranged from .06 to 1 (Median = .40, SD = .32). In other words, just under half of all voice types contained live music. We also found that 30 infants (.86) encountered a live voice as their rank-one voice type. These results indicate that live voice types were largely at the top of the mix of all unique voices.

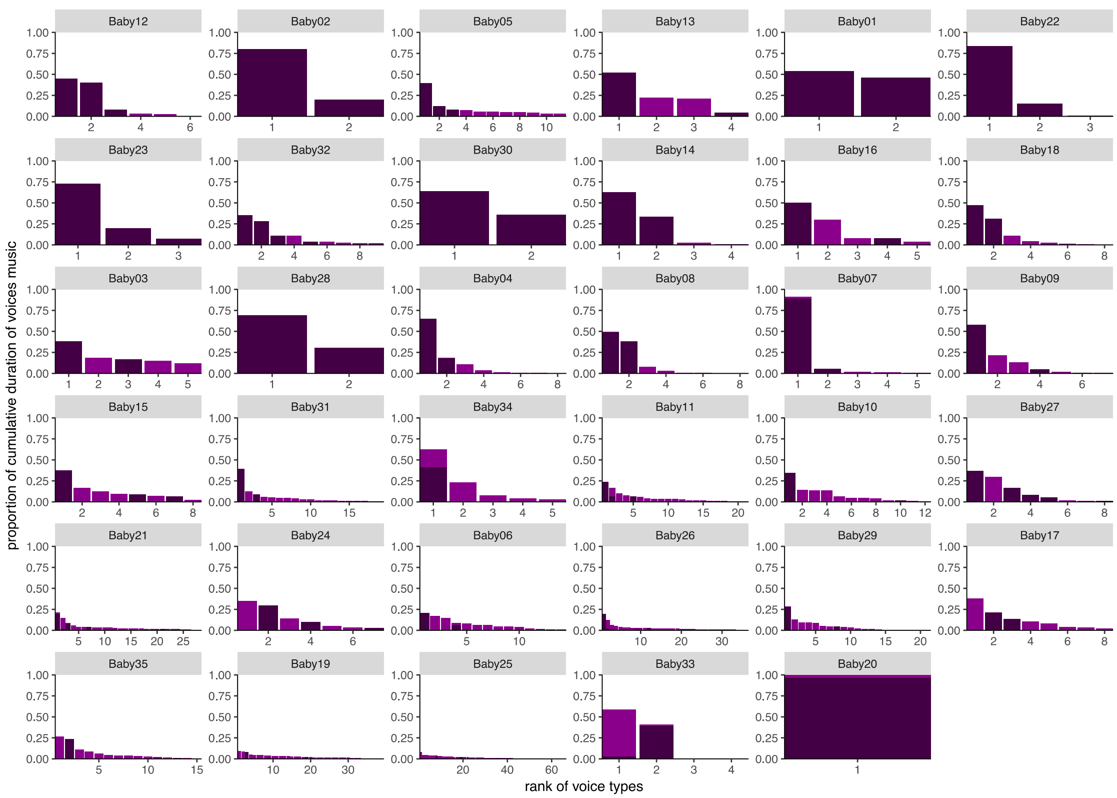

Figure 3.25. Live voices were commonly at the top of the frequency distributions of all voice types that occurred in the music of infants' everyday environments. Each bar represents one unique voice, and the darker bars depict the unique voices that were live. The height of the bars indicates the proportion of the cumulative duration of music that was accounted for by each voice type.

DISCUSSION

In this section of the dissertation, we answered the broad question: In what ways did infants encounter consistency, diversity, and social quality in the tunes and voices that occurred in their everyday musical input? We discovered that infants encountered diversity in the tunes and voices that occurred in the music of their everyday lives – infants encountered multiple unique tunes and multiple unique voices in the music that occurred in their everyday lives. We also found that infants encountered consistency in the tunes and voices that occurred in the music of their everyday lives, such that some specific tunes and some specific voices were more available than other tunes and other voices that occurred in infants' everyday music. In particular, we examined the single tune and single voice that occurred for the longest cumulative duration in the music of each infant's day. Infants encountered these "top tunes" and "top voices" for considerably longer than would be expected if each unique tune and each unique voice occurred for an equal duration. Taken together, we found evidence that infants encountered consistency bias amid diversity in the tunes and voices that occurred in the musical input available in their everyday environments.

We further examined infants' everyday musical input, assessing the social quality of the music that infants encountered. Across infants, the proportion of tunes music that was both live and vocal and the proportion of voices music that was live varied widely. Within the live-vocal music that occurred, most infants encountered consistency such that some specific live-vocal tunes and some specific live voices were more available than others in their everyday musical input. Many infants had one live voice, in particular, that occurred for a longer duration than would be expected if all the live voice types had occurred for an equal duration. Finally, for tunes, infants varied as to where live-vocal tunes occurred in the mix of all tunes. For almost half of the infants, their rank-one tune type occurred in a live-vocal rendition at least some of the time. In contrast, live voices were largely at the top of the mix of all voices. For most infants (.86), their rank-one voice type was a voice that occurred live at least some the time. In conclusion, we discovered converging evidence in favor of consistency of tunes and voices in infants' everyday live musical input available to infants.

What does it mean to encounter consistency biases in everyday musical input?

In this section of the thesis, we have reported that infants encountered consistency in the tunes, voices, live-vocal tunes, and live voices that occurred within the music available in their everyday environments – some specific tunes and some specific voices were more available (i.e., occurred for longer cumulative durations) than others. Because infants encountered biased frequency distributions of tunes and of voices, one specific tune and one specific voice were more available than others. Infants had the most opportunity to learn about their top tunes and top voices. It is possible that infants' top tunes and top voices could have a more powerful impact on infants' subsequent musical learning than other tunes and other voices that infants encountered in their input. Furthermore, the particular properties of infants' top tunes and top voices could determine the particular way in which infants' encountered musical input shapes their subsequent processing of musical stimuli. Such effects would be consistent with laboratory-based research showing that the nature of frequently-encountered item influenced how infants formed category boundaries (e.g., Oakes & Spalding, 1997).

The results presented in this section of the thesis lend plausibility to the mechanistic hypothesis that one highly frequent tune or one highly frequent voice might guide infants' processing of other tunes and voices that occur in the music they encounter (e.g., Bortfeld et al., 2005; Kurumada, Meylan, & Frank, 2013; Valian & Coulson, 1988). In order for this *perceptual anchor hypothesis* to potentially explain how infants build knowledge about tunes and voices that occur in their everyday musical input, infants would have to encounter a range of different tunes and different voices that occurred for differing cumulative durations. We discovered just such a distributional structure for both tunes and voices that occurred in music captured in day-long recordings of infants' everyday auditory environments. That infants' top voice was often a live voice may add power to its role as a perceptual anchor, by engaging and sustaining infants' attention to a greater degree than recorded voices (e.g., Belin, Zatorre, & Ahad, 2002; Levy, Granot, & Bentin, 2001; Nakata & Trehub, 2004).

Outstanding Questions and Future Directions.

The present findings raised several additional questions that have potential to be answered with further research. One logical question is: how were voices connected to tunes? In our research, by definition, any voice that produced music generated a tune. So, we could ask: how many different unique tunes did an infant's rank-one voice produce? For example, did mom sing *Itsy Bitsy Spider* each time she produced music? Or did she sing a mix of different tunes? Correspondingly, how many unique voices produced the same tune? Or, further still, how often did the same unique voice generate the same unique tune? In language learning, experiencing the same word spoken by multiple different voices tends to boost infants' learning of that word (e.g., Rost & McMurray, 2009, 2010; Twomey, Ranson, & Horst, 2013). Because repetition and variability both impact learning and memory in different ways (e.g., Horst, Parsons, & Bryan, 2011; Rost & McMurray, 2009), it would be of interest to determine the extent of consistency (or variability) of the tunes associated with the rank-one voice and vice versa (the voices associated with the rank-one tune).

Another question was: how many separate instances of music contained the top tune or the top voice? Because we defined the rank-one type as the type with the longest duration relative to other types, it was possible that infants encountered their rank-one voice (or tune) in one long, continuous music bout. It was also possible that their rank-one voice (or tune) occurred in multiple, separate music bouts throughout the infants' day. Each of these possibilities would potentially shape infants' learning differently. For example, encountering the rank-one voice in one long, continuous music bout might engage and sustain infants' attention (e.g., Belin, Zatorre, & Ahad, 2002; Levy, Granot, & Bentin, 2001; Nakata & Trehub, 2004), providing the opportunity for the infant to learn more about additional aspects of the musical content (e.g., pitches, rhythms, melodies, lyrics, etc.) of that music bout (e.g., Margulis, 2014). Encountering the rank-one voice in multiple, separate instances might result in other unique voices occurring in between repeated instances of the top voice, which would provide opportunities for the infant to compare and contrast their top voice with other voices (Carvalho & Goldstone, 2015). Determining the answer to this question is vital for further evaluating the plausibility of the perceptual anchor hypothesis. In Chapter 3, we explored these possibilities, examining how repeated instances of voices and tunes occurred *across time* in infants' recorded days.

A third question is: to what extent can music that was coded as both 'live' and 'vocal' be considered "high-quality" musical input? There are several dimensions that could make live-vocal music qualify as high-quality input. For one, it could be infant-directed, meaning it involves higher-pitched voices, slower tempos, and greater emotional engagement (Trehub et al., 1997; Longhi, 2009; Bergeson & Trehub, 1999). For another, it could be produced by a caregiver contingently responding to their infant's vocalizations (e.g., Goldstein, King, & West, 2003). It could also be a clear auditory signal (i.e., not cluttered) that is physically close to the infant. To the extent that live-vocal music reflects these properties of "high-quality" auditory input, then it might more powerfully capture infants' attention and subsequently boost their learning of musical content. Assessing the infant-directedness of infants' daily musical input would require additional coding. To our knowledge, there is no automatic way to assess the infant-directedness of the musical sounds recorded in everyday environments. Human listeners may be able to distinguish which musical sounds were infant-directed and which were not, as adult participants have successfully made this distinction for recordings of mothers singing in a laboratory setting either to their infants or without their infants (e.g., Trainor, 1996). The LENA automatic analyses provided estimates of the number of infant vocalizations, the distance from the recorder to the sound source, and the amount of noise, which would yield data about how physically close and cluttered the auditory signal was to the infant (properties relevant to determining the degree of quality). We intend to pursue this coding and use these automatic values in future research, as this question poses an exciting direction for future work.

Challenges and Limitations.

One limitation of the current study is that we did not include the music bouts that contained multiple tunes or multiple voices. It is unknown to what extent and in what way the tunes and voices that occurred during these music bouts would alter the frequency distributions of tune types and of voice types that we reported here. We had reason to predict that the overall pattern of results we reported is unlikely to change dramatically if we were to include the tunes and voices from the excluded music bouts. For one, the present results are based on the majority of the data. For another, we have preliminary insights from the coders about the various situations that yielded music bouts with multiple tunes or multiple voices. Multiple voices often occurred together in the same music bout because caregivers sang along either with each other or with a recorded music source, like the radio. Caregivers' voices were already the most frequently occurring voice for most infants, so this scenario would simply increase the cumulative duration of the most frequently occurring voice. The cumulative duration of additional voices would also increase, but it is likely that there would still be one voice that occurred for a much longer cumulative duration than the others voices. Multiple tunes often occurred together in the same music bout either because infants were playing with toys that produce music or because caregivers were flipping through radio stations. Many infant toys play the same small set of tunes repeatedly, which has potential to add further consistency to the mix of tunes. Caregivers flipping through radio stations results in tunes with very short durations. These tunes might add to diversity of tunes by increasing the range of unique tunes; but, they are unlikely to change the biased shape of the frequency distribution of tune types to the point that there is no longer one tune that accounts for a larger proportion of the distribution than other tunes. In on-going work, we are further coding the music bouts with multiple tunes and multiple voices, so we can ultimately incorporate the tunes and the voices that occurred in these bouts in our analyses.

One challenge of the current work was in defining what counted as a 'tune', and then what counted as a tune occurring again. In other words, when did a tune instance count as the same versus as different? In the current coding scheme, if the music contained any portion of a tune, it was given the same tune title. For example, if an infant's mom sings, "twinkle twinkle little star" in one music event, and then later sings, "up above the world so high", then both music events would be identified as *Twinkle Twinkle Little Star*. However, the particular pattern of pitches is actually different in each of those musical phrases. You can see in Figure 3.26 that the general pattern across the first 7 notes of *Twinkle Twinkle Little Star* (i.e., the first phrase) is that the notes go up; while the general pattern across the next 7 notes (i.e., the second phrase) is that they go down. As adults, we readily recognize the "twinkle twinkle little star" and "up above the world so high" are two phrases that are part of the same tune – we aggregate these phrases despite their differences in pitch patterns. This is why we opted for the current coding scheme. However, pitch patterns – or melodic contour – are particularly salient to infants (e.g., Chang & Trehub, 1977a). Thus, infants might not initially aggregate two musical instances that contained different pitch patterns, even if those instances are truly two sections of the same tune. It would be of interest to code pitch patterns in this corpus of music to determine how frequently this scenario arose in the everyday music available to infants. Then, we would have a more complete picture of the musical input that presumably drives infants' music learning.

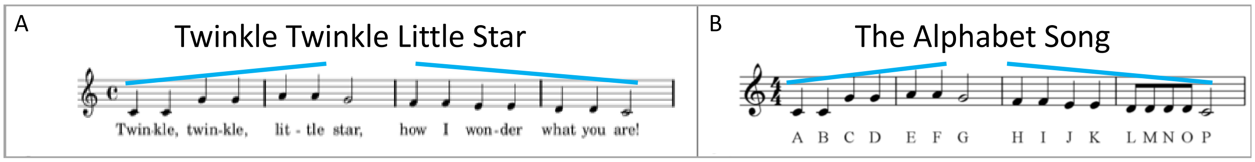

Figure 3.26. Musical notation for the first two phrases of (A) *Twinkle Twinkle Little Star* and (B) *The Alphabet Song.* The blue lines show the overall direction of the pitch frequencies.

It was also possible for two tunes to have been given different titles but to have actually contained the same set of pitches and rhythms. For example, *Twinkle Twinkle Little Star*, *The Alphabet Song*, and *Baa Baa Black Sheep* all share the same basic melody – the same pattern of pitches and rhythms; yet, each would be given its own unique title according to our current coding system. Again, you can see in Figure 3.26 that the first 7 notes of *Twinkle Twinkle Little Star* are the same as the first 7 notes of *The Alphabet Song.* Thus, it will ultimately be important to identify the specific patterns of pitches and rhythms of the tunes in this corpus of infant-available music. It is possible, especially as they are still acquiring language, that infants might aggregate two musical instances that contained the same pitch patterns even if the lyrics were different. Future research examining this level of music in the raw auditory data of the present corpus (e.g., pitch and rhythm patterns, tempos) would yield further insight into the extent to which infants' everyday musical input exhibits consistency.

In conclusion, we have provided evidence of consistency in the tunes and voices in infants' everyday music, such that some specific tunes and some specific voices occurred highly frequently among a mix of other tunes and voices. Live voices and live-vocal tunes displayed a comparable consistency bias, and live voices were commonly at the top of the frequency distributions of all voices. The mechanistic relevance of these findings for music learning depends on how the tunes and voices occurred in time, which we examined in the following section of this thesis.

CHAPTER IV

THE TEMPORAL DYNAMICS OF INFANTS' EVERYDAY MUSICAL INPUT ACROSS A DAY

INTRODUCTION

In the first two chapters of this thesis, we have reported key distributional properties about music and its contents – tunes and voices – that infants encountered in their everyday auditory environments. We discovered that infants encountered about one hour of music during their days (see Chapter 1). Infants' daily musical input contained multiple tunes and multiple voices. Further, most infants encountered one specific tune and one specific voice for considerably longer cumulative durations than the other tunes and other voices that occurred in music during their days (see Chapter 2). In other words, the "top" tune and the "top" voice were particularly available in infants' everyday musical input. How does encountering such consistency in the tunes and in the voices of their everyday music shape infants' learning in the domain of music?

The crux of how infants' everyday musical input may ultimately shape their music learning lies within in the temporal dynamics of infants' everyday musical input. In this chapter, we addressed the broad question: How were individual instances of music and its contents distributed *in time* across infants' days? Does music occur at regular or random intervals over the course of infants' days? For example, if an infant first encounters music at 7:00am, then does music next occur at 8:00am, and then at 9:00am, and then again at 10:00am, and so on once every hour? How much time passes from one music bout to the next? If the infant encounters music at 7:00am, then do they next encounter music one second later at 7:00:01am, one minute later at 7:01am, or one hour later at 8:00am? Furthermore, in what order do the contents of music (i.e., tunes and voices) occur across the day? If at 7:00am the infant encounters *Itsy Bitsy Spider*, then the next time music occurs, is it another instance of *Itsy Bitsy Spider* or is it an instance of *The Wheels on the Bus*? If *Itsy Bitsy Spider* is the infant's top tune, then how often does it occur adjacent to another less frequent tune?

These questions highlight two key temporal properties that have been shown in laboratory-based research to impact when and how people integrate information to build knowledge across multiple domains. These two properties are **temporal interval** (e.g., Childers & Tomasello, 2002; Hintzman, 1969; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Rubin-Rabson, 1940; Simmons, 2012; Schwab & Lew-Williams, 2016; Vlach & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008) and **sequential order** (e.g., Carvalho & Goldstone, 2013, 2017; Elio & Anderson, 1984; Mather & Plunkett, 2011; Zeithamova & Maddox, 2009; Mathy & Feldman, 2009; for recent review of related work, see Carvalho & Goldstone, 2015). Here, we first review research that has examined temporal properties of to-be-learned input in other domains, to inform our predictions about the nature of the temporal dynamics of infants' everyday musical input and about how different temporal properties of infants' everyday musical input may shape infants' subsequent learning in the domain of music. Then, we examine the temporal intervals at which music occurred and the sequential order of the tunes and the voices that infants encountered in their everyday musical input. Describing how music (as well as tunes and voices) occurred *in time* over the course of infants' days is an important step in understanding how infants integrate information from their encountered everyday input and ultimately build knowledge in the domain of music.

Bursty temporal dynamics of natural human behaviors.

Human behaviors and communication are streams of rich, complex events that unfold across time. Recent research has discovered that many human behaviors, such as sending emails (e.g., Karsai, Kaski, Barabási, & Kertész, 2012) and checking out library books (e.g., Vázquez, Oliveira, Dezsö, Goh, Kondor, & Barabási, 2006), occur in bursty temporal patterns (for recent review, see Karsai, Jo, & Kaski, 2018), as opposed to occurring regularly or randomly in time. Bursty temporal patterns are characterized by having some events that occur clustered together in time as well as long lulls during which no events occur. In other words, the durations of the intervals between events are heterogeneous. Bursty temporal patterns have also been discovered in both everyday and naturalistic human social interactions – words (Altmann, Pierrehumbert, & Motter, 2009), utterances (Slone, Abney, Smith, & Yu, 2017), and verbal dialog (Abney, Dale, Louwerse, & Kello, 2018) occur in bursty temporal patterns.

Music, like language, is a complex, human-generated phenomenon that unfolds across time. Like language, musical sounds have also been shown to exhibit temporal clustering (e.g., Falk & Kello, 2017; Kello, Dalla Bella, Médé, & Balasubramaniam, 2017). For example, Falk and Kello (2009) recorded mothers reading a children's story aloud and singing a popular playsong either to their infant or to an experimenter. They then analyzed the temporal clustering of acoustic energy in the audio recordings, and they discovered that infant-directed speech and singing exhibited greater temporal clustering than did adult-directed speech and singing. While this study provided initial evidence of temporal clustering in musical input to infants, it was limited to one particular tune that was selected by the researchers. Kello and colleagues (2017) analyzed temporal clustering in music across multiple musical genres; but, this study analyzed publicly available audio recordings, rather than everyday musical sounds available in natural auditory environments. It remains unknown to what extent musical input in infants' everyday auditory environments also occurs in a temporally clustered manner. Therefore, our first research question was: did music in infants' everyday auditory environments occur in regular, random, or bursty temporal patterns across infants' recorded days?

Attention and memory matter for learning.

At the core of learning across all domains are attention and memory. To build knowledge in any domain a learner must encode and recall encountered items. Then, across separate encounters, learns must integrate repetitions about the same item and separate those from repeated instances of different items. Comparing and contrasting items from different categories promotes learning about those categories (e.g., Carvalho & Goldstone, 2013, 2017). How items occur in time and in sequential order has attentional consequences which in turn impact learning and memory (e.g., Hintzman, 1969). The particular impact of temporal structure on attention, learning, and memory depends on many factors including (but not limited to) the particular task demands (e.g., Carvalho & Goldstone, 2017), the learner's level of expertise (e.g., Braithwaite & Goldstone, 2015), and the nature of the encountered input (e.g., Oakes & Spalding, 1997). Memory integration in young infants may pose a particular challenge as the neural structures that support memory are still developing (Gómez & Edgin, 2016). It is possible that infants' environments may reduce integration demands by presenting infants with simple, repeated input. This would be comparable to the simpler, more redundant speech input available to infants and young children compared to that available to older children and adults (e.g., Snow, 1972). In the current work, we examined the temporal dynamics of musical input available to infants, to determine whether such properties might plausibly reduce the integration demands and support infants' successful learning in the domain of music.

Temporal spacing of to-be-learned content shapes learning.

One common approach in cognitive and developmental psychology has been to manipulate the temporal interval at which learners encounter repeated instances of the same content (e.g., Childers & Tomasello, 2002; Hintzman, 1969; Vlach, Sandhofer, & Kornell, 2008). A temporal interval, or "gap", is the duration of time from the end of one event to the beginning of the next event. "Spaced" schedules typically involve repeated instances of the same content that occur with at least a one-second delay between each presentation. "Massed" schedules typically involve one continuous presentation of the content or repeated instances of the same content separated by less than 1 second. The best available evidence from infants, children, and adults has shown that spacing information out in time typically benefits learning and memory across a wide variety of domains (e.g., Childers & Tomasello, 2002; Ebbinghaus, 1885/1964; Hintzman, 1969; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Roy, Frank, DeCamp, Miller, & Roy, 2015; Schwab & Lew-Williams, 2016; Vlach & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008) including music (Rubin-Rabson, 1940; Simmons, 2012; for reviews/meta-analyses, see Cepeda, Pahler, Vul, Wixted, & Rohrer, 2006; Donovan & Radosevich, 1999; Janiszewski, 2003).

In one clear laboratory-based example of this effect (Vlach, Sandhofer, & Kornell, 2008), young children encountered novel objects with a novel nonsense label (e.g., “wug”). The researchers manipulated the temporal schedule of these presentations: children encountered these objects either in immediate succession (“massed”) or with 30 seconds between each presentation (“spaced”). The researchers also manipulated the type of learning tested: some children saw the same novel object paired with the same label three times. These children were then given four objects – one identical to the originally shown and three novel ones – and they were asked to choose the ‘wug’ among these objects (a test of their recognition memory). Other children saw three similar but different objects (category exemplars) each paired with the same one word. These children were then given four objects – a novel within-category exemplar and three novel objects – and they were asked to choose the ‘wug’ among these objects (a test of their category generalization). All children performed better on both test types after they had received the spaced presentation schedule, meaning that children’s recognition and generalization benefitted from having encountered information spread out in time. One theory about why spacing boosts learning is that encountering a stimulus at spaced temporal intervals provides opportunities for learners to forget and then retrieve information about that stimulus over time, which then improves their memory and promotes their learning of that information (e.g., Bjork, 1988; Cuddy & Jacoby, 1982; Hintzman, 1984; Hintzman & Ludlam, 1980).

Across this body of research, spaced schedules typically are designed with regular (i.e., homogenous) temporal intervals, as in the example above (Vlach, Sandhofer, & Kornell, 2008) where children encountered one item at regularly spaced 30-second intervals. The fact that many natural phenomenon and human behaviors occur with bursty temporal dynamics (i.e., heterogeneous temporal intervals; e.g. Abney, Dale, Louwerse, & Kello, 2018; Altmann, Pierrehumbert, & Motter, 2009; Falk & Kello, 2009; Karsai, Kaski, Barabási, & Kertész, 2012; Kello, Dalla Bella, Médé, & Balasubramaniam, 2017; Slone, Abney, Smith, & Yu, 2017; Vázquez, Oliveira, Dezsö, Goh, Kondor, & Barabási, 2006; for review see Karsai, Jo, & Kaski, 2018) raises questions about the extent to which human learners encounter to-be-learned items at regular temporal intervals in their real-world environments. In one extensive case study, researchers recorded the linguistic input available to one child during the first 3 years of their life (Roy, Frank, DeCamp, Miller, & Roy, 2015). The researchers measured the frequency of specific words as they occurred in one-hour time intervals throughout the day. They created an index of temporal distinctiveness – the extent to which a word occurred during a specific part of the day (as opposed to occurrences being regularly distributed across all hours of the day). Words that occurred in a more temporally distinctive pattern over the course of a day were produced at an earlier age by the child. The researchers observed that these temporally distinct words occurred at the same time across different days. These results provided evidence that encountering to-be-learned information – in this case, words –in a temporally distinct pattern within a day as well as at a regular temporal interval across days was positively associated with subsequent learning.

In the domain of music, prior research has demonstrated that adults who practiced their instrument according to a spaced schedule later performed more proficiently on subsequent test of musical production compared to adults who followed a massed practice schedule (e.g., Rubin-Rabson, 1940; Simmons, 2012). Since young infants are primarily honing their musical perception skills, rather than mastering musical production (for reviews, see Trehub, 2006; Trehub & Degé, 2015), it is unclear to what extent the findings about spaced instrumental practiced schedules are applicable. For most infants, the question is how the temporal dynamics of their everyday musical input shape their music perception skills. To the best of our knowledge, only one prior study has provided any data on how musical input occurs in time during everyday life – in an experience-sampling study, college students reported their everyday emotional experiences that either were or were not connected to music that had occurred (Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008). The researchers analyzed the mean number of 'musical emotion episodes' per participant that occurred in the morning, in the afternoon, and in the evening, discovering that a significantly higher number of musical emotion episodes occurred in the evening than in the morning. This index – the average number of musical emotion episodes per three time windows – provided only a limited picture of the temporal dynamics of adults' musical input. Although several additional studies have used the experience-sampling method to capture and characterize music in everyday environments (Greasley & Lamont, 2011; North, Hargreaves, & Hargreaves, 2004; Sloboda, O'Neill, & Ivaldi, 2001), including one study in which parents responded about the music available in the everyday environments of their toddlers (Lamont, 2008), none has reported any data on the temporal patterns at which music occurred over the course of the sampled days. These studies have revealed the proportion of sampled episodes in which participants reported hearing music – based on these proportions, it is possible to estimate the rate at which music occurred. For example, in research by Sloboda and colleagues (2001), participants received a prompt via pager at a random time during every two-hour window of the day from 8:00am to 10:00pm for 7 days. Participants reported experiencing music in 155 out of the 356 episodes. One could estimate, then, that music occurred roughly once every 4-5 hours; however, this assumes a constant rate of experiencing music. Given the temporal clustering patterns of many other human behaviors and communication, including music (e.g., Abney, Dale, Louwerse, & Kello, 2018; Altmann, Pierrehumbert, & Motter, 2009; Falk & Kello, 2009; Karsai, Kaski, Barabási, & Kertész, 2012; Kello, Dalla Bella, Médé, & Balasubramaniam, 2017; Slone, Abney, Smith, & Yu, 2017; Vázquez, Oliveira, Dezsö, Goh, Kondor, & Barabási, 2006; for review see Karsai, Jo, & Kaski, 2018), it seems unwise to assume music will occur at regular temporal intervals.

In sum, existing research has offered limited insight into the temporal dynamics of music that occurs in everyday environments for adults, and no data at all on the temporal structure of infant-available music in natural contexts. In other words, we do not yet know at what temporal intervals music occurs in infants' everyday environments. Does music occur on the order of seconds or minutes or hours? We must answer this question before we can start to understand how the temporal intervals at which music occurs might impact infants' integration of encountered musical content. In the current research, we addressed this question by asking: how long were the gaps between music bouts in infants' everyday auditory environments?

Sequential order of to-be-learned content shapes learning.

Considerable research has demonstrated that presentation order also shapes how infants, children, and adults encode and recall information in many domains (e.g., Carvalho & Goldstone, 2013, 2017; Elio & Anderson, 1984; Mather & Plunkett, 2011; Zeithamova & Maddox, 2009; Mathy & Feldman, 2009; for recent review of related work, see Carvalho & Goldstone, 2015). Laboratory-based research has largely focused on two types of sequential orders: blocked, in which learners encounter repeated instances of the same item (or multiple exemplars from the same category), and interleaved, in which learners encounter two items (or exemplars from two categories) that alternate. Specifically, when learners encountered a blocked stimulus presentation, then they attend to the features that were the same across instances, more strongly encoding that shared features of category exemplars (e.g., Carvalho & Goldstone, 2017). Blocked stimulus presentation might also promote infants' discovery of additional details about that category – in music, repeated exposure to the same musical piece leads listeners to discover new information (e.g., rhythm patterns, lyrics, tempo changes, etc.) that they may not have previously noticed (e.g., Margulis, 2014). On the other hand, when learners encounter an interleaved order, then they attended to differences between encountered categories (e.g., Carvalho & Goldstone, 2017). Comparing and contrasting the stimuli strengthened their representations of those items (categories), particularly the features that differentiate the items (categories) (e.g., Carvalho & Goldstone, 2017).

As one concrete example of these effects, Carvalho and Goldstone (2017) presented adult participants with categories of cartoon aliens. Each alien had 5 features that could vary (i.e., arms, legs, eyes, mouths, and antennae). Participants studied two categories of aliens in a blocked design, and they studied two different categories of aliens in an interleaved design. Some features systematically differentiated the categories of aliens. One feature was 'characteristic' in that it was common to the two presented categories of aliens. Participants were then tested on their ability to categorize previously studied aliens, novel aliens with the characteristic feature preserved, and novel aliens with the characteristic feature changed. When participants had studied blocked categories, they more accurately categorized the novel alien exemplars that preserved the characteristic feature compared to the novel alien exemplars that did not – presumably, the blocked presentation led participants to attend more to features that were similar within and across categories, which would make the characteristic feature particularly salient. When that common feature was preserved, then these participants could more effectively identify the discriminating features; but when that common feature was changed, then these participants were distracted by the missing feature and their categorization performance dropped. When participants had studied interleaved categories, they were equally accurate in categorizing both types of novel aliens (characteristic-feature preserved and characteristic-feature changed), presumably because the interleaved presentation led them to attend more to the discriminating features (so a change in the characteristic feature had less of an impact in this case than after blocked study). In other words, learners attended to and integrated different aspects of their encountered input based on the particular sequential order of that input – blocked or interleaved – that they experienced.

In natural environments, sequential order may be more complex, as learners may encounter more than two items (or two categories of items). For example, based our findings reported in Chapter 2, individual infants encountered from 14 to 213 unique tunes (Median = 51, SD = 41.74) and from 1 to 66 unique voices (Median = 8, SD = 13.16) in the music that occurred during their recorded days. Thus, a small number of unique items (i.e., tunes or voices) occurred highly frequently, while most unique items occurred relatively infrequently – in other words, infants encountered "biased" frequency distributions (i.e., non-uniform) of tunes and voices (for details, see Chapter 2). The perceptual anchor hypothesis proposes that encountering high-frequency 'anchor' items might bootstrap subsequent learning of other low-frequency items (i.e., diversity; e.g., Kurumada, Meylan, & Frank, 2013; Oaks & Spalding, 1997; Valian & Coulson, 1988; for relevant review on infants' natural visual input, see Smith, Jayaraman, Clerkin, & Yu, 2018). The proposed mechanism of this perceptual anchor hypothesis is that encountering such a biased mix of high- and low-frequency items would present learners with opportunities to compare and contrast many different low-frequency items with a small set of highly frequent (and thus more familiar) anchor items. One possible structure of input that would provide such opportunities would be for learners to encounter a high-frequency 'anchor' item paired *in sequential order* with a low-frequency item. Therefore, we addressed the question: How often did the rank-one tune (voice) occur paired in sequential order with other unique tunes (voices)? In addition, as a more global measure of change, we examined the rate at which infants encountered new unique tune types and new unique voice types from one instance to the next, in the same sequential order they occurred over the course of the day. Our final question was: How often did unique tunes (voices) switch over the course of infants' days?

Current Research

The field currently lacks information about the temporal dynamics of infants' everyday musical input. In surveys and interviews, caregivers have reported using music on a daily basis with their infants (e.g., Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003). Because 'daily' is the finest grained temporal interval that has been queried in these studies, little more is known about the detailed temporal patterns at which music and its contents (i.e., tunes and voices) occur across infants' days. A small number of prior studies have recorded raw musical input in infants' at-home environments (Addessi, 2009; Eckerdal & Merker, 2009; Koops, 2014). The recordings in these studies only captured about two hours in a day and families were generally aware that the researchers were studying music, so these studies offer limited insight into how infants' everyday musical input occurs in time over the course of a day. Infants must recall and integrate information after encountering some amount of musical input at some time intervals and in some orders, but there is no guiding data available about the particular details. Therefore, analyzing the temporal dynamics of music, tunes, and voices in the present corpus of infants' everyday musical input has potential to reveal altogether new information about how music occurs in time over the course of a day in infants' everyday lives. In the current section of this thesis, we described the temporal dynamics of infants' everyday musical input. We first characterized the temporal structure of music bouts, separately for music bouts in the tunes music corpus and for music bouts in the voices music corpus. Then, we separately determined the temporal structure of instances of infants' top tunes and top voices. Finally, we examined the sequential order at which top tunes and other specific tunes (and separately, top voices and other specific voices) occurred across infants' recorded days.

Temporal structure of music bouts. We determined the temporal pattern at which music bouts occurred over the course of infants' days, separately for the music bouts in the tunes music corpus and for the music bouts in the voices music corpus. We examined three possible temporal patterns: regular, random, or bursty (i.e., clustered).

Figure 4.1 depicts a schematic representation of these three possible temporal patterns of music bouts in the voices music corpus. The magenta bars each represent an individual music bout as it occurred in time from 7:00am to 5:00pm. To evaluate which of the three possible temporal patterns characterized the music bouts in each infants' day, we examined the inter-onset intervals (IOIs) – the number of seconds between *the start* of one music bout and *the start* of the next music bout (shown by the blue brackets on the left side of Figure 4.1.). If music bouts occurred in the regular temporal pattern (e.g., once every hour; Row A), then the IOIs would all be equal in duration (e.g., 3,600 seconds). If music bouts occurred in a random temporal pattern (Row B), then the IOI durations would be distributed exponentially. If music bouts occurred in a bursty (i.e., clustered) temporal pattern (Row C; e.g., Goh & Barabási, 2008; for recent review, see Karsai, Jo, & Kaski, 2018), then the IOI durations would occur in a heavy-tailed distribution.

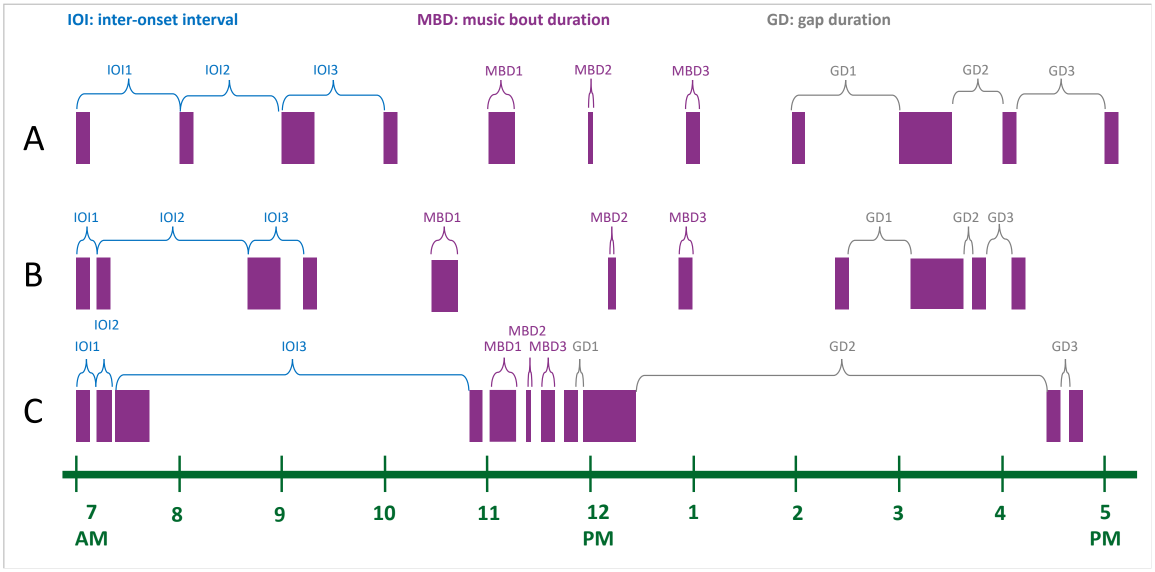

Figure 4.1. A schematic depicting 3 possible temporal structures of how music bouts (magenta bars) may have occurred in time over the course of infants' days: (A) regular (i.e., one music bout every hour), (B) random (i.e., the durations of inter-onset intervals (IOIs) are exponentially distributed), or (C) bursty (i.e., the distribution of IOIs is heavy-tailed).

Burstiness is one of many indices of temporal clustering (e.g., Abney, Kello, & Balasubramaniam, 2017; Warren Liao, 2005; for relevant discussion of measures, see Karsai, Jo, & Kaski, 2018). We opted to calculate the burstiness of music and its contents for three reasons. First, we aimed to characterize the temporal structure of music and its contents over day-long (i.e., 10+ hours) timescales. The burstiness parameter has successfully captured the complex temporal structure of heterogeneous temporal processes that unfold across days, months, and years (e.g., Karsai, Kaski, Barabási, & Kertész, 2012; Vázquez, Oliveira, Dezsö, Goh, Kondor, & Barabási, 2006). Second, it was advantageous that the burstiness parameter classifies individual event series as having either regular, random, or bursty temporal patterns (e.g., Goh & Barabási, 2008; Kim & Jo, 2016), given the paucity of data to inform a prediction about which particular temporal pattern might best characterize infant-available music in everyday auditory environments. Third, it is increasingly becoming a priority in developmental science to characterize properties of infants' natural experiences (Lee, Cole, Golenia, & Adolph, 2017; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017; for relevant review on infants' natural visual input, see Smith, Jayaraman, Clerkin, & Yu, 2018). Burstiness has been one measure recently used to analyze the temporal patterns of infants' natural input in visual and speech input (Abney, Jayaraman, Fausey, Slone, & Smith, 2017; Slone, Abney, Smith, & Yu, 2017). Thus, burstiness is well-suited to the current state of developmental science and will advance understanding beyond what is currently known.

In addition to burstiness and IOIs, we also examined the gap durations (GDs) between music bouts (i.e., the number of seconds from *the end* of one music bout until *the start* of the next music bout), because gap durations have been more commonly examined in prior research (e.g., Childers & Tomasello, 2002; Ebbinghaus, 1885/1964; Hintzman, 1969; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980; Roy, Frank, DeCamp, Miller, & Roy, 2015; Schwab & Lew-Williams, 2016; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017; Vlach & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008). We were particularly interested in determining the temporal pattern of these gaps, because recent research demonstrated that the linguistic input available to infants during a 45-minute free-play session in their natural environment contained a considerably larger amount of silence compared to the linguistic input available to infants in a 5-minute structured laboratory setting (Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Figure 4.1. shows GDs in gray brackets on the right side of the figure. If we ignore the duration of individual music bouts for the moment, then as with IOIs, if music bouts occurred regularly in time (e.g., once every hour; Row A), then each gap duration would each have an equal duration. If music bouts occurred randomly in time (Row B), then the distribution of gap durations would be exponential. If music bouts occurred in a bursty temporal pattern (Row C), then the distribution of gap durations would be heavy-tailed.

Finally, rather than ignoring individual music-bout durations (MBDs; i.e., the number of seconds from the start to the end of one music bout), we directly measured them. Figure 4.1. shows MBDs in magenta brackets in the center of the figure. Music-bout durations and gap durations are the two components of inter-onset intervals (i.e., one inter-onset interval is equal to the music-bout duration plus the following gap duration). Neither music-bout durations nor gap durations are taken directly into account in the burstiness calculation (rather, their combination, the IOIs, is used). We were interested in the duration of individual music bouts, because this duration constrained the amount of musical information available to infants. It is possible that music-bout durations were mainly short (i.e., on the order of seconds). This would mean infants encountered mainly brief tunes and/or partial phrases of longer tunes. For example, a typical rendition of *Happy Birthday* can be sung in about 14 seconds. It is also possible that music-bout durations were mainly long (i.e., on the order of minutes). For example, Beyoncé's hit song *Run the World (Girls)* has a duration of 235 seconds (~ 4 minutes). Longer pieces of music certainly exist – as one example, Tchaikovsky's *Symphony No. 5 in E minor, Op. 64* lasts for about 2,706 seconds (~ 45 minutes).

Very little prior research exists to guide our predictions about how long the durations of music bouts and the durations of the gaps between them might be in infants' everyday musical input. To the best of our knowledge, two recent case studies (Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Costa-Giomi & Sun, 2016) have provided the only available evidence about the durations of music bouts and the gaps between music bouts as they occur in infants' everyday environments. These studies identified music available in day-long auditory recordings of one 15-month-old infant (Costa-Giomi & Benetti, 2017) and of two 11-month-old twin infants (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016) while they were at home in their natural environments. Costa-Giomi and colleagues described the music in both of these families as having occurred in brief bouts (i.e., mostly on the order of seconds) with a mix of both short and extended pauses in between them. Short music bouts seemed particularly characteristic of music produced by live voices. Only one family had recorded music in their day (Costa-Giomi, 2016; Costa-Giomi & Sun, 2016) – Costa-Giomi and colleagues described the TV being on in the background for the majority of the recording period, with many musical sounds occurring for a cumulative duration of over 8 hours. It is unclear how long any one musical instance was from the sources of recorded music (i.e., TV, radio, toys, game consoles) in this family's day. It is also unknown the extent to which the findings from these two case studies generalize to the broader population. Our research extends this work, providing more detailed information about the duration of music bouts and of the gaps between them that occurred across day-long recordings of the natural auditory environments of multiple infants.

Temporal structure of top tunes and top voices. Next, we zoomed in from considering all music bouts in the tunes music corpus and all music bouts in the voices music corpus to focusing on specific tunes and specific voices within each separate corpus. Tunes and voices both present rich content that infants could learn. Tunes consist of melodies, rhythms, phrase structures, lyrics (in some cases). Voices vary in their timbre, pitch range, and tempos. How infants encountered instances of a specific tune and instances of a specific voice in time across their days could shape what they learned about those tunes and voices.

Specifically, we first focused on infants' top tunes and instances of infants' top voices. As in Chapter 2 of this thesis, we defined the top tune and the top voice as the specific tune and the specific voice that each occurred for the longest cumulative duration relative to other unique tunes and other unique voices in each infant's day. We focused on infants' top tunes and top voices, defined in this way, because infants had the greatest cumulative opportunity to learn about these specific tunes and these specific voices compared to any other single unique tune or single unique voice that occurred in the music of their days. It was possible that the top tune (voice) could have occurred in a single, long, continuous instance. In this case, there would be no temporal dynamics to analyze. Encountering a single instance of their rank-one tune (voice) would present the infant with very local integration demands – tracking the rank-one tune (voice) for the duration of its single instance. It could be that, in these cases, the top tune (voice) occurred again at timescales longer than one day – a possibility we could evaluate in future research using the Trio Corpus (Fausey & Mendoza, in prep) in which each family contributed 3 recorded days within one week. If the infants' top tune (voice) occurred in two or more separate instances, then we could examine the temporal dynamics of how these instances occurred over the course of infants' days. Even though the top-tune instances were a subset of the music bouts in the tunes music corpus, and the top-voice instances were a subset of the music bouts in the voices music corpus, the temporal dynamics of the top tunes and top voices did not necessarily have to reflect the temporal dynamics of the superset of music bouts. Top-tune (top-voice) instances could have occurred in regular, random, or bursty temporal structures, independently of whichever temporal pattern characterized their superset of music bouts. Therefore, we examined the same set of measures of temporal structure separately for infants' top tunes and top voices as we did for music bouts: burstiness (and inter-onset intervals), gap durations, and the durations of individual top-tune instances and of individual top-voice instances.

Top tune and top voice instance pairs. Because top tunes and top voices occurred for the longest cumulative durations in infants' days, they had the most potential to be well-learned and thus the most potential to serve as a perceptual anchor (e.g., Kurumada, Meylan, & Frank, 2013; Oaks & Spalding, 1997; Valian & Coulson, 1988; for relevant review on infants' natural visual input, see Smith, Jayaraman, Clerkin, & Yu, 2018) that could guide infants' processing of other tunes and other voices that they encountered in their everyday musical input. One way top tunes (voices) could plausibly serve in the role of perceptual anchor would be for instances of infants' top tunes (voices) to occur adjacent to instances of other unique tunes (voices). To evaluate this, we examined tune (voice) instance pairs, which included one tune (voice) instance and the immediately subsequent tune (voice) instance, in the sequential order they occurred in during infants' days. We determined how often infants encountered an instance of their top tune (voice) followed immediately by an instance of a different unique tune.

Sequential order of specific tunes and specific voices. Finally, we zoomed out to examine the sequential order of all unique tunes and of all unique voices. In laboratory-based research, sequential order is typically defined as either blocked (i.e., repetition of the same item or of exemplars from the same category) or interleaved (i.e., alternating two items or alternating exemplars from two categories). In the current study, infants encountered a mix of multiple (i.e., more than two) unique tunes and multiple unique voices (for detail, see Chapter 2 of this thesis). These data could not be cleanly classified into the extremes of "blocked" or "interleaved" orders. To evaluate the sequential order of unique tunes and unique voices that occurred in infants' everyday musical input, we calculated how often infants encountered a new unique tune (voice) from one tune (voice) instance to the next, in the same sequential order they occurred over the course of the day. This tunes (voices) "switch rate" revealed how often infants encountered a change in tune (voice) type across the tune (voice) instance pairs of their day.

To summarize, in the current section of the thesis, we addressed the following broad question: How were individual instances of music and its contents distributed *in time* across infants' days? We first evaluated the extent to which music, tunes, and voices occurred spaced out across time during infants' days. Separately for music bouts in the tunes music corpus and for music bouts in the voices music corpus, we answered the following questions about the temporal structure: (1) Did music bouts occur in regular, random, or bursty temporal patterns over infants' days?, (2) How long were the gaps between music bouts?, and (3) How long were individual music bouts?. Then, we focused on the top tunes and the top voices infants encountered, because infants had the greatest opportunity to learn about these specific tunes and specific voices. We addressed the following questions about infants' top tunes ("rank-one tunes") and top voices ("rank-one voices"): Did infants' rank-one tunes (voices) occur in regular, random, or bursty temporal patterns?, (5) How long were the gaps between instances of infants' top tune (voice)?, (6) How long were individual instances of infants' top tune (voice)?. Finally, we examined the extent to which specific tunes and specific voices changed (as opposed to repeated) in order across infants' days. We answered the following questions about the sequential order of tunes and voices: (7) How often did the top tune (voice) occur paired in sequential order with other tunes (voices)? and (8) How often did tune (voice) types switch over the course of infants' days? Determining the temporal dynamics of music, tunes, and voices that occurred in infants' everyday musical input yielded data on the nature of the integration challenges that infants faced in their everyday music learning.

METHOD

Details about the participants and the procedures for collecting, pre-processing, and coding music in the data are described in Chapter 1. Details about identifying voices and tunes in the music data are described in Chapter 2.

Data Analysis

Music bouts in two corpora. As in Chapter 2 of this thesis, we analyzed the set of music bouts that contained exactly one unique tune ("tunes music corpus") and the set of music bouts that contained exactly one unique voice ("voices music corpus"), excluding the music bouts that had multiple tunes and/or multiple voices (see Chapter 1 for additional details). We examined each dependent measure of temporal structure separately for these two corpora.

Burstiness and inter-onset intervals. We first examined the temporal clustering of music bouts as they occurred across infants' days using the burstiness parameter (e.g., Goh & Barabási, 2008; Kim & Jo, 2016). Bursty temporal patterns are characterized by many events that occur within a short period of time (i.e., bursts or clusters), as well as long lulls during which no events occur. Burstiness is especially well suited to characterize temporal dynamics in sequences of events that do not occur regularly in time (e.g., Karsai, Jo, & Kaski, 2018). The burstiness parameter (*B*) is typically computed as follows (e.g., Goh & Barabási, 2008; Kim & Jo, 2016),

$$B = \frac{\sigma - \mu}{\sigma + \mu} \qquad \text{(eq. 1)}$$

where σ is the standard deviation of the inter-onset intervals (IOIs) and μ is the mean of the IOIs. Burstiness (*B*) ranges from -1 to 1. A value of 0 indicates that events occur randomly across time, and values closer to -1 indicate that events occur periodically. Values closer to 1 indicate a bursty (or clustered) temporal structure of events.

Because this equation (eq. 1) yields a less accurate estimate of burstiness for small set sizes of events (i.e., less than ~100 events; Kim & Jo, 2016), we used a modified equation that corrects for having small event set sizes (Kim & Jo, 2016):

$$A_n(r) = \frac{\sqrt{n+1}r - \sqrt{n-1}}{(\sqrt{n+1} - 2)r + \sqrt{n-1}} \qquad \text{(eq. 2)}$$

where r is equal to the standard deviation of IOIs divided by the mean of IOIs (i.e., σ / μ) and n is equal to the total number of events. Like the original burstiness equation (Goh & Barabási, 2008), $A_n(r)$ ranges from periodic (-1) to random (0) to bursty (1). This measure requires at least 3 IOIs; therefore, we set a criterion that infants must have encountered at least 4 separate events (e.g., music bouts) to be included in our burstiness analyses.

If the burstiness value of music bouts is high, then this would indicate that some music bouts occurred close together in time while other music bouts occurred with longer delays between them. If music bouts occurred in a bursty temporal pattern, then the distribution of IOIs would be heavy-tailed. If the burstiness value of music bouts is low, then this would indicate that music bouts occurred in a temporally regular (i.e., periodic) pattern over the course of infants' days. If music bouts occurred in a regular temporal pattern, then every IOI in that infant's day would have an equal duration. If the burstiness value of music bouts is near zero, then this would mean music bouts occurred randomly in time across infants' days (see the following section for how we defined the bounds the 'random' category). If music bouts occurred randomly, then the distribution of IOIs would be exponential. These three possibilities each portray different temporal structures that would result in very different integration demands for infants as they build musical knowledge across a day. We determined the proportion of infants that encountered music bouts in each of these three temporal structures, which will provide data about the nature of the integration challenge infants encounter in their real-world musical input. We also evaluated the burstiness and IOIs of instances of infants' top tunes and separately of instances of infants' top voices. We predicted that most infants would encounter music bouts, top tunes, and top voices in a bursty temporal pattern.

Simulated bounds of the "random" category for burstiness. We determined the lower and upper bounds of the random category for the burstiness analyses, modifying the simulation procedure outlined in recent research on the temporal dynamics of human communication (Abney, Dale, Louwerse, & Kello, 2018). We simulated one data stream per infant from an exponential distribution, using sample sizes matched to the number of events in each infant's real data. These values were then used as IOI durations in the burstiness calculation. In an exponential distribution, the times are assumed to be independent, so the resulting burstiness value should be $A_n(r) = 0$. We calculated the burstiness parameter for each simulated data stream. We then repeated this process 100 times. Based on the resulting burstiness estimates, we determined the mean burstiness value ($A_n(r)$) and the lower and upper 95% confidence intervals (CIs). The CIs defined the lower and upper bounds of the random category in our burstiness analyses. We conducted this entire procedure separately for each of our main event types of interest (i.e., music bouts, vocal music bouts, top tunes, and top voices). For each event type, the number of simulated data streams matched the number of infants whose real data contained at least 4 events for that event type.

Gap durations. To further address questions about spacing, we also calculated "gap" durations – the number of seconds from the end of one music bout to the start of the subsequent music bout. For any one infant, if all of the gap durations between music bouts were short, then the music bouts would have occurred close together in time. Conversely, if all of the gap durations were long, then the music bouts would have occurred spread out across time. Each of these possibilities poses different integration demands. As we did for burstiness, we separately examined gap durations between music bouts in the tunes music corpus, music bouts in the voices music corpus, top-tune instances, and top-voice instances. Because we expected music bouts, top tunes, and top voices to occur in bursty temporal patterns, we predicted that infants would encounter gap durations of varying durations, such that many gap durations would be relatively short and some would be considerably longer in duration. Given this predicted structure, it would not be sensible to describe these data using standard measures of central tendency. Instead, we determined the proportion of gap durations in each infants' day that occurred within 3 different duration bins – less than 60 seconds (< 1 minute), 60-3,599 seconds (1 minute to < 1 hour), and greater than 3,600 seconds (> 1 hour). We selected these particular duration bins because we wanted to know whether instances of music and its contents occurred spaced out in time mainly on the order of seconds, minutes, or hours. Determining the proportion of gap durations that occurred within each of these duration bins will reveal the durations across which infants must potentially integrate information from their everyday musical input.

Durations of music bouts, top-tune instances, and top-voice instances. We also examined the durations of individual music bouts as well as the durations of individual instances of infants' top tunes and top voices. For each, we calculated the duration as the number of seconds from the start of the bout (instance) to the end of the bout (instance). We predicted that music bouts, top-tune instances, and top-voices instances would have mainly short durations (i.e., on the order or seconds). As we did for gap durations, we reported the proportion of music bout (top-tune and top-voice) durations in each infants' day that occurred within several different duration bins – less than 60 seconds (< 1 minute), 60-3,599 seconds (1 minute to < 1 hour), and greater than 3,600 seconds (> 1 hour). We selected these particular duration bins because we wanted to know whether instances of music and its contents extended for seconds, minutes, or hours in time during infants' days.

Rank-one temporal pairs. Next, we examined how rank-one tune (voice) instances were related to other tune (voice) types. Specifically, we examined pairs of tune (voice) instances immediately adjacent in sequential order. We focused on forward pairs – a tune (voice) instance paired with the tune (voice) instance that immediately followed it. (We find the same pattern of results if we included only backward pairs or if we included both forward and backward pairs; see Supplemental Materials). We categorized tune (voice) instance pairs into 3 groups: 1) *Rank-one repetitions* were pairs in which one instance of the infant's rank-one tune (voice) type was followed by another instance of the infant's rank-one tune (voice) type. 2) *Rank-one anchors* were pairs in which one instance of the infant's rank-one tune (voice) type was followed by an instance of a different tune (voice) type. 3) *Non-rank-one pairs* were pairs in which neither instance was of the infant's rank-one tune (voice) type. We calculated the proportion of all adjacent pairs of tune (voice) instances that were each of these 3 categories for each infant. Figure 4.2 illustrates an example of how we calculated these 3 proportions. In Panel B, the teal-striped bars depict instances of the rank-one tune. If an infant encountered a high proportion of *rank-one repetitions*, then this would indicate that many of their consecutive tune (voice) types were repeated instances of their rank-one tune – this could provide periods dense with instances of their rank-one tune (voice), which might promote learning the details of one specific tune (voice) type. If an infant encountered a high proportion of *rank-one anchors*, then this would mean their rank-one tune (voice) often occurred paired in sequential order with other tunes (voices) – this could provide opportunities for the infant to compare and contrast between one dominant tune (voice) type and other, less-frequently-occurring tune (voice) types, which might foster learning about multiple tunes.

If an infant encountered a high proportion of *non-rank-one pairs*, then this would mean their rank-one tune (voice) did not frequently occur in their tune (voice) instance pairs – this structure would not engage the predicted spacing and interleaving learning mechanisms – and it might pose a larger challenge of an unstructured mix of input.

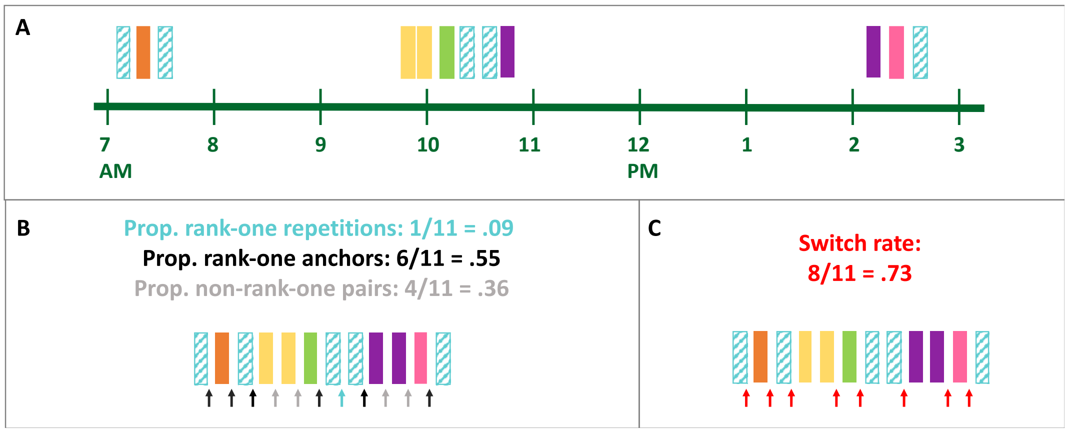

Figure 4.2. A schematic depicting (A) one example sequence of tune instances (bars) and their associated tune types (colors) as they occurred in time from 7:00am to 3:00pm. B) The rank-one tune is depicted by the teal, striped bars, and the instances of each rank-one pair category are indicated by the arrows – the teal arrow shows the rank-one repetition, black arrows show the rank-one anchors, and gray arrows show the non-rank-one pairs. The resulting proportions of each category are printed. (C) The sequential order of tune types in the same sequence of tune instance is also depicted divorced from the time of day. Red arrows indicate switches in tune type from one instance to the next. The resulting tune switch rate is shown (i.e., the proportion of all transitions that were switches).

Switch rate. To examine the sequential order of music in infants' everyday environments, we focused on the specific content of music bouts – the tunes and voices. We first made a list of all the tune (voice) instances and their associated tune (voice) types that occurred in each infant's day. We preserved the order in which the tune (voice) instances occurred in time from the beginning to the end of that day. For this analysis, we ignored both the duration of each instance and the actual time of day at which the instances occurred. Then, we determined whether each tune (voice) instance was a repetition or a switch in tune (voice) type compared to the tune (voice) instance that occurred immediately prior in the sequential order. We calculated the proportion of tune (voice) transitions that involved a switch from one tune (voice) type to a different tune (voice) type – the *switch rate*. As a proportion, *switch rate* could range from 0 to 1, with larger values indicating overall more switches than repetitions of tune (voice) types across tune (voice) instances. If infants had higher tune (voice) switch rates, then this would indicate that infants would more likely encounter a change in tune (voice) type from one tune (voice) instance to the next than a repetition of the same tune (voice) type – this would potentially present greater opportunities for infants to attend to differences between adjacent unique tunes (voices), comparing and contrasting unique tunes (voices) that occurred in the music of their everyday lives. If infants had lower tune (voice) switch rates, then this would indicate that from one tune (voice) instance to the next, infants would more likely encountered a repetition of the same unique tune (voice) – this would potentially present greater opportunities for infants to recognize what was the same across instances, building up detailed knowledge about one unique tune (voice). Figure 4.2 provides a schematic to illustrate how we calculated the switch rate. Panel A shows a set of music bouts – colored to depict the different tune types that occurred in each music bout – as they occur in time across a day that starts at 7:00am and ends at 3:00pm. Panel C preserves the sequential order of these tune instances but divorces them from the time of day. The red arrows indicate where the tune type (color) changes from one tune instance to the next. In this example, there were 8 switches out of 11 transitions, so the tune switch rate was .73.

Shuffled streams. As part of each analysis reported in this section, we compared the infants' real data to random data. The burstiness measure had this comparison built-in. For the other measures, we compared infants' real data to randomly shuffled streams.

To generate these shuffled streams, we retained the total number of music bouts, the total duration of individual music bouts, and the associated content (i.e., tune type) of each music bout from infants' real data. We first sampled randomly from the set of music bouts, without replacement to create a new order of music bouts (this step was important for subsequent analyses that relate to the order of musical content). Then, we placed the music bouts randomly in time within each infant's recorded time span (this step was important for subsequent analyses that relate to the temporal intervals of music). Doing so resulted in streams of music bouts with gaps of random durations between the music bouts and a randomized order of tunes and voices. For each analysis, we calculated the measure(s) of interest based on the shuffled streams. We repeated this process 100 times, calculating the median of each measure across all iterations for each infant. We used paired t-tests and two-sample Kolmogorov-Smirnov tests to compare findings based on infants' actual data to those based on the shuffled streams.

RESULTS

---------------------------------- TUNES MUSIC CORPUS ------------------------------------

In the present research, we aimed to discover the details of how music and its contents occurred in time across one day in infants' everyday lives, and we report the resulting distributions of each measure. We first answered each research question using the music bouts and the tunes in the tunes music corpus – the set of music bouts with exactly one tune. As reported in Chapter 2 of this thesis, across all 35 infants, the tunes music corpus contained 98,960 seconds of music that occurred in 4,023 separate music bouts. Based on these data, we found that individual infants encountered 19 to 343 separate music bouts during their days (Median = 98, SD = 73.24). Here, we examined the temporal pattern of how those music bouts occurred in time over the course of infants' days. For the remainder of this "Tunes Music Corpus" section, when we refer to "music bouts" we mean the individual instances of music that occurred within the tunes music corpus, and when we refer to "music", we mean the cumulative duration of the music instances within the tunes music corpus.

Did music bouts occur in regular, random, or bursty temporal patterns across infants' days?

To answer this question, we first calculated the inter-onset intervals ("IOIs") from the start of one individual music bout to the start of the subsequent music bout. Construing all IOIs across all infants as one dataset, we found that IOIs ranged from 3 seconds to 42,627 seconds (Median = 57 seconds, SD = 1,575.88 seconds). The durations of individual IOIs clearly varied such that most IOIs were short in duration on the order of seconds (e.g., .51 of all the IOIs were less than 60 seconds) while a small proportion of individual IOIs were considerably longer in duration on the order of hours (e.g., .02 of all the IOIs were greater than 3,600 seconds). The IOIs within each individual infant's day exhibited comparable variability – each infant encountered a biased mix of IOI durations, mostly shorter and some longer, between the music bouts that occurred over the course of their days (see Supplemental Materials). These findings warranted using burstiness to analyze the temporal clustering of music bouts in individual infants' days.

We next used each infant's set of IOIs to compute the burstiness parameter (eq. 2 here; Kim & Jo, 2016, equation 22). All infants encountered at least 4 music bouts, so data from all infants were included in this analysis. We discovered that music bouts occurred in a temporally bursty pattern across infants' days – the burstiness ($A_n(r)$) values of music bouts ranged from .23 to .90 (Median = .62, SD = .14; see Figure 4.3).

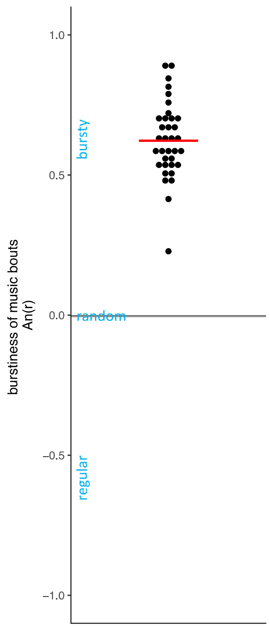

Figure 4.3. Infants encountered music bouts in the tunes music corpus in a bursty temporal pattern throughout their days, rather than a random or regular temporal pattern. The median burstiness ($A_n(r)$) value across infants (.62) is indicated by the red horizontal line. The longer gray horizontal lines depict the 95% CIs around the mean $A_n(r)$) value based on the simulated data.

To interpret these results, we compared these values to the values obtained by calculating the burstiness of simulated data streams that we generated from an exponential distribution. The mean burstiness across these simulated data streams was -.004, the lower 95% CI was -.006, and the upper 95% CI was -.002. We used these CI values as the lower and upper bounds of the random category in our bursty analysis of infants' actual music bout data. All of the burstiness values of infants' real music bout data were greater than the upper bound of the random category ($A_n(r)$ = -.002), which means music bouts occurred in a bursty temporal pattern, rather than in random or regular temporal patterns, across each infant's day. As infants advanced through their days, they encountered many music bouts close together in time as well as longer lulls during which no music bouts occurred.

How long were the gaps between music bouts in the tunes music corpus?

We also calculated the gap durations between music bouts – the number of seconds from the end of one music bout to the start of the next music bout. We first construed all of the gaps between music bouts across all infants' days (n=3,988 gaps) as one corpus (see Figure 4.4), and we discovered that music-bout gap durations ranged from 1 second to 42,566 seconds (Median = 29 seconds, SD = 1574.38 seconds). Most gaps between music bouts were short in duration on the order of seconds (e.g., .62 of the music-bout gap durations were shorter than 60 seconds), while some were considerably longer in duration on the order of hours (e.g., .04 of music-bout gap durations were 3,600 seconds or more). We noted a similar pattern of results for the IOIs between music bouts – it is reasonable, but not requisite, that IOIs and gaps would exhibit comparable patterns.

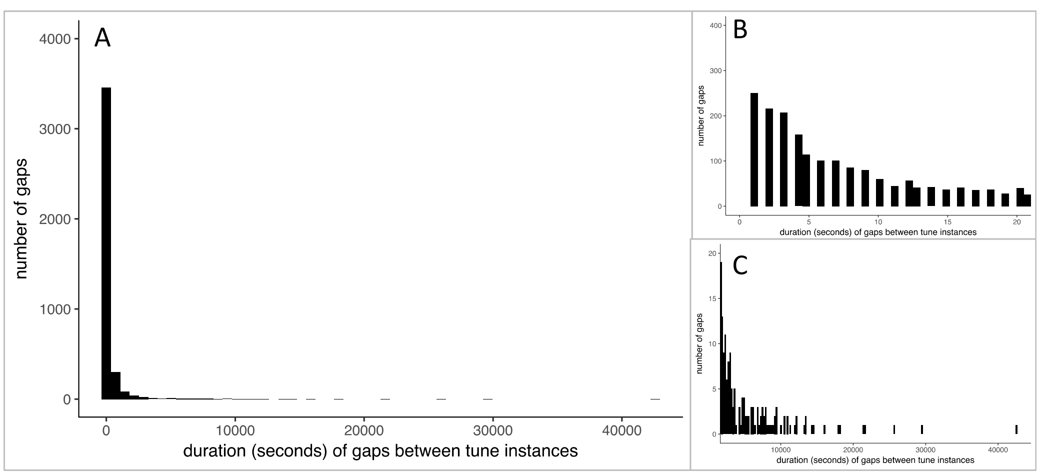

Figure 4.4. Most gaps between music bouts were very short in duration, while a some were considerably longer. Of (A) all of the gaps between music bouts across all infants' days (n=3,988), (B) .62 were shorter than 60 seconds in duration while (C) .04 were 3,600 seconds or more in duration.

We compared the corpus of infant's real music-bout gap data to the data of randomly shuffled streams. Each shuffled stream contained the same number of music bouts and the same total duration of individual music bouts, as that of one infant's real music data. We placed the music bouts randomly in time within each infant's recorded time span, creating a shuffled order of music bouts with gaps of random durations between the music bouts. We repeated this process 100 times, calculating the median of each measure across all iterations for each infant. We then compared the gap durations in the real data across infants to the gap durations of the shuffled data, using a two-sample Kolmogorov-Smirnov ("KS") test. The null hypothesis was that the two samples were drawn from the same distribution. The KS test results confirmed that, at the corpus level, the infants' real music-bout gap-duration data differed significantly from that in the shuffled streams (D = .64, p < .001). To gain further information about how these two distributions differed, we also calculated the proportion of these gaps between the shuffled music bouts that occurred in several duration windows. Table 4.1 reports the proportions of gaps between music bouts in each of several duration windows in infants' real data compared to the shuffled streams. In general, infants' real data consisted of a higher proportion of short-duration gaps between music bouts compared to data in the shuffled streams.

Table 4.1.
A higher proportion of music-bout gap durations were short (i.e., seconds level) in infants' real data relative to the shuffled data.

TUNES	REAL DATA		SHUFFLED DATA	
	Median	SD	Median	SD
seconds	0.56	0.11	0.13	0.09
minutes	0.42	0.09	0.84	0.07
hours	0.02	0.04	0.00	0.05

We also examined the gap durations between music bouts for individual infants. The median music-bout gap duration for each infant ranged from 4 seconds to 702 seconds. We also examined the proportion of music-bout gap durations within each infant's day that occurred for several different duration bins. Figure 4.5 clearly revealed that for each infant, approximately half of their music-bout gaps with relatively short durations (i.e., shorter than 60 seconds). For each infant, a small number of music-bout gaps also had considerably longer durations (i.e., longer than 3600 seconds). Discovering that each infant encountered a mix of gap durations between their music bouts is consistent with finding that music bouts occurred in a bursty temporal pattern. Thus, we provided congruent evidence that many music bouts occurred close together in time while some music bouts were separated by long lulls during which no music occurred.

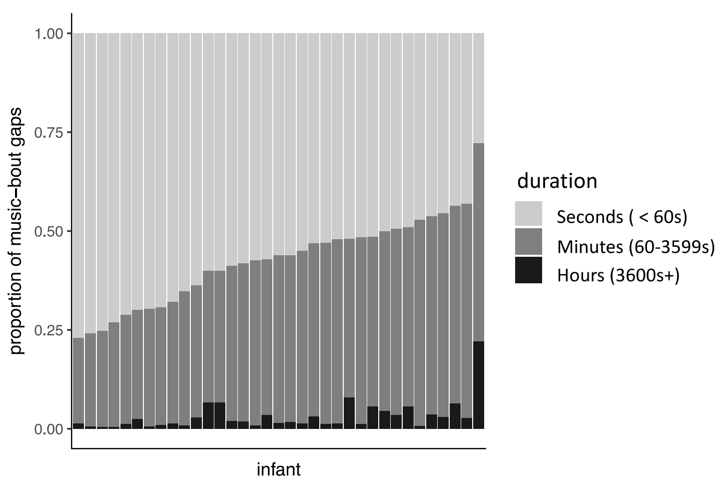

Figure 4.5. Individual infants (vertical bars) each encountered music-bout gaps with a mix of durations, ranging from seconds (light gray) to minutes (dark gray) to hours (black). Within each infant's day, most music bout gaps were of shorter durations (i.e., seconds and minutes), while a small proportion were considerably longer (i.e., hours).

How long were the individual music bouts in the tunes music corpus?

Across the entire tunes music corpus, infants encountered 4,023 separate music bouts. The duration of these individual music bouts ranged from 2 seconds to 673 seconds (Median = 12 seconds, SD = 40.24 seconds). Most music bouts (.91) in this tunes-music corpus had durations on the order of seconds (i.e., < 60 seconds), and there were no music bouts with durations on the order of hours. For individual infants, we reported the proportion of music-bout durations that occurred for three different duration bins. Figure 4.6 shows that each infant encountered mostly short music-bout durations (i.e., on the order of seconds). Because the bulk of the music bout durations in each infant's day were short, the bursty temporal patterns observed were unlikely to be a result of varying durations of the music bouts and were therefore more strongly tied to the durations of gaps between the music bouts.

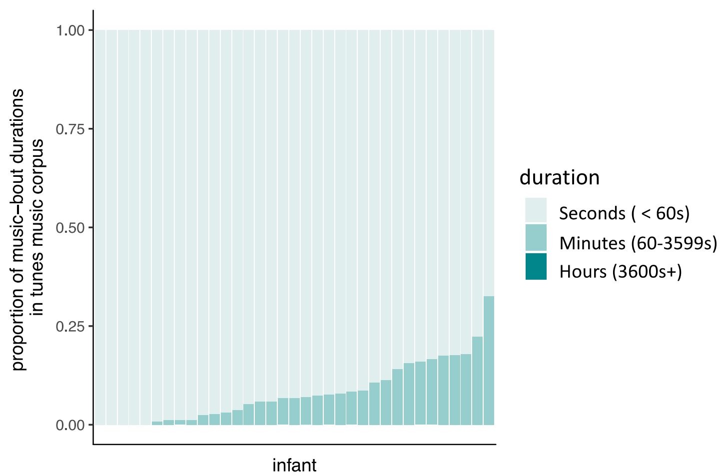

Figure 4.6. Individual infants (vertical bars) encountered music bouts of mainly short durations (i.e., < 60 seconds; lightest teal) in the tunes music that occurred in their everyday environments. Some infants encountered a small proportion of music bouts with durations on the order of minutes (medium teal) but no music bouts with durations of more than an hour (dark teal).

A day of bursty music bouts in the life of one baby.

To give a sense of the measures we have reported up to this point, let's look at a day in the life of one infant who encountered 953 seconds of music across 85 music bouts during their recorded day. This number of music bouts was approximately the same as the median number of music bouts (in the tunes music corpus) across all infants (Median

= 91 music bouts). Panel A of Figure 4.7 shows the frequency distribution of this infant's music-bout durations. Most music bouts were very short in duration (Median = 7 seconds, SD = 11.36 seconds). All but one of this infant's music bouts were shorter than 60 seconds. Panel B of Figure 4.7 shows the frequency distribution of the gap durations between the music bouts that occurred in this infant's day. Most of the gap durations were short (i.e., .37 were shorter than 20 seconds in duration), and two gap durations were very long (i.e., .02 were longer than 3 hours). These frequency distributions of music-bout durations and gap durations suggest that this infant's music bouts occurred in a bursty temporal pattern. They did: the burstiness parameter for the music bouts in this infant's day was $A_n(r)$ = .71. What does this really mean? Panel C of Figure 4.7 depicts the infant's music as it occurred in time from 5:00am to 10:00pm during the infant's recorded day. Each bar depicts one second in the day, and the color of each bar shows what was occurring during that second. This family started recording around 5:45am and left the recorder on continuously until about 9:45pm. The dark gray portions at the beginning and end of the time series show when the recorded was turned off. The turquoise bars show each second of music as it occurred in time over the course of this infant's day. The light peach shows portions of the day that were coded (but not judged to be music), and the light gray depicts portions of the day that fell below the decibel threshold and were therefore not coded. The bursty temporal pattern of music bouts in this infant's day pops out by examining how music bouts and the gaps between music bouts occurred in time. The short durations of individual music bouts are clearly depicted by the very narrow horizontal extent of each section of the turquoise bars. The two very long gaps clearly occurred between the music bouts that occurred before and after the long, very quiet portions of the day (i.e., not coded, in gray). These very long gaps may signal coherence in that they set the outer bounds of how activities might be separated as they occurred across this infant's day. The remaining gaps were much shorter, which is particularly clear in Panel C between about 7:00am and 9:00am where each of the light peach sections between the music bouts (turquoise) is very narrow in its horizontal extent. Taken together, this infant encountered many short bouts of music early in the morning and then a long lull during which no music occurred. Then, they encountered another round of many short music bouts in the early afternoon, followed by another long lull during which no music occurred. Their day ended with a few additional short music bouts in the evening, and then one final lull where no music occurred before the recorded was turned off.

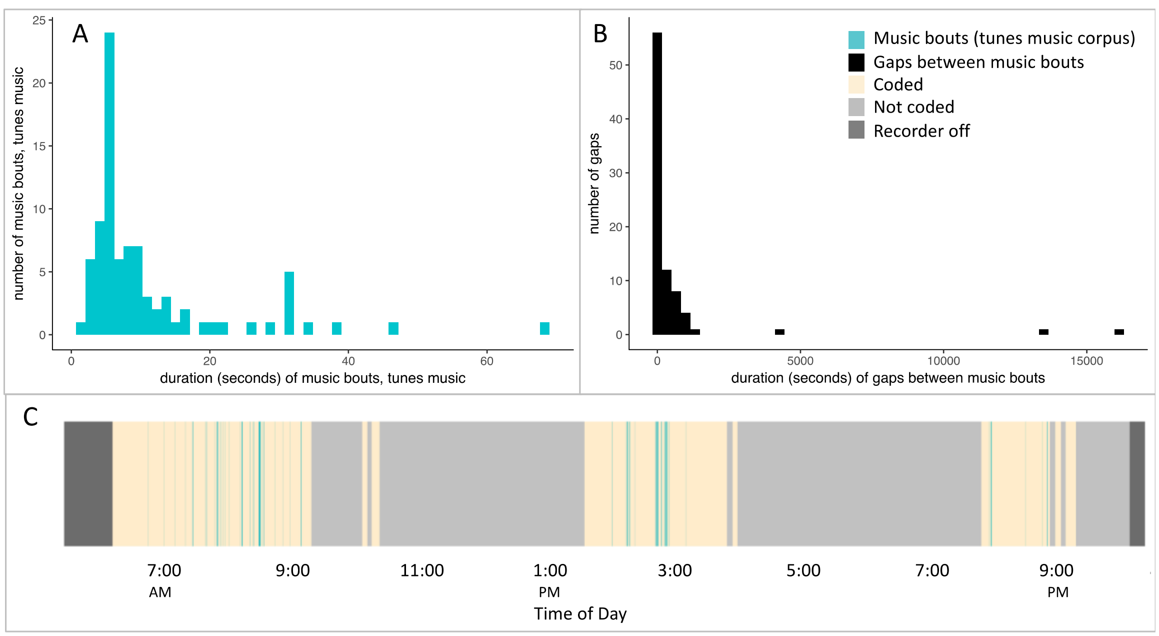

Figure 4.7. Music bouts (tunes music corpus) in one example infant's day were mostly of short durations (A) with some short and some longer gaps between them (B). The bursty temporal pattern ($A_n(r)$ = .71) of the music bouts in this infant's day is clear – in (C), music (turquoise) is plotted in time as it occurred over the course of infant's day from 5:00am to 10:00pm. (Dark gray = recorder off, Light gray = not coded, light peach = coded, but not music).

Did infants' top tunes occur in regular, random, or bursty temporal patterns across infants' days?

We have established that daily music occurred in a bursty temporal pattern. We were also interested in how the specific content of music occurred in time over the course of infants' days. Therefore, we next examined the temporal dynamics of one specific tune in each infant's day. As one example, we selected infants' top tunes (i.e., their "rank-one" tune type) for this analysis. It did not have to be the case that infants' top tunes also occurred in a bursty temporal pattern. To assess whether infants' rank-one tune type occurred in regular, random, or bursty temporal patterns, we first determined the number of separate instances that each infant's rank-one tune occurred. Infants encountered their rank-one tune from 1 to 88 times in the tunes music that occurred in their day (Median = 6 instances, SD = 18.00 instances). Because we defined the rank-one tune type as the tune type that occurred for the longest cumulative duration, it was possible for infants to encounter their rank-one tune in one single, continuous instance. This was true for 6 infants (.17). After their rank-one tune type occurred once, these infants never encountered another instance of their rank-one tune – at least not within their one recorded day. Instead, they had only very local integration demands that involved tracking this tune for the duration of its single instance. Since these infants had no IOIs for their rank-one tune, we did not include them in the burstiness analysis. Six additional infants encountered fewer than 4 instances of their rank-one tune, so we also excluded data from these infants from the following analysis of burstiness.

Thus, the burstiness analysis for infants' rank-one tune instances included data from 23 infants (.66). The burstiness ($A_n(r)$) values for infants' rank-one tunes ranged from -.60 to .96 (Median = .35, SD = .41, see Figure 4.8). As we did for the music bouts burstiness analysis, we compared these values to those of simulated data streams generated from an exponential distribution. We generated 23 simulated data streams with the same number of simulated times as the number of IOIs in each of the included real data streams. The mean burstiness of the simulated data was .006, the lower 95% CI was -.004, and the upper 95% CI was .016. Given these bounds, most of these infants (.91) encountered their rank-one tune in a bursty temporal pattern across their days. No infants encountered instances of their rank-one tune randomly in time. For two infants (.09), the rank-one tune instances occurred in a more regular temporal pattern.

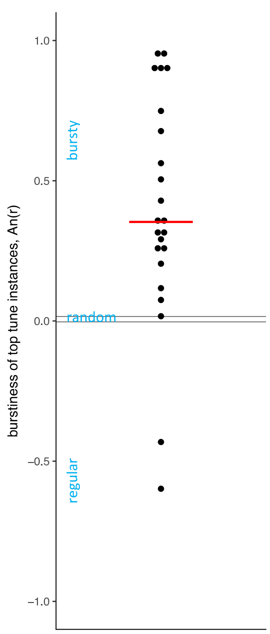

Figure 4.8. Infants encountered instances of their top tune in a bursty temporal pattern throughout their days, rather than a random or regular temporal pattern. The median burstiness ($A_n(r)$) value across infants is indicated by the red horizontal line. The longer gray horizontal lines depict the 95% CIs around the mean $A_n(r)$) value based on the simulated data.

Taken together, these analyses revealed that as infants moved through their days, they encountered many music bouts close together in time and also longer lulls during which no music bouts occurred. Likewise, most infants encountered many repeated instances of their top tune close together in time followed by long lulls during which their top tune did not occur.

How long were the gaps between instances of infants' top tunes?

Here, again, we were interested in how the content of music occurred in time over the course of infants' days. We examined the duration of gaps between instances of infants' top tune. Construing all of the gaps between instances of infants' rank-one tune across all days (n=348) as one corpus (see Figure 4.9), we discovered that rank-one tune gap durations ranged from 1 second to 32,711 seconds (Median = 5 seconds, SD = 3,320.21 seconds). A large proportion of rank-one tune gaps durations were short (e.g., .78 of the rank-one tune gap durations were shorter than 60 seconds), while others were considerably longer in duration (e.g., .03 of rank-one tune gap durations were 3,600 or more seconds).

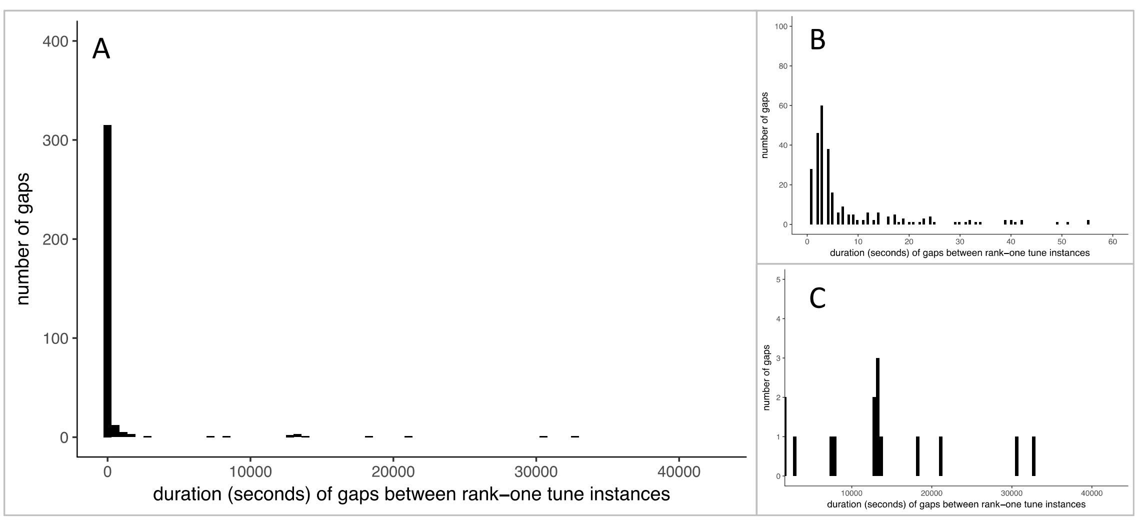

Figure 4.9. Most gaps between rank-one tune instances were very short in duration, while a some were considerably longer. Of (A) all of the gaps between rank-one tune instances across all infants' days, (B) most (.78) were 60 seconds or less while (C) some (.02) were 3,600 seconds or more in duration.

We also compared the corpus rank-one tune gap durations across infants' data to those of the randomly shuffled streams. We used the same randomly shuffled streams that we generated to compare to infants' real music bout data. Because we preserved the duration of each music bout and the links between duration and tune type when we generated the shuffled streams, the rank-one tune types in the infants' real data were also the tune types that occurred for the longest cumulative duration in the shuffled streams. Therefore, we determined the gap durations between the instances of these shuffled rank-one tune types. As we did for gap durations between music bouts, we calculated the median across iterations for the gap durations between each rank-one tune instance in the shuffled corpus. We then compared the gap durations in the corpus of infants' real data to the durations of the shuffled data. The KS test results revealed that distribution of the corpus of infants' real rank-one tune gap-duration data differed significantly from that of the shuffled stream data (D = .93, p < .001). We further examined this difference by determining the proportion of rank-one tune gap durations that were within the same set of duration windows we defined for music-bout gap durations. Table 4.2 reports the proportions of gaps between rank-one tune instances in infants' real data compared to those of the shuffled streams. Infants' real data consisted of a higher proportion of short-duration gaps between rank-one tune instances than did the data in the shuffled streams.

Table 4.2.
A higher proportion of rank-one tune gap durations were short (i.e., seconds level) in infants' real data than in the shuffled data.

TOP TUNES	**REAL DATA**	**SHUFFLED DATA**		
	Median	SD	Median	SD
seconds	0.83	0.31	0.00	0.02
minutes	0.01	0.23	0.40	0.28
hours	0.00	0.20	0.60	0.30

We also examined the proportion of gaps between instances of infants' rank-one tunes that occurred within each duration bins for each individual infant. It is clear in Figure 4.10 that gap durations between instances of infants' rank-one tunes varied considerably across individual infants. Even with this variation, for most infants, the majority of gap durations between instances of infants' rank-one tune occurred on the order of seconds. Six infants (.17) had more than half of their rank-one tune gap durations occur on the order of minutes and/or hours.

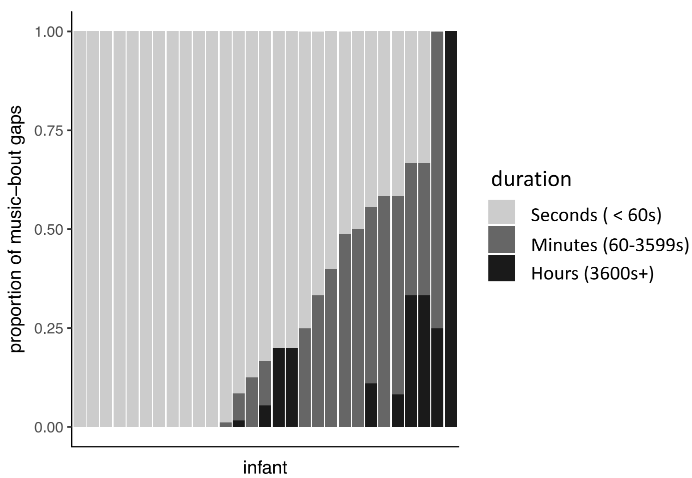

Figure 4.10. Individual infants (vertical bars) each encountered rank-one tune gaps with a mix of durations, ranging from seconds (light gray) to minutes (dark gray) to hours (black). Within each infant's day, most rank-one tune gap durations were short (i.e., seconds), while some were considerably longer (i.e., minutes or hours).

How often did the top tune occur paired in sequential order with other tunes?

Since the gaps between instances of infants' top tune were mostly short in duration, it raised the possibility that infants were encountering lots of repeated instances of their top tune. We also discovered evidence of longer gap durations between instances of infants' top tune, which raises the question: what, if anything, did infants encounter in between those instances? To address these questions, we examined the proportion of tune instance pairs (i.e., 2 consecutive tune instances) that were *rank-one repetitions* (i.e., an instance of the rank-one tune followed by another instance of the rank-one tune), rank-one anchors (i.e., an instance of the rank-one tune followed by an instance of a different tune type), or *non-rank-one pairs* (i.e., 2 consecutive tune instances that were not of the rank-one tune). Figure 4.11 shows the proportion of *rank-one repetitions* (turquoise) and of *rank-one anchors* (black) that occurred in each infants' day (the remaining proportion represents the proportion of non-rank-one pairs). We discovered that most tune instance pairs in infants' everyday music actually did not involve an instance of the rank-one tune (Median = .94, SD = .10, range: .56 - .996). When the rank-one tune was involved, it was more likely to occur followed by a repeated instance of itself (i.e., *rank-one repetitions*, Median = .04, SD = .10, range: 0-.44), than to occur followed by an instance of a different tune type (i.e., *rank-one anchors*, Median = .02, SD = .03, range: .003-.13).

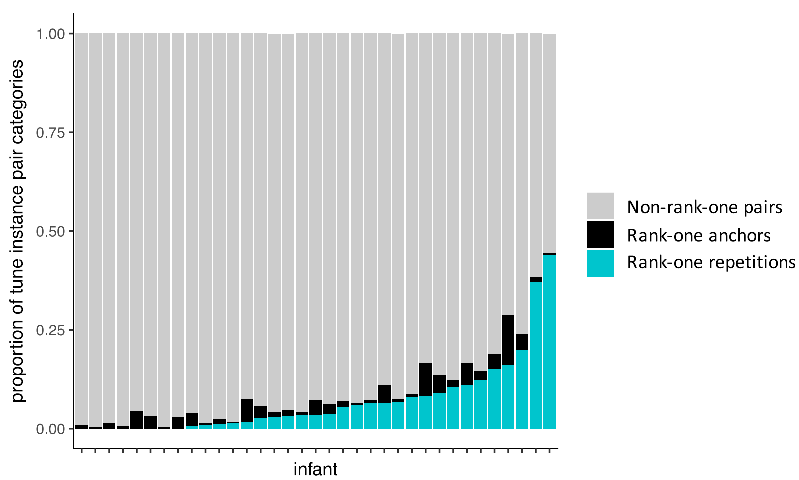

Figure 4.11. Infants mostly encountered pairs of tune instances that did not involve their rank-one tune (*non-rank-one pairs*, gray). When they did encounter their rank-one tune, most infants were more likely to encounter an instance of their rank-one tune paired in sequential order with itself (*rank-one repetition*, turquoise) than an instance of their rank-one tune paired in order with a different tune type (*rank-one anchor*, black).

Figure 4.12 shows infants' rank-one tunes in sequential order across each infant's day. Each row represents one infant, and each dot represents one tune instance. The x-axis is the tune instance number, meaning that instances appear in the sequential order they occurred during infants' days (divorced from the actual time of day). Rank-one tune instances are shown in turquoise, and instances of all other tune types are shown in black. Thus, any instance of the rank-one tune that followed another instance of the rank-one tune (i.e., two consecutive turquoise dots) depicts a *rank-one repetition.* Any instance of a different unique tune that followed an instance of the rank-one tune (i.e., a black dot following a turquoise dot) shows a *rank-one anchor*. *Non-rank-one pairs* are depicted by consecutive black dots. Baby 35 is the top-most row. The first (left-most) dot depicts the first tune instance that occurred in Baby 35's day. The next dot depicts the second tune instance in Baby 35's day. Both of these dots are turquoise, which means this infant encountered 2 instances of their rank-one tune paired in sequence at the beginning of their day. The second instance was thus a *rank-one repetition*. Baby 35 encountered 5 *rank-one repetitions* in a row in their first 6 tune instances. The 7^{th} tune instance that occurred was a different tune type than their rank-one tune (depicted as a black dot). Because it followed an instance of the infant's rank-one tune, it was a *rank-one anchor*. Across all infants (all rows), it is clear that infants' rank-one tune did not occur frequently; when it did, infants encountered many instances of their rank-one tune in a row. This contributed to the higher proportions *of rank-one repetitions* relative to the proportions of *rank-one anchors.*

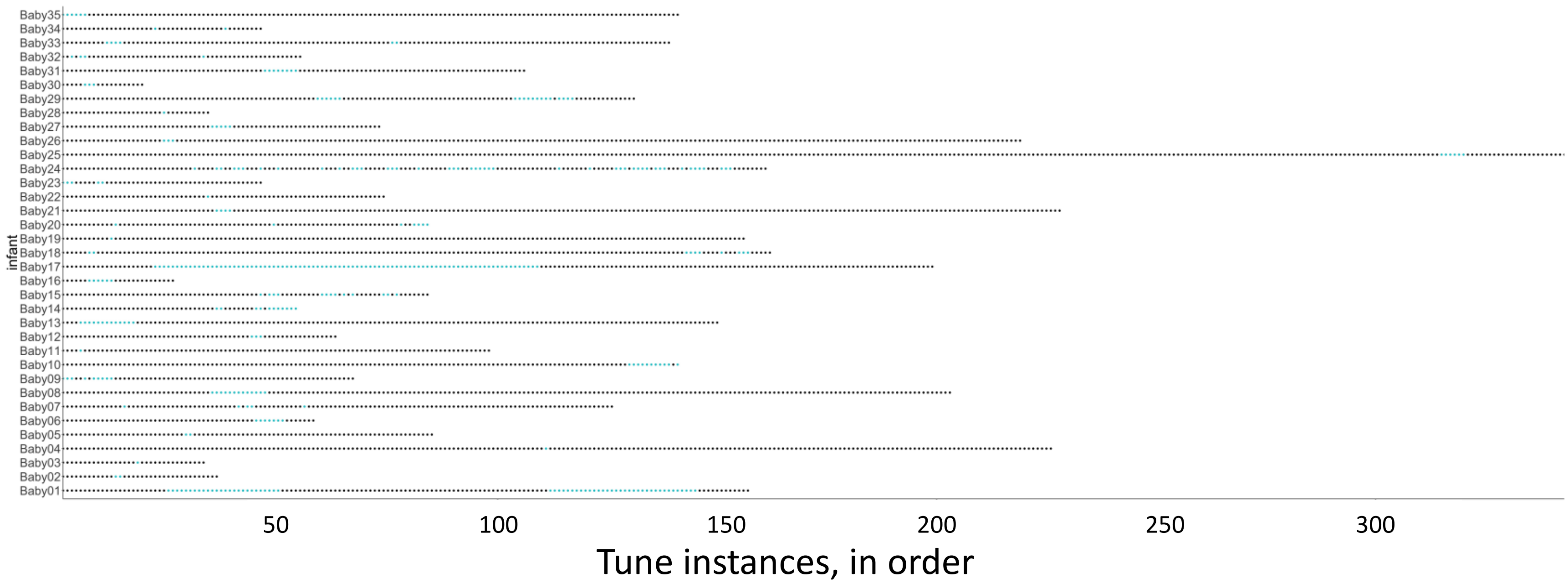

Figure 3.12. By plotting tune instances as they occurred in order in each infants' day, with rank-one tune instances highlighted (teal dots), it is clear that rank-one repetitions (two sequential teal dots) and rank-one anchors (teal dot followed by a black dot) were both relatively rare.

We also compared rank-one tune pairs in infants' real tune data to that of randomly shuffled data streams. We examined rank-one tune pairs on the same shuffled data streams we created for evaluating the tune switch rate. As described earlier, this process yielded the same set of tune instances and associated tune types in a randomly shuffled order. Here, we calculated the proportion of each tune instance pair type (i.e., *rank-one repetitions*, *rank-one anchors*, and *non-rank-one pairs*), considering the same rank-one tune type in infant's real data to be the *shuffled rank-one tune type* in the matched shuffled data stream. We repeated this process 100 times, calculating the median proportions of each pair type across all iterations for each infant. Table 4.3 shows how the infants' real data compared to the shuffled streams for each type of tune instance pair. The bulk of the distribution in both infants' real data and in the shuffled data was approaching the upper limit. Therefore, we next analyzed the overall daily switch rate of tunes to characterize the repetition and change that occurred for the whole mix of tunes.

Table 4.3.
Proportions of rank-one pair types in the tune instance pairs of infants' real data and in the shuffled data.

TUNES	INFANTS' REAL DATA		SHUFFLED DATA	
	Median	SD	Median	SD
proportion rank-one repetitions	0.04	0.10	0.00	0.01
proportion rank-one anchors	0.02	0.03	0.02	0.03
proportion non-rank-one pairs	0.94	0.10	0.98	0.04

Taken together, these findings revealed that as an infant advanced through their day, whenever they encountered their top tune, they were more likely to encounter it again the very next time a tune occurred than to encounter a different specific tune the next time a tune occurred. These results also highlight that infants did not have many separate opportunities to encounter their top tune in their day. Together, these findings raise more questions than they answer about the potential (somewhat limited) opportunities for infants to compare and contrast their top tune with other tunes over the course of their days. We will further address this topic in the discussion.

How often did tune types switch over the course of infants' days?

Since most tune instance pairs did not involve infants' top tunes, then this raised the question of what did occur from one tune instance to the next. To gain further insight into the nature of the non-rank-one pairs, we answered the question: how often did tune types switch over the course of infants' days? For this analysis, we did not consider the time of day or the duration between tune instances. We calculated the number of times the tune type changed from one tune instance to the next. Because each infant encountered a different cumulative number of tune instances over the course of their days, we calculated the *tune switch rate* as a proportion: the cumulative number of tune instances that were a change from the immediately prior tune instance divided by the cumulative number of all tune instances in the infant's day. The *tune switch rate* ranged from .33 to .94 (Median = .67, SD = .16; see Figure 4.13). This finding means that infants more often encountered a change in tune type from one tune instance to the next than a repeated instance of the same tune type. For example, if an infant encountered an instance of *Itsy Bitsy Spider*, then the next time a tune occurred, they would be more likely to encounter an instance of *Wheels on the Bus* or *If You're Happy and You Know It* than another instance of *Itsy Bitsy Spider*. Infants encountered a high switch rate for tunes, because a high proportion of their daily tune types occurred in only one instance (Median = .68, SD = .15, range: .30 - .90), and therefore could never occur as a repeated instance.

While most infants encountered a high tune switch rate, some infants (.20) encountered the reverse – they were more likely to encounter the same tune type repeated across consecutive tune instances than to encounter switches in tune type from one tune instance to the next.

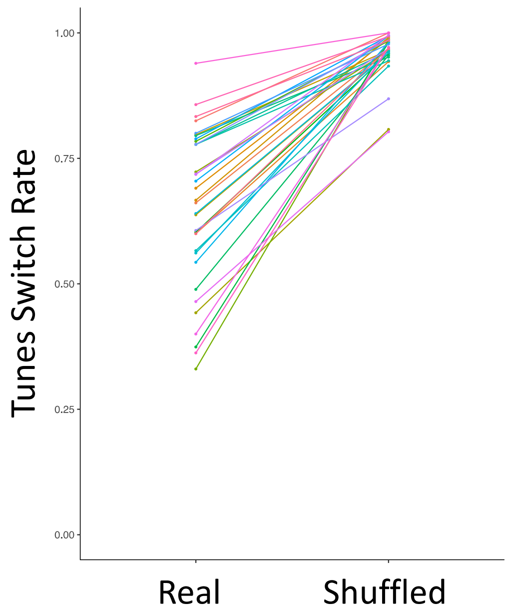

Figure 4.13. Even though infants were more likely to encounter a change in tune type from one tune instance to the next than a repeated instance of the same tune type, they still encountered more repetition of tune types (i.e., lower switch rates) in their real tune data compared to the shuffled streams.

As we did for the analyses of gap durations, we also compared switch rate of infants' real tune data to that of randomly shuffled data streams. We used the same randomly shuffled data streams as described earlier. By shuffling the order of the music bouts when we created the shuffled streams, we generated new orders of tune instances and their associated tune types. We determined the tune switch rate based on these shuffled orders in 100 iterations, and then we calculated the median *shuffled tune switch rate* across iterations for each shuffled stream. The resulting *shuffled tune switch rate* was even higher than the switch rate for tunes in infants' real data, ranging from .80 to 1 (Median = .97, SD = .05). This higher switch rate occurred because the random shuffling disrupted the sequences of repeated instances of tune types that did occur more than once. A paired t-test ($t(34) = -12.16$, $p < .001$) confirmed that the shuffled tune data had a significantly high switch rate (Median = .97) compared to infants' real tune data (Median = .67). Therefore, even though infants encountered a high proportion of tune instances changed in tune type from one tune instance to the next, they also encountered more repetitions of the same tune type across tune instances than would be expected randomly.

Taken together, our analyses of the music bouts, tunes, and top tunes yielded data about the temporal properties of infants' everyday musical input that likely matter for how infants build up knowledge about music and tunes in their everyday lives. Infants encountered most music and tunes within seconds of when they had last encountered music and tunes, and they also encountered long lulls, on the order of hours, during which no music occurred. As infants moved through their days, they encountered mostly different tunes from one tune instance to the next. However, infants also more often encountered repetitions of the same tune than would be expected by chance. Thus, the temporal intervals and sequential orders at which music and infants' top tunes occurred varied across infants' days. We next answered the same set of questions about temporal dynamics, focusing on the voices music corpus and top voices that infants encountered across their days.

---------------------------------- **VOICES MUSIC CORPUS** ------------------------------------

We next addressed the same set of questions about the music bouts and the voices in the voices music corpus – the set of music bouts with exactly one voice. As presented in Chapter 1 of this thesis, across all 35 infants, the voices music corpus contained 63,490 seconds of music that occurred in 2,149 separate music bouts. Based on these data, we discovered that individual infants encountered 7 to 188 vocal music bouts that occurred during their days (Median = 44, SD = 45.69). Here, we examined the temporal pattern of how the music bouts of the voices music corpus occurred in time over the course of infants' days. For the remainder of this "Voices Music Corpus" section, when we refer to "music", we mean the voices music corpus, and when we refer to "music bouts" we mean the individual instances of music that occurred within the voices music corpus.

Did music bouts occur in regular, random, or bursty temporal patterns across infants' days?

First, we calculated the inter-onset intervals ("IOIs") from the start of one individual music bout to the start of the subsequent vocal music bout. Construing all IOIs across all infants as one dataset, we found that IOIs for music bouts in the voices music ranged from 4 seconds to 43,127 seconds (Median = 96 seconds, SD = 2,226.13 seconds). This general pattern of a mix of shorter and longer IOIs between vocal music bouts appeared within each individual infant's day (see Supplemental Materials). Thus, we used the burstiness parameter to analyze temporal clustering of vocal music bouts that occurred in infants' days.

We next determined the temporal clustering of music bouts in the voices music, using each infant's set of IOIs to compute the burstiness parameter (eq. 2, here; Kim & Jo, 2016, equation 22). All infants encountered at least 4 music bouts, so we included data from all infants in this analysis. Music bouts in infants' voices music occurred in a temporally bursty pattern across infants' days – burstiness ($A_n(r)$) ranged from .21 to .96 (Median = .53, SD = .15; see Figure 4.14).

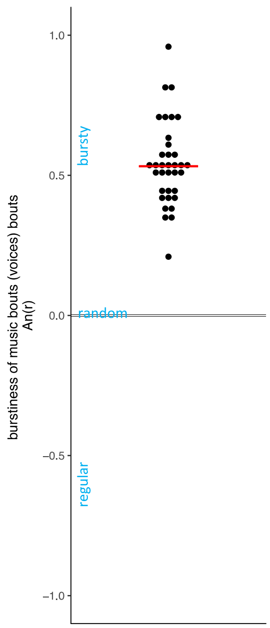

Figure 4.14. Infants encountered music bouts in the voices music corpus in a bursty temporal pattern throughout their days, rather than a random or regular temporal pattern. The median burstiness ($A_n(r)$) value across infants is indicated by the red horizontal line. The gray horizontal lines depict the 95% CIs around the mean $A_n(r)$) value based on the simulated data.

To interpret these results, we compared these values to the values obtained by calculating the burstiness of simulated data streams that were generated from an exponential distribution. For simulations based on infants' vocal music data, the mean burstiness ($A_n(r)$) was -.0001, the lower 95% CI was -.003, and the upper 95% CI was .003. We used these CI values as the lower and upper bounds of the random category in our bursty analysis of infants' actual vocal music bout data. All of the burstiness values of infants' real music bout data were greater than the upper bound ($A_n(r)$ = .003) – thus, voices music bouts occurred in a bursty temporal pattern, rather than in random or regular temporal patterns, across each infant's day. As infants moved through their days, they encountered many voices music bouts close together in time followed by relatively long lulls during which no voices music bouts occurred.

How long were the gaps between music bouts in the voices music corpus?

As we did for tunes, we next examined the duration of the gaps between music bouts in the voices music corpus – the number of seconds from the end of one music bout to the start of the next music bout. Construing all of the gaps between music bouts that occurred across all infants' days (n=2,114 gaps) as one corpus (see Figure 4.15), we found that gap durations between music bouts ranged from 1 second to 43,066 seconds (Median = 60 seconds, SD = 2,223.40 seconds). A considerable proportion of gaps between music bouts were short in duration (e.g., .49 of the voices-music-bout gap durations were shorter than 60 seconds), while others were much longer in duration (e.g., .04 of voices-music-bout gap durations were 3,600 or more seconds).

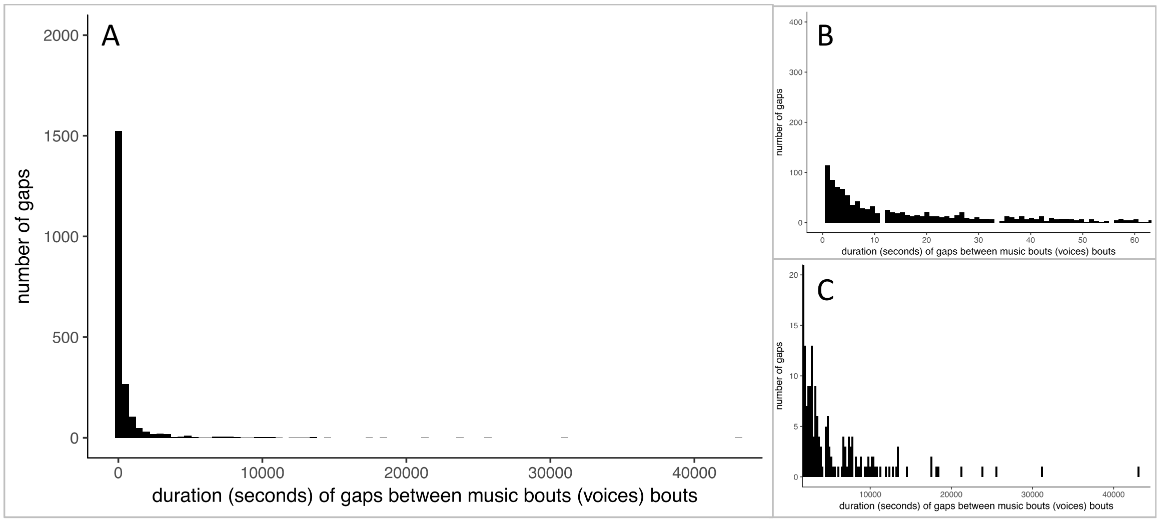

Figure 4.15. Many gaps between voices music bouts were short in duration, while some were much longer. Of (A) all of the gaps between voices music bouts across all infants' days (n=2,114 gaps), (B) Nearly half (.49) were shorter than 60 seconds in duration while (C) .04 were 3,600 seconds or more in duration.

We also compared the corpus of infant's real music-bout gap data to that of the randomly shuffled streams. We used the same shuffled streams that we generated earlier for music bouts, and we calculated the proportion of gaps between the real and shuffled music bouts that occurred in each duration window. We repeated this process 100 times, calculating the median of each measure across all iterations for each infant. We calculated the median across iterations for the gap durations between each music bout in the shuffled corpus. Then, we used a two-sample Kolmogorov-Smirnov ("KS") test to compare the gap durations in the infants' real data to the median durations of the shuffled data. The null hypothesis was that the two samples were drawn from the same distribution. The KS test results revealed that the infants' real music-bout gap-duration data differed significantly from that in the shuffled streams (D = .59, p < .001). For additional information about how these two distributions differed, we also examined the proportion of gap durations that occurred in each duration window based on the shuffled streams. In Table 4.4, we listed the results for how the proportion of gaps between voices music bouts in infants' real data compared to the shuffled streams. In general, infants' real data consisted of a higher proportion of short-duration gaps and a higher proportion of very long gap durations between music bouts compared to data in the shuffled streams.

Table 4.4.
Infants encountered a higher proportion of short durations of gaps between voices music bouts compared to the gaps that occurred in the shuffled data.

VOICES	**REAL DATA**		**SHUFFLED DATA**	
	Median	SD	Median	SD
seconds	0.42	0.15	0.05	0.06
minutes	0.56	0.12	0.87	0.13
hours	0.05	0.06	0.02	0.13

We then examined the duration of gaps between music bouts that occurred in individual infant's days. The median gap duration between music bouts for each infant ranged from 14 seconds to 954.5 seconds. We also examined the proportion of gap durations between music bouts that occurred for several different windows of durations. Figure 4.16 shows that for most infants, roughly half of the gap durations were shorter (i.e., on the order of seconds) and roughly half were longer (i.e., on the order of minutes). Every infant also encountered a small proportion of gap durations between music bouts that were considerably longer (i.e., on the order of hours). The finding that each infant encountered a mix of durations of the gaps between their voices music bouts is consistent with the result that voices music bouts occurred in a bursty temporal pattern. These results reveal that many voices music bouts occurred close together in time while some music bouts were separated by much longer lulls during which no voices music occurred.

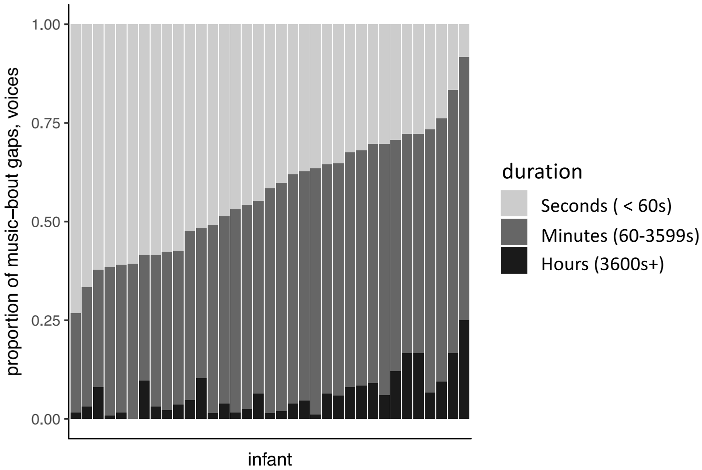

Figure 4.16. Individual infants (vertical bars) each encountered voices-music-bout gaps with a mix of durations, ranging from seconds (light gray) to minutes (dark gray) to hours (black). Within each infant's day, the bulk of the voices music bout gaps were of shorter durations (i.e., seconds and minutes), and a small proportion were on the order of hours.

How long were the individual music bouts in the voices music corpus?

In the entire voices music corpus, infants encountered 2,149 individual music bouts. The duration of these separate music bouts ranged from 2 seconds to 827 seconds (Median = 14 seconds, SD = 51.62 seconds). Most music bouts (.89) in this voices-music corpus had durations on the order of seconds (i.e., < 60 seconds), and none of the individual music bouts had durations on the order of hours. For individual infants, we examined the proportion of music-bout durations that occurred for several different duration bins. Figure 4.17 reveals that each infant encountered mostly music bouts with relatively short durations (i.e., on the order of seconds). We reported the median proportions of music bouts in each duration window in the Supplemental Materials.

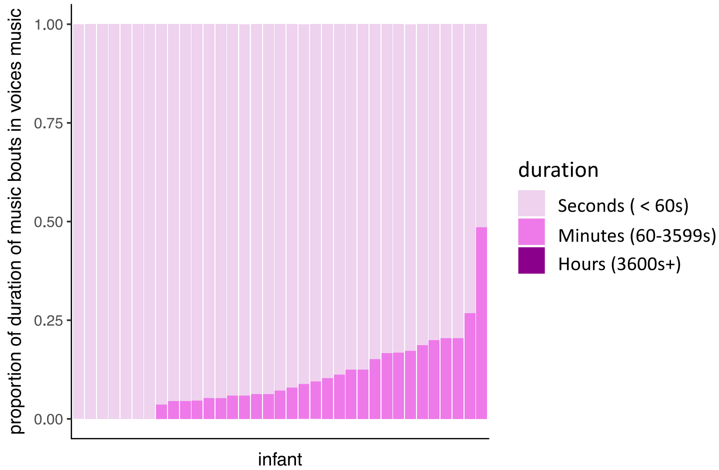

Figure 4.17. Infants encountered music bouts of short durations (i.e., < 60 seconds; lightest magenta) in the voices music that occurred in their everyday environments. Infants encountered a small proportion of music bouts with durations on the order of minutes (medium magenta) and no music bouts that had durations of an hour or more (dark magenta).

Did infants' top voices occur in regular, random, or bursty temporal patterns across infants' days?

As we did for tunes, we also examined the temporal dynamics of one specific voice in each infant's day, focusing on infants' top voices (i.e., their "rank-one" voice type). To determine whether infants' rank-one voice type occurred in regular, random, or bursty temporal patterns, we first found the number of separate instances that each infant's rank-one voice occurred. Infants encountered their rank-one voice from 1 to 138 times in their day (Median = 18 instances, SD = 25.39 instances). One infant (.03) encountered their rank-one voice in one single, continuous instance. This infant never re-encountered an instance of their rank-one voice after the first time it occurred in their day. We excluded the data from this infant from this burstiness analysis. Two additional infants encountered fewer than 4 instances of their rank-one voice, so we also excluded data from these infants in the following analysis of burstiness.

We computed the burstiness parameter for infants' rank-one voice instances for 32 infants (.91). The burstiness ($A_n(r)$) values for infants' rank-one voices ranged from .20 to .97 (Median = .57, SD = .20, see Figure 4.18). Again, we compared these values to those of simulated data streams generated from an exponential distribution. Based on the 32 simulated data streams we generated, the mean burstiness ($A_n(r)$) was -.005, the lower 95% CI was -.010, and the upper 95% CI was .0003. Given these bounds, all infants encountered their rank-one voice in a bursty temporal pattern across their days.

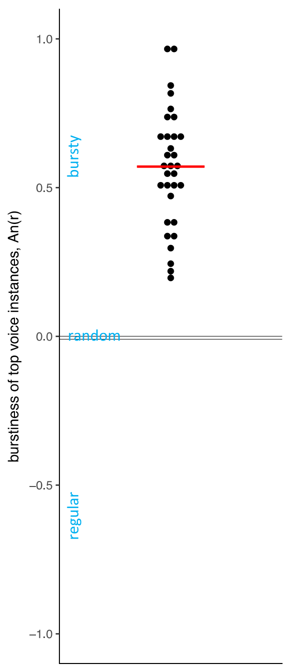

Figure 4.18. Infants encountered instances of their top voices in a bursty temporal pattern throughout their days, rather than a random or regular temporal pattern. The median burstiness ($A_n(r)$) value across infants is indicated by the red horizontal line. The longer gray horizontal lines depict the 95% CIs around the mean $A_n(r)$) value based on the simulated data.

In sum, these analyses revealed that as they advanced through their days, infants encountered many voices music bouts clustered in time, followed by longer lulls during which no voices music bouts occurred. Likewise, infants encountered many repeated instances of their top voice close together in time followed by long lulls in which their top voice did not occur in music.

How long were the gaps between instances of infants' top voices?

To illustrate how the content of voices music occurred in time across infants' days, we examined the duration of gaps between instances of infants' top voice. We construed all of the gaps between instances of infants' rank-one voice across all days as one corpus (see Figure 4.19), and we discovered that these gap durations ranged from 1 second to 44,050 seconds (Median = 85 seconds, SD = 3642.22 seconds). A large proportion of rank-one voice gaps durations were short (e.g., .44 of the rank-one voice gap durations were shorter than 60 seconds), while many others were considerably longer in duration (e.g., .08 of rank-one voice gap durations were 3,600 or more seconds).

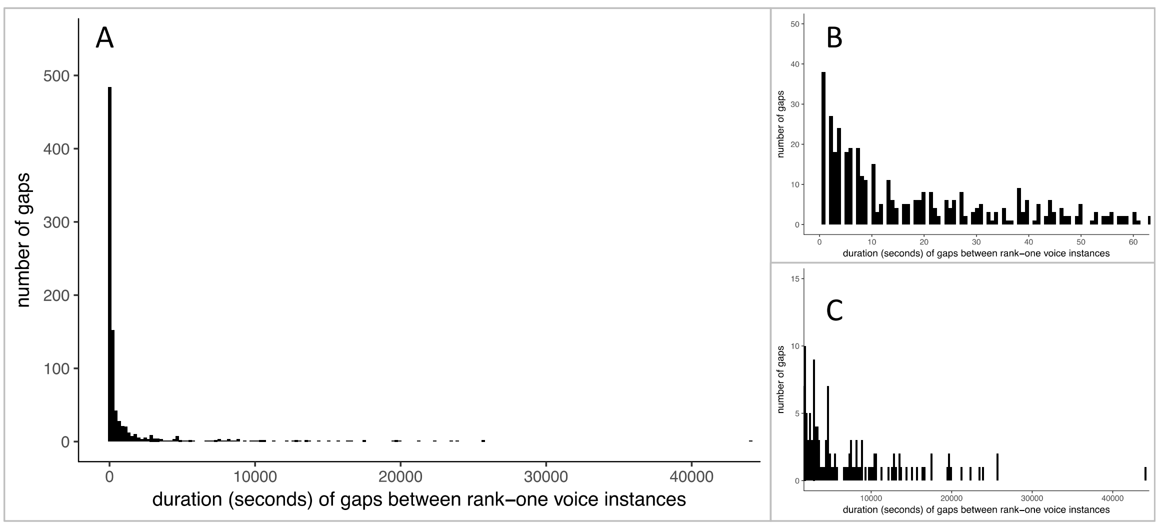

Figure 4.19. Many gaps between instances of infants' top voice were short in duration, and many were considerably longer. Of (A) all of the gaps between instances of infants' rank-one voice across all infants' days, (B) .44 were shorter than 60 seconds in duration and (C) .08 were 3,600 seconds or more in duration.

We also compared the corpus of infant's rank-one voice data to the data in the randomly shuffled streams, using the same randomly shuffled streams that we generated earlier. Because we preserved the duration of each music bout and the links between duration and voice type when we generated the shuffled streams, the rank-one voice types in the infants' real data were also the voice types that occurred for the longest cumulative duration in the shuffled streams. We determined the gap durations between the instances of these shuffled rank-one voice types. As we did for gap durations between music bouts, we calculated the median across iterations for the gap durations between each rank-one voice instance in the shuffled corpus. We then compared the gap durations in the infants' real data to the median durations of the shuffled data. The KS test results confirmed that the infants' real rank-one voice gap-duration data differed significantly from that in the shuffled streams ($D = .61$, $p < .001$). Table 4.5 presents the results (i.e., medians and standard deviations across infants) for the proportion of gaps between rank-one voice instances in infants' real data and for the shuffled streams. In general, infants' real data consisted of a higher proportion of short-duration gaps between rank-one voice instances relative to data in the shuffled streams.

Table 4.5.
A higher proportion of rank-one voice gap durations were short (i.e., seconds level) in infants' real data than in the shuffled data.

TOP VOICES	**REAL DATA**		**SHUFFLED DATA**	
	Median	SD	Median	SD
seconds	0.41	0.21	0.00	0.04
minutes	0.47	0.21	0.72	0.23
hours	0.09	0.11	0.24	0.24

We discovered a comparable pattern when we examined the gap durations between instances of the rank-one voice in each infant's data – within each infants' day, some gaps between instances of the infant's top voice were short in duration, while others were much longer (i.e., on the order of hours).

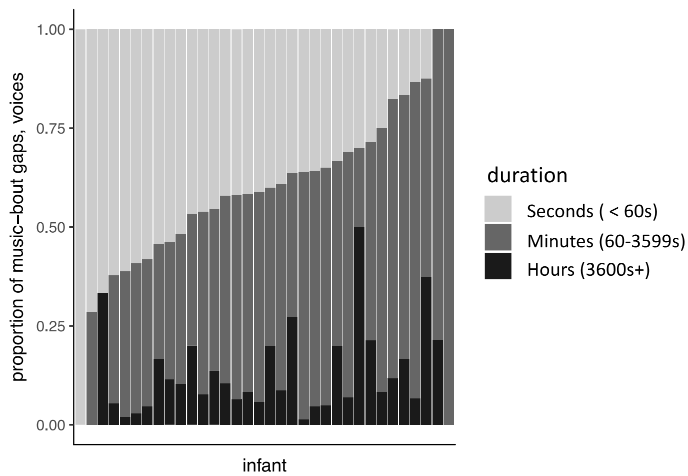

Figure 4.20. Individual infants (vertical bars) each encountered rank-one voice gaps with a mix of durations, ranging from seconds (light gray) to minutes (dark gray) to hours (black). For some infants, rank-one voice gap durations were mostly short (i.e., on the order of seconds), and for other infants, they were mostly somewhat longer (i.e., on the order of minutes). Almost all infants encountered a small proportion of considerably longer rank-one-voice gap durations (i.e., on the order of hours).

How often did the top voice occur paired in sequential order with other voices?

Next, we examined whether infants encountered repetitions of their top voice as well as what occurred during the long gaps between instances of their top voice. We determined the proportion of voice instance pairs (i.e., consecutive voice instances) that were *rank-one repetitions* (i.e., an instance of the rank-one voice followed by another instance of the rank-one voice), rank-one anchors (i.e., an instance of the rank-one voice followed by an instance of a different voice type), or *non-rank-one pairs* (i.e., two consecutive voice instances that were not of the rank-one voice type). Figure 4.21 shows the proportion of *rank-one repetitions* (magenta) and of *rank-one anchors* (black) that occurred in the voice instance pairs of each infants' day (the remaining proportion represents the proportion of non-rank-one pairs). We found that proportion of voice instance pairs that involved the rank-one voice type in any way varied widely across infants, ranging from .01 to 1 (Median = .49, SD = .74) in infants' everyday music. Across infants, the rank-one voice was more likely to occur followed by a repeated instance of itself (Median = .40, SD = .25, range: 0-1), than to occur followed by an instance of a different voice type (Median = .07, SD = .04, range: 0-.16).

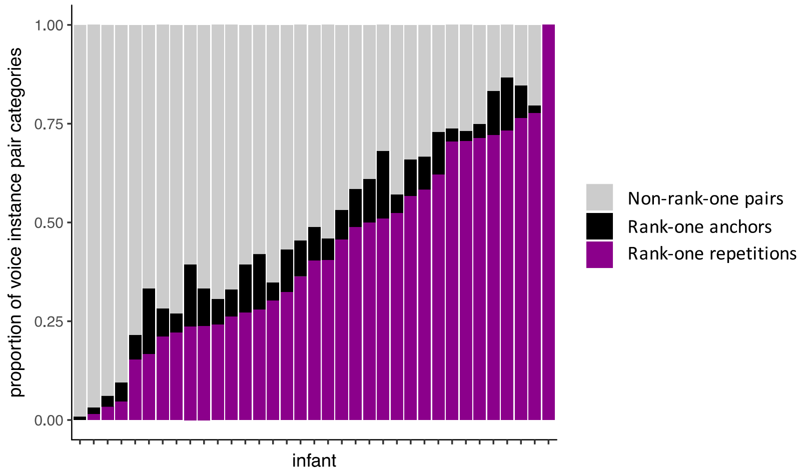

Figure 4.21. Infants mostly encountered their rank-one voice paired in sequential order with itself (*rank-one repetition*, magenta). Infants did encounter some *rank-one anchors*, where their rank-one voice was paired in order with a different unique voice (black) as well as many voice-instance pairs in which their rank-one voice was not involved (*non-rank-one pairs*, gray).

Figure 4.22 shows how infants' rank-one voice instances occurred in sequential order across each infant's day. Each row is one infant, and each dot is one voice instance. The x-axis is the voice instance number, meaning that instances appear in the sequential order they occurred during infants' days (divorced from the actual time of day). Rank-one voice instances are shown in magenta, and instances of all other voice types are shown in black. Any instance of the rank-one voice that followed another instance of the rank-one voice (i.e., two consecutive magenta dots) depicts a *rank-one repetition.* Any instance of a different unique voice that followed an instance of the rank-one voice (i.e., a black dot following a magenta dot) shows a *rank-one anchor*. *Non-rank-one pairs* are depicted by consecutive black dots. Baby 34 is the second row down from the top. The first (left-most) dot depicts the first voice instance that occurred in Baby 34's day. The next dot depicts the second voice instance in Baby 34's day. Both of these dots are magenta, which means this infant encountered 2 instances of their rank-one voice paired in sequential order at the beginning of their day. The second instance was thus a *rank-one repetition*. The 3rd voice instance in Baby 34's day was a different voice type than their rank-one voice, making it a *rank-one anchor*. Across all infants (all rows), it was evident that the rank-one voice occurred frequently paired in sequential order with itself across infants' days.

In sum, as infants progressed through their days, whenever they encountered their top voice, almost half of the time, the next voice they were likely to encounter was a repeated instance of that same voice – their top voice. This local repetition meant that infants had many opportunities to build up knowledge about one specific voice. Infants had relatively fewer opportunities to compare and contrast (sequentially) their top voice with other voices that occurred in their days.

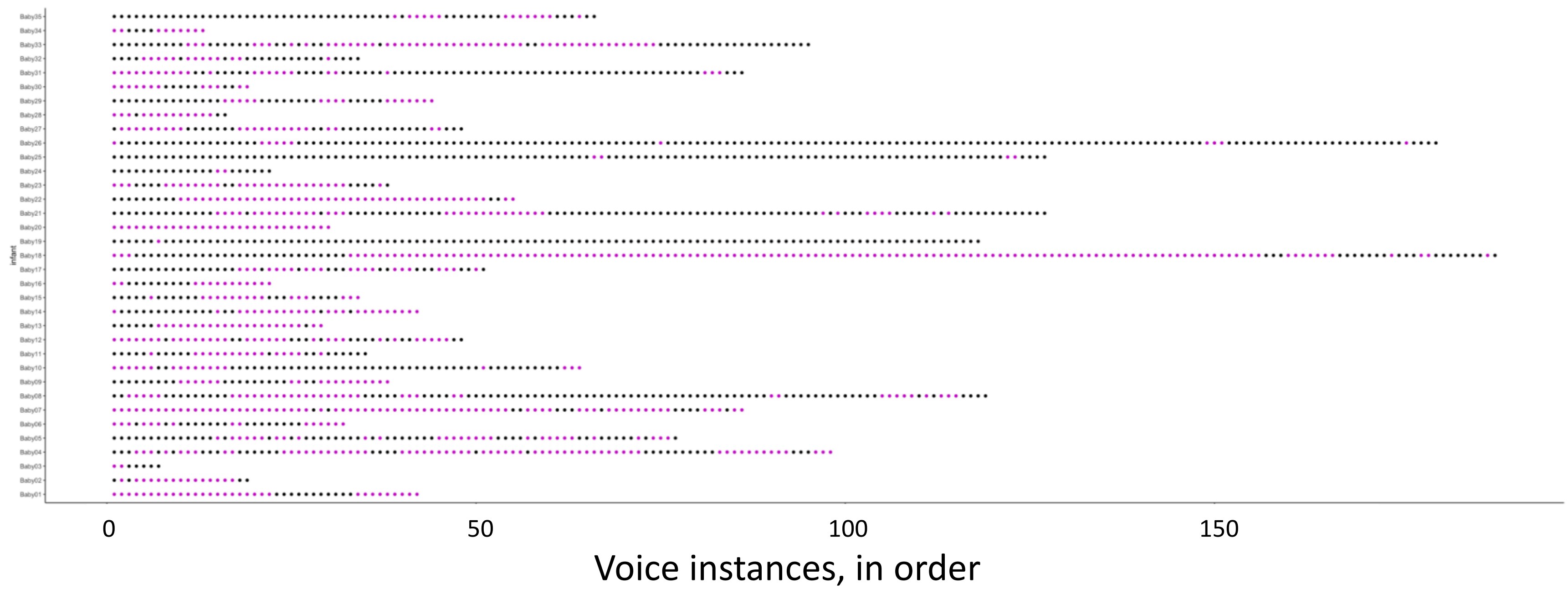

Figure 3.22. By plotting voice instances in order in each infants' day, with rank-one voice instances highlighted (magentaa dots), it is clear that rank-one repetitions (two sequential magenta dots) and rank-one anchors (magenta dot followed by a black dot) were both relatively rare.

As we did for infants' tune data, we also compared rank-one voice pairs in infants' real voice data to that of the randomly shuffled data streams. We calculated the proportion of each voice instance pair type (i.e., *rank-one repetitions*, *rank-one anchors*, and *non-rank-one pairs*), considering the same rank-one voice type in infant's real data to be the shuffled rank-one voice type in the matched shuffled data stream. We repeated this process 100 times, calculating the median proportions of each pair type across all iterations for each infant. Table 4.6 shows how the infants' real data compared to the shuffled streams for each type of voice instance pair. In general, it appeared that infants encountered a greater proportion of rank-one repetitions in their real data compared to in the shuffled data.

Table 4.6.
Proportion of rank-one pair types that occurred in infants' real voices data and in the shuffled streams.

VOICES	INFANTS' REAL DATA		SHUFFLED DATA	
	Median	SD	Median	SD
proportion rank-one repetitions	0.40	0.25	0.12	0.27
proportion rank-one anchors	0.07	0.04	0.18	0.08
proportion non-rank-one pairs	0.51	0.26	0.62	0.28

How often did voice types switch over the course of infants' days?

Finally, we determined how often voice types changed over the course of infants' days. As with tunes, for this analysis, we did not take into account the time of day or the duration between voice instances. Rather, we calculated how often the voice type changed from one voice instance to the next. Each infant encountered a different cumulative number of voice instances over the course of their days, so we calculated the *voice switch rate* as a proportion: the cumulative number of voice instances that were a change from the immediately prior voice instance divided by the cumulative number of all voice instances in the infant's day. The *voice switch rate* ranged from 0 to .83 (Median

= .32, SD = .18; see Figure 4.23). On average, this was a relatively low switch rate, meaning that infants did not often encounter a change in voice type from one voice instance to the next. In fact, in nearly two-thirds of their voice instances, infants encountered a repetition of the same voice type from one voice instance to the next. Thus, if an infant encountered an instance of *Mom* then the next time a voice occurred, they would be more likely to encounter another instance of *Mom* than an instance of *Sister* or *Daniel Tiger*. While this overall pattern characterized the voice data for most infants, a few infants (.14) encountered the reverse – they were more likely to encounter a change in voice type from one voice instance to the next than to encounter the same voice type across consecutive voice instances.

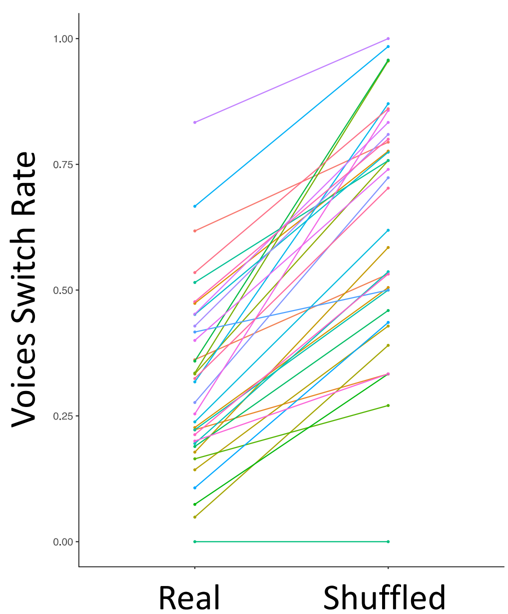

Figure 4.23. Compared to the shuffled streams, most infants encountered a lower switch rate of voices, indicating that they were more likely to encounter a repetition of the same unique voice from one instance to the next than a change in voice type.

We also compared switch rate of infants' real voices data to that of randomly shuffled data streams. As we did for tunes, we began with each infant's actual data – the list of voice types per voice instance ordered as they occurred in time across each infant's day. We then sampled randomly from that list, without replacement, to generate a new order of the same content. Then, we calculated the voice switch rate based on this shuffled sequential order of voice instances and their associated voice types. We repeated this process 100 times, calculating the median *shuffled voice switch rate* across all iterations for each infant. Across infants, the resulting *shuffled voice switch rate* ranged from 0 to .97 (Median = .68, SD = .23). A paired t-test ($t(34) = -13.05$, $p < .001$) revealed that infants' real voice data exhibited significantly less frequent switching compared to the randomly shuffled data.

This set of analyses yielded data about the temporal dynamics of infants' everyday vocal musical input that may shape how infants build up knowledge about vocal music and the voices that produce it in their everyday lives. In the voices music corpus, infants mostly encountered music within seconds of when they had last encountered music, and they also encountered many long lulls, on the order of hours, during which no music occurred. As infants advanced through their days, they encountered mostly repetitions of specific voices from one voice instance to the next. Repeated instances of infants' top voices were frequently paired together in time.

DISCUSSION

In this section of the dissertation, we examined the temporal dynamics of music and its contents (i.e., tunes and voices) as they occurred over the course of infants' recorded days. We answered the following questions: (1) Did instances of tunes music (and separately, voices music, top tunes, and top voices) occur in regular, random, or bursty temporal patterns over infants' days?, (2) How long were the gaps between individual instances of music (and separately, vocal music, top tunes, and top voices)?, (3) How often did tune types (and separately, voice types) switch from one instance to the next across infants' days?, and (4) How often did the top tune (and separately, the top voice) occur in sequential order paired with another instance of itself versus with an instance of a different tune (voice) type? We discovered that infants encountered many music bouts clustered close together in time (i.e., durations between music bouts on the order of seconds) in combination with long lulls (i.e., durations over an hour) during which no music bouts occurred. This bursty temporal pattern also characterized how music with exactly one voice, top tunes, and top voices occurred in time across infants' days. We also found greater repetition (i.e., less switching) of specific tunes and specific voices than would be expected by chance. In particular, repeated instances of infants' top tunes and top voices often occurred immediately adjacent in sequential order. These patterns of results were particularly evident in infants' real voices data, because compared to tunes, infants encountered fewer total unique voices and a lower proportion of unique voices that occurred in only one instance.

Integration demands in infants' everyday musical input

The set of results presented in this section of the thesis provide the first quantitative data on the nature of the integration challenges that infants face in their everyday music learning. Based on these findings, we now know that infants mostly encountered music and tunes separated by delays on the order of seconds (Median gap duration for music = 29 seconds; Median gap duration for top tunes = 5 seconds), and they mostly encountered vocal music and voices separated by delays on the order of minutes (Median gap duration for vocal music = 1.0 minutes; Median gap duration for top voices = 1.42 minutes). Evidence across many domains points to temporal proximity as one clear mechanism by which learners build knowledge (for relevant review, see Goldstein, et al., 2010). Encountering to-be-learned content close in time enables learners to integrate that content without information having to survive in their memory over extended durations. We have discovered that in the context of infants' everyday music, "proximity in time" means roughly within 30 seconds. Although a considerable amount of laboratory-based research has been conducted on the effects of spacing, little is known about how encountering content at temporal intervals varying from about 10 seconds to 90 seconds might differentially impact learning and memory over the course of a day. The bulk of extant laboratory-based research has presented information at temporal intervals on the order of seconds and then tested learning and memory of that information after delays of only seconds or minutes, not hours over the course of a day (for reviews, see Cepeda, Pahler, Vul, Wixted, & Rohrer, 2006; Rovee-Collier, 1999). Thus, there is a clear gap in existing research on the temporal dynamics that occur within and across one full day and the impact of that temporal structure on subsequent learning.

While infants mostly encountered music and its contents at relatively short temporal intervals, they sometimes also encountered music and its contents separated by considerably longer delays, on the order of hours. It is possible that these longer temporal intervals introduced 'desirable difficulty' (e.g., Bjork & Kroll, 2015) – this could facilitate infants' learning by forcing infants to engage cognitive systems for encoding and recalling information across longer time periods. In this way, it is possible that the bursty temporal patterns of infants' everyday musical input optimally support infants' musical learning. A clear next step would be to directly test this possibility, manipulating the temporal dynamics at which infants encounter musical content, and then testing their subsequent learning of that content.

Unresolved plausibility of the perceptual anchor hypothesis.

We discovered that infants encountered their top tune paired in sequential order with other tunes (*rank-one anchors*) in a relatively small proportion of tune-instance transitions (range: .003-.13). As infants progressed through their days, they only occasionally encountered an instance of their top tune followed by an instance of a less frequently occurring tune type. Discovering this low proportion of *rank-one anchors* potentially rules out the perceptual anchor hypothesis as a likely explanation for how infants build up knowledge about tunes in the music they encounter in their everyday lives. For voices, the proportion of voice-instance transitions that involved a top voice paired in order with a different unique voice varied across infants (range: 0 - .17). For 12 infants (.34), rank-one anchors occurred in 10% or more of their rank-one pairs, which may be a non-negligible amount. For these infants, then, it would be more plausible that their top voice may have served as a perceptual anchor to guide their processing of less frequently occurring voices in the music of their everyday lives. Additional research is needed to directly test the perceptual anchor hypothesis for both tunes and voices. It is possible that the proportion of *rank-one anchor pairs* must reach some particular threshold in order to serve as an effective perceptual anchor. It is also possible that there just needs to be a non-zero occurrence of a top tune (voice) paired in order with a different unique tune (voice) for the top tune (voice) to serve as an effective perceptual anchor. Another possibility is that the anchoring effect does not have to be limited to tunes (voices) that occur consecutively in order (or close together in time, for that matter). It is unknown to what extent one instance of a top tune (voice) might impact infants' processing of subsequent tunes (voices) that occur seconds, minutes, and hours later in their day. It is also unknown how individual differences in the proportion of rank-one anchors might matter for infants' learning about tunes and voices that occurred in the music of their everyday auditory environments. A productive next step in this line of inquiry would be to present infants with a set of tunes, directly manipulate how often infants encounter a top tune (voice) paired in sequential order with a less-frequent tune (voice), and then measure how the manipulation impacted infants' learning of the less-frequent tune (voice).

Outstanding questions and future directions.

As the first study to examine how musical input occurred across time in the day in the everyday lives of multiple infants, these results raise several key questions for future research. We highlight two outstanding questions that we see as important next steps in this line of research.

What do different burstiness values mean? Burstiness has been widely used in the field of physics (for recent review, see Karsai, Jo, & Kaski, 2018), but it is a relatively new measure in the domain of developmental psychology. Therefore, limited data are available to guide our interpretation of the burstiness results in the context of how everyday input relates to children's learning and memory. One possibility is that only the categorical distinction is meaningful – if an infant's $A_n(r)$ value fell above the random cutoff (as determined by the upper CI based on the simulated data), then the temporal structure for the music in that infant's day was bursty, and the particular value of $A_n(r)$ would not provide any further information about the structure of music in that infant's day. In this scenario, the burstiness parameter would not provide a useful to differentiate the temporal patterns of music and its contents across infants. For all event types (i.e., music, vocal music, top tunes, and top voices) all infants $A_n(r)$ values were clearly in the bursty category (with the exception of two infants who had burstiness values that fell in the regular category for their rank-one tune). Another possibility is that the specific values of $A_n(r)$ within the range the bursty category signify meaningful differences in temporal structure (i.e., less bursty to more bursty). In this scenario, then, we could further explore differences across infants – for example, what does $A_n(r) = .12$ versus $A_n(r) = .94$ mean about how music occurred in time in two infants' days? Further, do lower versus higher burstiness values differently impact infants' subsequent learning in the domain of music? Do other factors about the infants' days and/or about their musical encounters systematically relate to their level of burstiness? We do not yet know the answers to these questions, but the questions themselves clearly motivate the next steps to take in future research.

How we interpret these burstiness results might also depend on the underlying processes that generate the bursty temporal patterns. By some accounts, burstiness of human behaviors relates to rhythms of human activities and demands of prioritizing among various tasks (e.g., Barabási, 2005). By other accounts, bursty temporal patterns are related to circadian and/or weekly rhythms (for relevant review, see Karsai, Jo, & Kaski, 2018). Based on existing survey data, music is often connected to infants' daily activities, such as playing, bathing, and preparing for sleep (Johnson-Green & Custodero, 2002; Trehub, Unyk, et al., 1997; Valerio, Reynolds, Morgan, & McNair, 2012). In addition, the small set of activities that infants engage in repeat and change frequently over the course of their days (Fausey, Jayaraman, & Smith, 2015). It is plausible that bursty temporal patterns in infants' everyday musical input could arise as a result of infants' daily activity rhythms. For example, infants might encounter multiple music events close together in time as their caregivers attempt to put them to sleep for a nap. Then, this would then be followed by a long period during which no music occurred (e.g., presumably the duration of the nap plus however long until music occurred again). This poses another fruitful avenue for future research.

What is the nature of rank-one repetitions? The proportion of rank-one repetitions for both tunes and voices was higher than would be expected randomly, indicating that infants encountered considerable local repetition of their top tunes and of their top voices. In other words, not only were infants' top tunes (voices) the most available tune (voice) type that occurred in infants' days, but they also often occurred closer together in time. What was the nature of these repeated instances? Did the top tune start from the beginning of the tune in each separate instance? Or did the top tune continue to advance, split across multiple instances because it was interleaved with other noises in the infants' environments? Based on our current coding scheme, both of these scenarios would appear as repeated instances of the top tune. These situations would actually present different learning opportunities for infants. In the former situation, infants would encounter their top tune in a manner akin to a blocked presentation of one item, in which each instance of the stimulus is exactly the same as the previous instance. This would provide opportunities for infants to learn about the details of what was the same across the repeated instances (e.g., Carvalho & Goldstone, 2013, 2017) – the pitches, the rhythms, the tempos, and so forth. In the latter scenario, infants would encounter instances of their top tune that were a little bit different each time. This presents a different integration challenge for infants, making it harder to know which sounds should be aggregated as part of the same tune and which should be separated as belonging to two different tunes. In some ways, this latter scenario could be comparable to encountering different exemplars of the same category – infants must discover and integrate the set of sounds are associated with the tune *Twinkle Twinkle Little Star* and separate those from the sets of sounds that are associate with other tunes, like *If You're Happy and You Know It*. Rank-one repetitions likely occurred in some combination of these two possibilities. Encountering some exact repetitions of a top tune as well as some more variable 'repetitions' might actually facilitate infants' learning to recognize the full length of their top tune and to differentiate it from other tunes. Future research could code acoustic properties of the raw audio signal, such as absolute pitch and tempo, to determine the more detailed nature of rank-one repetitions.

Likewise, for voices, in our current coding scheme a voice would be given the same label regardless of the musical content it was producing. We have not yet examined how similar or different any one instance of an infant's top voice was from any other instance. For many infants (.46), the rank-one voice was *Mom*. How often did *Mom* sing the same tune? If *Mom* sang the same tune more than once, then how consistent was one rendition to the next? Laboratory-based research has demonstrated that mothers' singing the same tune to their infants across two separate recording sessions was remarkably stable in absolute pitch and tempo (Bergeson & Trehub, 2002). In future research, we could examine the extent to which top voices produce stable pitches and tempos in infants' everyday musical input.

In conclusion, we have described multiple aspects of how infants' everyday musical input occurred in time over the course of their recorded days. Our findings provide initial landmarks that should move theory forward and guide future research about how infants' everyday musical input shapes their learning in the domain of music.

CHAPTER V

GENERAL DISCUSSION

In this dissertation, we addressed three broad questions about the nature of music available to infants in their everyday lives: (1) In one full day, how much music did infants encounter and how often did they encounter it?, (2) In what ways did infants encounter consistency, diversity, and social quality in the tunes and voices that occurred in their everyday musical input?, and (3) How were individual instances of music and its contents distributed in time across infants' days?

We identified music in day-long auditory recordings from infants and caregivers at home in their natural environments. Thus, we successfully created a corpus of infant-available, everyday musical sounds. We analyzed the distributional and temporal properties of infants' everyday music and its contents (i.e., the tunes and voices that occurred in infants' everyday music). We discovered that infants in this sample encountered roughly one hour of music distributed across multiple separate instances throughout their days. Infants also encountered some instances of live music produced by caregivers, siblings, and other adults in their environments. Within their daily music, infants encountered multiple unique tunes and multiple unique voices, some of which occurred for much longer cumulative durations than others. As infants progressed through their days, they encountered many music bouts close together in time as well as some music bouts separated by much longer lulls. This bursty temporal pattern also characterized how infants encountered instances of their top tune and their top voice – the specific tune and specific voice that occurred for the longest cumulative duration in each infant’s day. Finally, infants encountered many pairs of consecutive music bouts with repeated content – the same tune or the same voice. Taken together, we discovered that infants' everyday musical input was biased toward repeated tunes and voices occurring close together in time.

A first-of-its-kind corpus of infant-available music.

In the current research, we successfully collected a corpus of infant-available musical input identified in one day-long auditory recording from each of 35 infants and their caregivers at home in their natural environments. Across the whole corpus, caregivers recorded a total about 467 hours, in which we identified about 42 hours of music. The corpus we have collected is the first of its kind in that it contains infant-available music identified in day-long natural auditory data from a sizable number of families. Our research was an important advance compared to prior research that has opted to collect either in-depth data for one or two infants or to characterize music based on survey data.

Survey studies have recruited large numbers of participants (i.e., hundreds of participants) to complete questionnaires and/or interviews about their infants’ everyday musical experiences (e.g., Custodero & Johnson-Green, 2003; Custodero, Britto, & Brooks-Gunn, 2003; Ilari, 2005; Rideout, 2013; Rideout, Vandewater, & Wartella, 2003) – the main advantage of these studies is that they have polled such a sizable sample that their results were likely to be representative of the general population. The main limitation of these large-scale survey studies is that they provided only coarse estimates of some aspects of infants’ everyday musical input. For example, this research has primarily asked caregivers to report the frequency of music that infants encounter in the past week (e.g., Custodero, Britto, & Brooks-Gunn, 2003) or in the past day (e.g., Rideout, 2013). In these studies, it was not possible to collect in-depth information about every participating family's everyday musical activities, such as the how often within a day infants heard music, which particular songs were available, who sang songs, and what specific pitches and rhythmic patterns were in the music that occurred. One recent case study has explicitly demonstrated the limitation of relying on caregiver report (Costa-Giomi, 2016) – the researchers recorded one day in the life of a family with twin infants. The caregiver told the researchers that they sang "all the time" with their infants, but when the researchers analyzed the musical input that had occurred during the recorded day, the caregiver had only sung to one infant for 35 seconds. Thus, survey and interview studies have provided limited insight into the nature of infants' everyday musical experiences. On the other hand, targeted case studies have recorded infants' auditory or audio-visual environments and analyzed in detail the quantity and quality of the available sensory input (Addessi, 2009; Bergeson & Trehub, 1999, 2002; Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017; Eckerdal & Merker; 2009; Koops, 2014). While these studies have provided rich information about the participating infants' musical experiences, this research has been limited in other ways. For example, researchers have studied only one family (Costa-Giomi, 2016; Costa-Giomi & Benetti, 2017) or a small number of families (i.e., under 10 families; Addessi, 2009; Koops, 2014), or they have collected data during restricted, particularly musical, activities of infants' days (Addessi, 2009; Bergeson & Trehub, 1999, 2002; Eckerdal & Merker; 2009; Koops, 2014).

In the present research, we combined these two approaches to achieve a balance among these competing goals. By collecting data from 35 infants and their families who were intentionally not recruited due to any particular involvement in music, we have increased the generalizability of our results relative to studying one or a small number of infants. By harnessing recent innovations in wearable recording technology to collect day-long auditory recordings, we have captured and characterized the natural complexity of music available in infants' everyday environments.

In addition, our research followed in line with recent work that documented how studying infants' natural, complex, messy environments and behaviors yielded strikingly different results than studying infants in brief, highly structured laboratory-based experiments (Lee, Cole, Golenia, & Adolph, 2017; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017; see also Smith, Jayaraman, Clerkin, & Yu, 2018). For example, during 5 minutes of structured play in which infants and caregivers were instructed to play with a set of experimenter-provided toys, caregivers provided dense linguistic input to their infants. In contrast, during 45 minutes of their unstructured, natural routines, caregivers exhibited greater moment-to-moment variability in the amount of linguistic input they generated. Notably the natural context contained a much higher proportion of silence than did the structured-play context.

In the current study, the corpus of music available in infants' natural environments that we have created has enabled us to begin answering questions that were previously beyond the scope of possibilities. It will continue to be a rich source of information about the nature of music in infants' everyday auditory environments that will fuel new lines of research.

What is the pathway from frequent to familiar?

One potentially fruitful avenue to pursue in future research is how infants' everyday musical input shapes their song recognition. Prior research points to the impact of encountered input in shaping infants' song recognition (e.g., Costa-Giomi & Davila, 2014; Ilari & Polka, 2006; Plantinga & Trainor, 2005; Saffran, Loman, & Robertson, 2000; Volkova, Trehub, & Schellenberg, 2006; Weiss, Trehub, & Schellenberg, 2012; Weiss, Schellenberg, Trehub, & Dawber, 2015), but many methodological differences across these laboratory-based studies (i.e., amount of exposure, duration of delay between exposures, number of test trials, etc.) have made it difficult to determine precisely which circumstances promote infants learning and memory for tunes. In the present research, we detailed several properties of how tunes occurred in infants' everyday musical input that may shape infants' subsequent learning.

We discovered that infants encountered some unique tunes and some unique voices for longer cumulative durations than the other unique tunes and other unique voices that occurred in the music of their everyday lives. The one specific tune and one specific voice that occurred for the longest duration in each infant's day was their "top tune" and "top voice". Most infants encountered multiple instances of their top tunes and top voices which occurred in bursty temporal patterns across infants' days. Moreover, from one music bout to the next, infants often encountered repeated instances of their top tune and their top voice. The present findings elucidate to what extent the stimuli and study designs used in prior research differ from infants' everyday musical input. In some cases, laboratory-based research has presented infants with multiple repetitions of one tune, and then immediately tested their processing of that tune versus another tune (e.g., Plantinga & Trainor, 2009; Trehub, Bull, & Thorpe, 1984). This design would be well matched to what we have discovered about infants' everyday musical input, as infants encountered only a small number of tunes in any 5-minute period (i.e., the typical timescale of these laboratory-based studies). In other cases, prior research has sent CDs of one or two tunes home with families to play once per day for 1-2 weeks (e.g., Ilari & Polka, 2006; Plantinga & Trainor, 2005; Saffran, Loman, & Robertson, 2000; Trainor, Wu, Tsang, 2004). This design would be less well matched to what we have found about infants' everyday musical input, as most infants encountered more than two unique tunes in a full day. In this way, the findings we have presented in the current research could serve as landmarks to guide future research. As researchers design future studies about infants' musical processing, they could take into consideration the extent to which their design matches (or does not match) the structure of what infants actually encounter in their everyday musical input.

The properties we discovered about how infants encounter tunes in their everyday musical input musical input vary across infants, and so it will be possible to empirically test hypotheses about how variation on these dimensions shape learning. For example, infants varied in their *rank-one consistency bias* – the degree of to which one tune occurred for a longer duration compared to other tunes they encountered in their everyday musical input. Further, some infants encountered their top tune in a single, continuous instance, while other infants encountered their top tune in multiple separate instances. Within the latter group of infants, some infants encountered all instances of their top tune close together in time with only brief delays (i.e., seconds) between each instance, while other infants encountered more variability in how instances of their top tune occurred in time, such that some instances occurred close together and other instances were spaced apart by minutes or hours. Differences in these properties of infants' top-tune encounters would relate to differences in infants' subsequent learning and memory in the domain of music, since variability shapes learning (e.g., Carvalho & Goldstone, 2013, 2017).

Recent laboratory-based research has demonstrated that familiar songs shape infants' behavior in social contexts (Cirelli & Trehub, 2018; Mehr, Song, & Spelke, 2016; Mehr & Spelke, 2017). For example, infants were familiarized with novel lullaby by having their caregivers sing it at home or encountering it being played from a toy. In both cases, caregivers estimated the number of times infants encountered the song. In a subsequent laboratory-based session, infants were shown two videos, one of a novel person singing the same lullaby and one of a novel person singing an unfamiliar tune. Infants who had encountered the lullaby being sung by their caregivers attended longer to a novel person singing this lullaby than to a novel person singing an unfamiliar song, while infant who had encountered the lullaby from a toy did not show this attentional preference (Mehr, Song, & Spelke, 2016). This study provided one example of how different properties of encountered input – in this case live versus recorded renditions of a tune – impact infants' subsequent learning and behavior. How does this operate in the context of infants' real-world music learning? For example, do infants who encounter more of their daily music as live-vocal tunes also show accelerated song generalization?

How do tunes that occur in infants' everyday environments become familiar? We propose that encountering one specific tune that cumulates to a longer duration than other tunes, that repeats across multiple, consecutive instances, and that occurs in a bursty temporal pattern across time in everyday musical input may be one route by which infants acquire a familiar song. One indication that a tune has become familiar to an infant would be for the infant to recognize the tune across variability. If an infant encountered their top tune in a different musical key, presented at a different tempo, played with some rhythmic variation (e.g., in a jazzy style), and/or sung by an unfamiliar voice, could they recognize it as their top tune and be able to differentiate it from other tunes? According to the above hypothesis, we predict that infants who encountered greater consistency, coherence, and clustering of their top-tune encounters would demonstrate superior ability to recognize their top tune across variability than infants whose top tunes occurred in a less consistent, coherent, and clustered manner. Our research revealed that infants encountered consistency in more than one dimension of their everyday musical experience, and we discovered preliminary evidence that the degree of consistency tracked across different dimensions. For example, the rank-one consistency bias for tunes was positively correlated with the proportion of tune-instance pairs that were rank-one repetitions ($r_s = .75$, $p < .001$; see Figure 5.1).

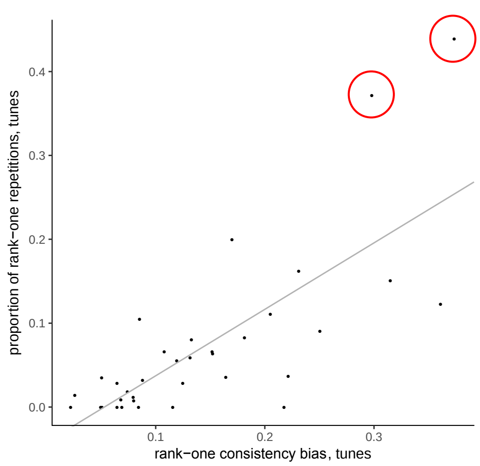

Figure 5.1. Infants encountered consistency in their musical input in multiple dimensions that tracked together, such as the degree to which one specific tune was more available (*rank-one consistency bias*) and how often infants encountered repeated instances of that specific tune across consecutive music bouts (*rank-one repetitions*).

In other words, the degree to which one specific tune was disproportionately available in the infants' everyday musical input was associated with how often infants encountered repetition of that specific tune from one music bout to the next. Repetition and consistency are likely to be particularly helpful for young learners, who in the early stages of building up knowledge in the domain of music (e.g., Roy, Frank, DeCamp, Miller, & Roy, 2015; Vlach & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008). This natural pattern leads to specific, testable hypotheses such as: Do infants who encounter their top tune with high degrees of repetition and consistency (e.g., circled in red, Figure 5.1) start singing this tune earlier that infants who encounter their top tune with more variability (as for first words in Roy et al., 2015)?

A multi-dimensional view of music in infants' everyday environments.

At first it may seem like a simple task to design a study to manipulate properties of the available input and then measure the effects of that manipulation on infants' learning and memory. But which property of the input should be selected to manipulate? And is it possible to manipulate one property without impacting others? The multidimensional structure of infants' everyday musical input has potential to pose a considerable challenge to future researchers who aim to test how the structure of the input shapes infants' learning in the domain of music. To overcome this challenge, it will be vital to gain a more complete understanding of how the various properties of infants' everyday music relate to one another. It will also be necessary to develop innovate methodologies that monitor infants as they encounter sensory input that reflects these real-world properties (for one example of an innovative approach to studying infants' motor development, see Lee, Cole, Golenia, & Adolph, 2017).

Part of the wealth in collecting data from infants in their natural, complex, messy environments is that we have captured multiple dimensions of infants' everyday musical input, and we can characterize how these dimensions were interrelated. As a first step, we examined pair-wise correlations among all of the dependent measures reported in this dissertation. We discovered sensible relationships among many of the variables related to infants' everyday musical input (see Supplemental Materials for additional details). To give one example, the number of music bouts was positively correlated with the duration of music in the voices music corpus ($r_s = .85$, $p < .001$; see Figure 5.2A). It did not have to be the case that a higher number of music bouts was associated with a longer duration of music. It could have been that infants encountered either lots of very short music bouts or a couple of very long music bouts – this scenario would have yielded a relationship in the opposite direction. Since the vast majority of music bouts that occurred were of similar, short durations, it was reasonable that infants who encountered a higher number of music bouts also encountered a longer duration of music. The more interesting (although still expected) findings were that the number of music bouts positively correlated with the number of unique voices ($r_s = .61$, $p < .01$; see Figure 5.2B), and in turn, the number of unique voices was positively correlated with the voices switch rate ($r_s = .67$, $p < .05$; see Figure 5.2C). These relationships suggested that encountering a longer cumulative duration of music and a larger number of music bouts provided infants with a greater number of opportunities for new unique voices to occur. In other words, infants who encountered more music also encountered greater diversity among the voices that occurred.

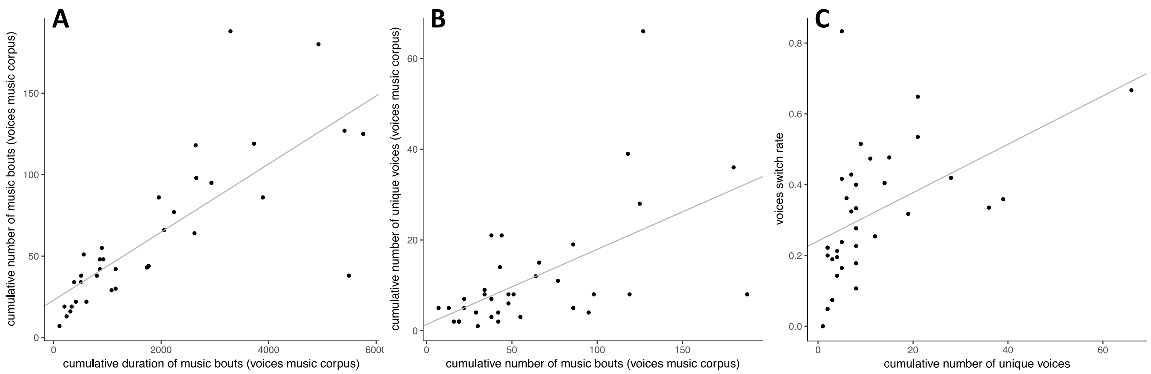

Figure 5.2. Infants who encountered a greater cumulative duration of music (in the voices music corpus) also encountered a greater number of music bouts (A). A greater number of music bouts was associated with a greater number of voices(B), which in turn was related to a higher switch rate of voices (C).

These findings raised new questions, such as how the relationships among these measures changed over the course of infants' days. As infants advanced through their days, the cumulative number of music bouts they encountered incremented across time, and with it, the cumulative duration of music increased. How did these changes relate to changes in other variables measured? For example, as infants encountered an incrementally larger number of music bouts over the course of their day, how did the switch rate of unique voices change? As we have defined it, the *voice switch rate* provided one value to characterize infants' entire days. Given that music bouts (and vocal music bouts) – and therefore tune (and voice) instances – did not occur at one regular rate throughout the day (i.e., music bouts occurred in a bursty temporal pattern), it is plausible that tune types and voice types may not have switched at one fixed rate throughout the day either. The full-day switch rates we reported were divorced from time of day. In future research, we could directly examine how the sequential orders of tunes and of voices order unfold in time across infants' days.

To give only an initial hint towards addressing this topic, we plotted the *voice switch rate* as it changed over the course of one infant's day (see Figure 5.3). In this plot, the x-axis depicts the seconds in a day. Each music bout (voices music corpus) that occurred in the infant's day is represented by a point in the plot. The height of the point (the y-axis) represents the proportion of cumulative instances up to that point in time that were switches. In this way, the denominator in the proportion changes from left to right across the x-axis. Values on the x-axis with no corresponding point indicate seconds during the day when no music bout occurred. If an infant encountered the same voice type across all of their music bouts, then every point in the resulting cumulative voice switch rate plot would occur at y = 0, because there would never be any voice switches. Conversely, if an infant encountered a different unique type for every music bout, then every point in the resulting cumulative voice switch rate plot would occur at y = 1 (i.e., 1 switch out of 1 transition = 1, 2 switches out of 2 transitions = 1, 3 switches out of 3 transitions = 1, etc.). In Figure 5.3, the infant's day began with 2 consecutive instances of the same voice (as indicated by the first 2 points occurring at y=0). After that, increasing proportion values indicated that as more music bouts occurred more voice switches occurred and decreasing proportion values indicated that as more music bouts occurred no additional voice switches occurred. Accordingly, this infant appears to have encountered mostly stretches of time during which the same voice type occurred repeatedly across music bouts as well as some stretches of time during which each music bout presented a change in voice type from the preceding music bout. Thus, the rate at which voice types changed as the infant encountered increasingly more music bouts was clearly not fixed over the course of this infant's day. This matters because it means that over the course of their day, this infant encountered a mix of some opportunities to build up knowledge about a repeated voice and other opportunities to compare and contrast different voices. We look forward to future research that characterizes multidimensional structure and how it changes as individual learners encounter new music over the course of their days.

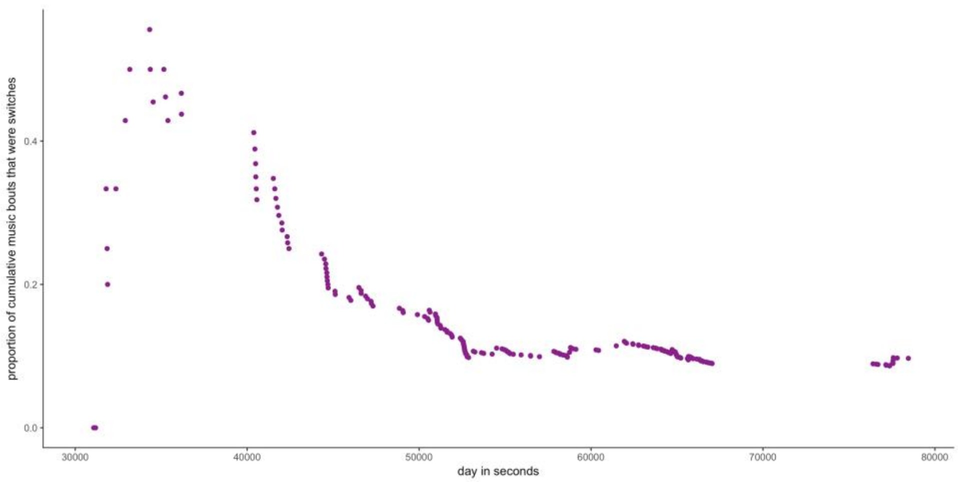

Figure 5.3. The rate at which voice types switched from one music bout to the next clearly changed across time in one infant's day.

Many paths from input to learning in the music domain.

One can rank order infants on each dimension of our newly available measurements of infants' everyday musical input. Which musical skill(s) does this variation predict? We recommend the following dimensions as starting points to empirically assess the relationships between input and learning, because they vary substantially across individual infants' experiences: *rank-one consistency bias*, the proportion of music produced by a live voice, and the switch rate of tunes and voices.

Musical input available to infants in their everyday environments might also shape their learning and behavior in domains beyond music. Based on prior research, logical domains to consider for potential targeted learning outcomes would include socioemotional development (e.g., Gerry, Unrau, & Trainor, 2012), executive function skills (e.g., Moreno, Bialystok, Barac, Schellenberg, Cepeda, & Chau, 2011), and language acquisition (e.g., Kraus, Hornickel, Strait, & Slater, 2014). Prominent scholars have proposed that music plays a substantial role in helping infants regulate their emotions (e.g., Corbeil, Trehub, & Peretz, 2015; for reviews, see Trehub, Hannon, & Schachner, 2010; Trainor & Trehub, 1998), and in promoting infants' attachment with their caregivers (e.g., Edwards, 2010). In one study, infants with formal musical training (i.e., participation in an early childhood education music class) not only displayed expedited acquisition of culture-specific musical knowledge but they also showed enhanced infants' socio-communicative development (Gerry, Unrau, & Trainor, 2012). Older children who were randomly assigned to participate in formal music training have shown enhanced IQ relative to children who were randomly assigned to participate in other activities, such as drama, science, or computer classes (e.g., Ireland, Parker, Foster, & Penhune, 2018; Schellenberg, 2004; Schellenberg, 2006; for reviews of research mainly with older children and adults see, Tierney & Kraus, 2013; Trainor & Hannon, 2013; Schellenberg & Weiss, 2013).

Prominent scholars have proposed that music experience may facilitate language learning (Patel, 2011). Recently, researchers have documented initial evidence that music experience may impact language learning. For example, older children who were enrolled in active music lessons exhibited faster and more robust brainstem responses to speech sounds compared to children enrolled in a music appreciation class (Kraus et al., 2014). Likewise, in a 2-year longitudinal study, children were randomly assigned to take either music or painting lessons. Children who participated in 12 months of music lessons showed enhanced neural activation for pre-attentive processing of speech sounds relative to children who were participated in the painting lessons (Chobert, François, Velay, & Besson, 2012). In infants, newborns exhibited neural responses to processing the artificial structure of speech syllables only when the stimuli were musically enhanced and not when the stimuli were presented with a flat contour, and this brain response was linked to the infants' later vocabulary size as assessed with a standard measure at age 18 months (François, Teixidó, Takerkart, Agut, Bosch, & Rodriguez-Fornells, 2017).

The findings we presented in this dissertation have potential to launch a new line of research in developmental psychology to examine how the nature of early music experience shapes infants' early learning and then to design evidence-based interventions to foster their learning and development. Multiple dimensions of content, timing, and social quality likely impact learning in multiple ways. Future research will use the discoveries presented here to test hypotheses about how a variety of 'musical diets' shape learning. We look forward to increasingly specific models about the role of early music environments in early learning.

Limits on generalization

Though our corpus of everyday musical input to young infants is a dramatic advance beyond prior efforts to characterize early musical experiences, we note several limitations that should be addressed in future work. First, this corpus includes music recorded inside families' homes (due to Oregon state law about audio recording in public). It is possible that music encountered in public spaces, as well as in transit between home and public spaces (e.g., the family's car), could change some aspects of input distributions reported here. We also note that the current sample of families is largely white, educated, and middle-to-upper socioeconomic status. We encourage future researchers to take full advantage of the ease in using LENA recorders to capture the musical experiences of families beyond this profile, including multiple languages, cultures, ethnicities, incomes, and education levels. Finally, we captured musical input available to infants ages 6-12 months. Future research should aim to capture the earliest, and even pre-natal, musical experiences given evidence that the human auditory system is tuned by experience from very early on (e.g., Partanen, Kujala, Tervaniemi, & Huotilainen, 2013).

Conclusion

In this dissertation, we collected and analyzed a first-of-its-kind corpus of music identified in day-long auditory recordings of 6- to 12-month-old infants and their caregivers at home in their natural environments. We discovered that infants encountered nearly an hour of cumulative music per day distributed across multiple instances. Infants encountered many different tunes and voices in their daily music. Some tunes and some voices occurred for considerably longer than others. Infants encountered many music instances close together in time that contained repeated content – the same unique tune or the same unique voice. Our research has potential to inform theory and future research examining how the nature of early music experience shapes infants' early learning.

REFERENCES CITED

Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and nonverbal communication. *Cognitive Science*, *143*(6), 2304. http://doi.org/10.1111/cogs.12612

Addessi, A. R. (2009). The musical dimension of daily routines with under-four children during diaper change, bedtime and free-play. *Early Child Development and Care, 179*(6), 747–768. http://doi.org/10.1080/03004430902944122

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE*, *4*(11) 1-7. http://doi.org/10.1371/journal.pone.0007678

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(02), 239–273. http://doi.org/10.1017/S030500091400049X

Aslin, R. N., & Newport, E. L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language Learning*, *64*(s2), 86–105. http://doi.org/10.1111/lang.12074

Baldwin, D. A., Markman, E., Bill, B., Desjardins, R. N., Irwin, J., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, *67*(6), 3135–3153. https://doi.org/10.1111/j.1467-8624.1996.tb01906.x

Baldwin, D., & Meyer, M. (2007). How inherently social is language? In E. Hoff & M. Shatz (Eds.) *Blackwell Handbook of Language Development.* Blackwell Publishing Ltd, Oxford, UK. http://doi.org/doi: 10.1002/9780470757833.ch5

Baldwin, D. A., & Moses, L. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development*, *10*(3), 309–329. https://doi.org/10.1111/1467-9507.00168

Barabási, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*(7039), 207-211. http://doi.org/10.1038/nature03459

Barrett, M. S. (2009). Sounding lives in and through music: A narrative inquiry of the ‘everyday' musical engagement of a young child. *Journal of Early Childhood Research*, *7*(2), 115–134. http://doi.org/10.1177/1476718X09102645

Belardi, K., Watson, L. R., Faldowski, R. A., Hazlett, H., Crais, E., Baranek, G. T., ... & Oller, D. K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with Fragile X Syndrome at 9–12 months of age. *Journal of Autism and Developmental Disorders*, *47*(4), 1193-1206. https://doi.org/10.1007/s10803-017-3033-4

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. http://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *PNAS*, *114*(49), 12916–12921. http://doi.org/10.1073/pnas.1712966114

Bergeson, T. R., & Trehub, S. E. (1999). Mothers' singing to infants and preschool children. *Infant Behavior and Development*, *22*(1), 51–64. https://doi.org/10.1016/S0163-6383(99)80005-8

Bergeson, T., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72–75. http://doi.org/10.1111/1467-9280.00413

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 397–401). New York: Academic Press.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, *128*(2), 241-252. http://doi.org/10.5406/amerjpsyc.128.2.0241

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. http://doi.org/10.1111/j.0956-7976.2005.01531.x

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. *Cognition and Instruction*, *33*(3), 226–256. http://doi.org/10.1080/07370008.2015.1067215

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513. http://dx.doi.org/10.1037/0003-066X.32.7.513

Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the Language ENvironment Analysis system (LENA™) in European French. *Behavior Research Methods*, *48*(3), 1109-1124. http://doi.org/10.3758/s13428-015-0634-8

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–11283. http://doi.org/10.1073/pnas.1309518110

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. http://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*, 1–12. http://doi.org/10.3389/fpsyg.2015.00505

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699-1719. http://dx.doi.org/10.1037/xlm0000406

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508. https://doi.org/10.1111/j.14677687.2005.00441.x

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, *128*(5), 910-916. http://doi.org/10.1542/peds.2011-0609

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380. http://doi.org/10.1037/0033-2909.132.3.354

Chang, H.-W., & Trehub, S. E. (1977a). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology*, *24*(2), 324–331. http://doi.org/10.1016/0022-0965(77)90010-8.

Chang, H.-W., & Trehub, S. E. (1977b). Infants' perception of temporal grouping in auditory patterns. *Child Development*, *48*(4) 1666–1670. http://doi.org/10.2307/1128532

Chang, L., de Barbaro, K., & Deák, G. (2017). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental Neuropsychology*, *41*(5-8), 342-361. https://doi.org/10.1080/87565641.2016.1274313

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. http://dx.doi.org/10.1037/0012-1649.38.6.967

Chobert, J., Francois, C., Velay, J. L., & Besson, M. (2012). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, *24*(4) 956-967. http://doi.org/10.1093/cercor/bhs377

Cirelli, L. K., & Trehub, S. E. (2018). Infants help singers of familiar songs. *Music & Science*, *1*, 1-11. http://doi.org/10.1177/2059204318761622

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. https://doi.org/10.1137/070710111

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711) 1471-2970. https://doi.org/10.1098/rstb.2016.0055

Cohen, A. J. (2008). Advancing interdisciplinary research in singing through a shared digital repository. *Journal of the Acoustical Society of America*. 123(5) 3380. https://doi.org/10.1121/1.2934017

Cole, W. G., Robinson, S. R., & Adolph, K. E. (2015). Bouts of steps: The organization of infant exploration. *Developmental Psychobiology*, *58*(3), 341–354. http://doi.org/10.1002/dev.21374

Corbeil, M., Trehub, S. E., & Peretz, I. (2015). Singing delays the onset of infant distress. *Infancy*, 1–19. http://doi.org/10.1111/infa.12114

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142–158. http://doi.org/doi:10.1111/desc.12100

Costa-Giomi, E. (2016). Infant home soundscapes: A case study of 11-month-old twins. *International Perspectives on Research in Music Education*, 70-78.

Costa-Giomi, E., & Benetti, L. (2017). Through a baby's ears: Musical interactions in a family community. *International Journal of Community Music*, *10*(3), 289–303. http://doi.org/10.1386/ijcm.10.3.289_1

Costa-Giomi, E., & Davila, Y. (2014). Infants' discrimination of female singing voices. *International Journal of Music Education*, *32*(3), 324-332. https://doi.org/10.1177/0255761413491211

Costa-Giomi, E., & Sun, X. (2016). Infants' home soundscape: A day in the life of a family. In J. Bugos (ed.). *Contemporary Research in Music Learning Across the Lifespan* (pp. 99-108). New York: Routledge.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, *56*(5), 815-822. http://doi.org/10.1037/0022-3514.56.5.815

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. Journal of Verbal Learning & Verbal Behavior, 21, 451–467.

Custodero, L. A., & Johnson-Green, E. A. (2003). Passing the cultural torch: Musical experience and musical parenting of infants. *Journal of Research in Music Education, 51(2),* 102–114. http://dx.doi.org/10.2307/3345844

Custodero, L. A., Britto, P. R., & Brooks-Gunn, J. (2003). Musical lives: A collective portrait of American parents and their young children. *Journal of Applied Developmental Psychology*, *24*(5), 553–572. http://dx.doi.org/doi:10.1016/j.appdev.2003.08.005

Dean, B. K. (2017). *A hidden world of song: Spontaneous singing in the everyday lives of three- and four-year-old children at home.* Unpublished doctoral dissertation. University of Exeter, United Kingdom.

Deutsch, D. (1975). Facilitation by repetition in recognition memory for tonal pitch. *Memory & Cognition*, *3*(3), 263–266. https://doi.org/10.3758/BF03212909

de Vries, P. (2007). The use of music CDs and DVDs in the home with the under-fives: what the parents say. *Australian Journal of Early Childhood*, *32*(4), 18–21.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795-805. http://doi.org/10.1037/0021-9010.84.5.795

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology. In H.A. Ruger, C.E. Bussenius, E.R. Hilgard (Eds.), Trans. New York: Dover Publications. (Original work published in 1885).

Eckerdal, P., & Merker, B. (2009). Music and the 'action song' in infant development: An interpretation. *Communicative musicality: Exploring the basis of human companionship,* Oxford, UK: Oxford University Press.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, *12*(1), 20–30.

Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition*, *163*, 80–86. http://doi.org/10.1016/j.cognition.2017.02.017

Fausey, C.M., Jayaraman, S., & Smith, L.B. (2015). The changing rhythms of life: Activity cycles in the first two years of everyday experience. Paper presented at the 2015 biennial meeting of the Society for Research in Child Development, Philadelphia.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005

Fausey, C.M. & Mendoza, J.K. (in prep). Trio Corpus. https://homebank.talkbank.org/access/

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E. Hartung, S., et al., (1993). *The MacArthur Communicative Development Inventories: user's guide and technical manual.* Baltimore, MD: Paul H. Brookes Publishing Co.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.1111/j.1467-8624.1993.tb02933.x

Ford, M., Baer, C. T., Xu, D., Yapanel, U., & Gray, S. (2008). The LENA™ language environment analysis system: Audio specifications of the DLP-0121. Retrieved from http://www.lenafoundation.org/wp-content/uploads/2014/10/LTR-03-2_Audio_Sp ecifications.pdf.

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., & Rodriguez-Fornells, A. (2017). Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. *Scientific Reports*, *7*(1), 12451-12451. https://doi.org/10.1038/s41598-017-12798-2

Ganek, H., & Eriks-Brophy, A. (2017). Language ENvironment Analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, *72*, 77-85. http://doi.org/10.1016/j.jcomdis.2017.12.005

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science*, *15*(3), 398–407. http://doi.org/10.1111/j.1467-7687.2012.01142.x

Gesell, A., (1940). *The first five years of life: A guide to the study of the preschool child, from the Yale Clinic of Child Development*. *115*(5), 408. Harper & brothers.

Gesell, A., Ilg, F. L., Learned, J., & Ames, L. B. (1943). *Infant and child in the culture of today: the guidance of development in home and nursery school.* Oxford, England: Harper.

Gilkerson, J., & Richards, J. A. (2008). The LENA Foundation Natural Language Study. Boulder, CO: LENA Foundation. Retrieved from http://www.lenafoundation.org/wp-content/uploads/2014/10/LTR-02-2_Natural_Language_ Study.pdf (accessed 18 November 2015).

Gilkerson, J., Richards, J. A., & Topping, K. J. (2017). The impact of book reading in the early years on parent-child language interaction. *Journal of Early Childhood Literacy*, *17*(1), 92-110. http://doi.org/10.1177/1468798415608907

Goh, K. I., & Barabási, A. L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, *81*(4), 48002. https://doi.org/10.1209/0295-5075/81/48002

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(03), 515-531. https://doi.org/10.1017/S0305000907008641

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, *80*(3), 636-644. https://doi.org/10.1111/j.1467-8624.2009.01287.x

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*(6), 249–258. http://doi.org/10.1016/j.tics.2010.02.004

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–437. https://doi.org/10.1111/1467-9280.00476

Gómez, R. L., & Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience*, *18*, 57-69. https://doi.org/10.1016/j.dcn.2015.08.009

Greasley, A. E., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, *15*(1), 45-71. https://doi.org/10.1177/1029864910393417

Hannon, E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*(11), 466–472. https://doi.org/10.1016/j.tics.2007.08.008

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48–55. https://doi.org/10.1111/j.0956-7976.2005.00779.x

Hannon, E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639 – 12643. https://doi.org/10.1073/pnas.0504254102

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD, US: Paul H Brookes Publishing Co.

Hawes, B. L. (1974). Folksongs and function: Some thoughts on the American lullaby. *Journal of American Folklore*, *87*(344), 140–148. https://doi.org/10.2307/539474

Hills, T. (2012). The company that words keep: comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language*, *40*(3), 586–604. http://doi.org/10.1017/S0305000912000165.

Hintzman, D. L. (1969). Recognition time: Effects of recency, frequency and the spacing of repetitions. *Journal of Experimental Psychology*, *79*(1, Pt. 1), 192-194. https://dx.doi.org/10.1037/h0026943

Hintzman, D. L. (1970). Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology*, *83*(3), 435-444.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297-306.

Hintzman, D. L., Grandy, C. A., & Gold, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(4), 231-240.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition*, *8*(4), 378-382.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071-1083. http://doi.org/10.1177/0956797615581493.

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language*, *12*(2), 367–385.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. doi:10.1016/j.dr.2005.11.002.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 1-17. http://doi.org/10.3389/fpsyg.2011.00017.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. http://doi.org/10.1037//0096-1523.26.5.1570

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236-248.

Ilari, B. (2005). On musical parenting of young children: Musical beliefs and behaviors of mothers and infants. *Early Child Development and Care*, *175*(7-8), 647–660.

Ilari, B., & Polka, L. (2006). Music cognition in early infancy: infants' preferences and long-term memory for Ravel. *International Journal of Music Education*, *24*(1), 7–20.

Ireland, K., Parker, A., Foster, N., & Penhune, V. (2018). Rhythm and melody tasks for school-aged children with and without musical training: Age-equivalent scores and reliability. *Frontiers in Psychology*, *9*, 1-14. http://doi.org/10.3389/fpsyg.2018.00426.s001

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLoS ONE*, *10*(5), e0123780.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A Meta-analysis of the Spacing Effect in Verbal Learning: Implications for Research on Advertising Repetition and Consumer Memory. *Journal of Consumer Research*, *30*(1), 138–149. http://doi.org/10.2307/1248262.

Johnson-Green, E., & Custodero, L. (2002). The toddler top 40: Musical preferences of babies, toddlers, and their parents. *Zero to Three*, *23*(1), 47-48.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, *8*(5), 668-683.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2016). Decisions at the brink: locomotor experience affects infants' use of social information on an adjustable drop-off. *Frontiers in Psychology*, *7*, 1-11.

Karsai, M., Jo, H. H., & Kaski, K. (2018). *Bursty human dynamics*. Cham, Switzerland: Springer.

Karsai, M., Kaski, K., Barabási, A. L., & Kertész, J. (2012). Universal features of correlated bursty behaviour. *Scientific Reports*, *2*, 1-7.

Kaye, K. (1980). Why we don't talk 'baby talk' to babies. *Journal of Child Language*, *7*(3), 489–507.

Kaye, K., & Charney, R. (1981). Conversational asymmetry between mothers and children. *Journal of Child Language*, *8*(1), 35-49.

Kello, C. T., Bella, S. D., Médé, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society Interface*, *14*(135), 2-11. http://doi.org/10.1016/j.tics.2010.02.005

Kim, E. K., & Jo, H. H. (2016). Measuring burstiness for finite event sequences. *Physical Review E*, *94*(3), 1-7.

Ko, E. S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of child language*, *43*(2), 284-309.

Konishi, H., Stahl, A. E., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Individual differences in nonlinguistic event categorization predict later motion verb comprehension. *Journal of Experimental Child Psychology*, *151*, 18-32.

Koops, L. H. (2014). Songs from the car seat: Exploring the early childhood music-making place of the family vehicle. *Journal of Research in Music Education*, *62*(1), 52–65. http://doi.org/10.1177/0022429413520007

Kraus, N., Hornickel, J., Strait, D. L., Slater, J., & Thompson, E. (2014). Engagement in community music classes sparks neuroplasticity and language development in children from disadvantaged backgrounds. *Frontiers in Psychology, 5*, 1-9.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *Journal of Neuroscience*, *34*(36), 11913-11918.

Krause, A. E., & North, A. C. (2015). Music listening in everyday life: Devices, selection methods, and digital technology. *Psychology of Music*, *44*(1), 129–147. http://doi.org/10.1111/j.0038-4941.2005.00301.x

Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental science*, *20*(4), e12421.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*, 1503-1518.

Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, *10*(1), 110–120. http://doi.org/10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *PNAS, 100*(15), 9096–9101. doi: 10.1073/pnas.1532872100.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, *127*(3), 439–453. doi:10.1016/j.cognition.2013.02.002.

Lamont, A. (2008). Young children's musical worlds: musical engagement in 3.5-year-olds. *Journal of Early Childhood Research*, *6*(3), 247–261. doi:10.1177/1476718X08094449

Lee, D. K., Cole, W. G., Golenia, L., & Adolph, K. E. (2017). The cost of simplifying complex developmental phenomena: a new perspective on learning to walk. *Developmental Science*, *23,* e12615. http://doi.org/10.1111/cogs.12366

LENA Research Foundation Team [LRF Team]. (2014, August 27). *LENA Training Series – Session 2: Record* [Video file]. Retrieved from https://youtu.be/E4esMiyaL4k.

Levy, D., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *NeuroReport*, *12*(12), 2653-2657.

Longhi, E. (2009). `Songese': maternal structuring of musical interaction with infants. *Psychology of Music*, *37*(2), 195–213. http://doi.org/10.1177/0305735608097042

Loui, P., & Wessel, D. L. (2008). Learning and liking an artificial musical system: Effects of set size and repeated exposure. *Musicae Scientiae*, *12*(2), 207–225.

Luo, R., & Tamis-LeMonda, C. S. (2016). Mothers' verbal and nonverbal strategies in relation to infants' object-directed actions in real time and across the first three years in ethnically diverse families. *Infancy*, *21*(1), 65-89.

Margulis, E. H. (2014). *On repeat: How music plays the mind.* New York, NY: Oxford University Press.

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition*, *119*(3), 438–447. http://doi.org/10.1016/j.cognition.2011.02.008

Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*(6), 1050–1057. http://doi.org/10.3758/PBR.16.6.1050

Maurer, D., & Werker, J. F. (2013). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. http://doi.org/10.1002/dev.21177

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*(3), 323-338. https://doi.org/10.1093/llc/15.3.323

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014-9019. https://doi.org/10.1073/pnas.1105040108

Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-month-old infants, melodies are social. *Psychological Science*, *27*(4), 486-501. http://doi.org/10.1177/0956797615626691

Mehr, S. A., & Spelke, E. S. (2018). Shared musical knowledge in 11-month-old infants. *Developmental Science*, *21*(2), e12542. https://doi.org/10.1111/desc.12542

Mendoza, J. K., & Fausey, C. M. (2015). Reducing the musical input: Favorite songs in early infancy. *Poster session to be presented at the International Conference on Interdisciplinary Advances in Statistical Learning*, San Sebastián, Spain.

Mendoza, J. K., & Fausey, C. M. (2017, October 20). Characterizing music in infants' everyday environment. Retrieved from osf.io/eb9pw

Moog, H. (1976). The development of musical experience in children of pre-school age. *Psychology of Music*, *4*(2), 38-45. https://doi.org/10.1177/030573567642005

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive Science*, 1–38. https://doi.org/10.1111/cogs.12592

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-Term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433. http://doi.org/10.1177/0956797611416999

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language*, *25*(1), 95–120.

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. *Music Perception: An Interdisciplinary Journal*, *22*(1), 41-77. https://doi.org/10.1525/mp.2004.22.1.41

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54-65. https://doi.org/10.1037/0278-7393.14.1.54.

Oakes, L. M., & Spalding, T. L. (1997). The role of exemplar distribution in infants' differentiation of categories. *Infant Behavior and Development, 20*(4), 457–475. https://doi.org/10.1016/S0163-6383(97)90036-9

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318-6323. https://doi.org/10.1073/pnas.1300337110

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Partanen, E., Kujala, T., Tervaniemi, M. M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. *PLoS ONE*, *8*(10), e78946. http://doi.org/10.1371/journal.pone.0078946.t002

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. *Science*, *296*(5571), 1321-1323. https://doi.org/10.1126/science.1070223

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. http://doi.org/10.1177/0956797610389189.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130. http://doi.org/10.3758/s13423-014-0585-6.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1–11. http://doi.org/10.1016/j.cognition.2004.09.008

Plantinga, J., & Trainor, L. J. (2009). Melody recognition by two-month-old infants. *The Journal of the Acoustical Society of America*, *125*(2), EL58-62. http://doi.org/10.1121/1.3049583

Politimou, N., Stewart, L., Müllensiefen, D., & Franco, F. (2018). Music@Home: A novel instrument to assess the home musical environment in the early years. *PLoS ONE*, *13*(4), e0193819. http://doi.org/10.1371/journal.pone.0193819

Putkinen, V., Tervaniemi, M. M., & Huotilainen, M. (2013). Informal musical activities are linked to auditory discrimination and attention in 2-3-year-old children: an event-related potential study. *European Journal of Neuroscience*, *37*(4), 654–661. http://doi.org/10.1111/ejn.12049

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. http://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look Who's Talking NOW! Parentese Speech, Social Context, and Language Development Across Time. *Frontiers in Psychology*, *8*, 1008. https://doi.org/10.3389/fpsyg.2017.01008

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language*, *14*(2), 201–209. https://doi.org/10.1017/S0305000900012885

Rideout, V. J. (2013). Zero to eight: Children's media use in America. *Common Sense Media*, 1–40.

Rideout, V. J., Vandewater, E. A., & Wartella, E. A. (2003). Zero to six: Electronic media in the lives of infants, toddlers and preschoolers. *Kaiser Family Foundation*, 1-35.

Rovee-Collier, C. K. (1999). The development of infant memory. *Current Directions in Psychological Science*, *8*(3), 80-85. https://doi.org/10.1111/1467-8721.00019.

Rovee-Collier, C. K., Sullivan, M. W., Enright, M., Lucas, D., & Fagen, J. W. (1980). Reactivation of infant memory. *Science*, *208*(4448), 1159–1161. https://doi.org/10.1126/science.7375924

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608-635. http://doi.org/10.1111/j.1532-7078.2010.00033.x

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762-1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. http://dx.doi.org/10.1073/pnas.1419773112

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, *5*(1), 2–14. https://doi.org/10.1207/S15327957PSPR0501_1

Rubin-Rabson, G. (1940). Studies in the psychology of memorizing piano music: II. A comparison of massed and distributed practice. *Journal of Educational Psychology*, *31*(4), 270–284. https://doi.org/10.1037/h0061174

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928. https://doi.org/10.126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Loman, M. M., & Robertson, R. R.W. (2000). Infant memory for musical experiences. *Cognition, 77*(1), B15-B23. https://doi.org/10.1016/S0010-0277(00)00095-0

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, *69*(1), 94–104. https://doi.org/10.1111/j/1467-8624.1998.tb06136.x

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511-514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*(2), 457-468. https://doi.org/10.1037/0022-0663.98.2.457

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In *The Psychology of Music (Third Edition)* (pp. 499-550). https://doi.org/10.1016/B978-0-12-381460-9.00012-2

Schwab, J. F., & Lew-Williams, C. (2016). Repetition Across Successive Sentences Facilitates Young Children's Word Learning. *Developmental Psychology, 52*(6), 879-886. http://doi.org/10.1037/dev0000125

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*(4), 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x

Simmons, A. L. (2012). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education*, *59*(4), 357–368. http://doi.org/10.1177/0022429411424798

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833–870. http://doi.org/10.1016/j.cognition.2007.05.002.

Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the Experience Sampling Method. *Musicae scientiae*, *5*(1), 9-32. http://dx.doi.org/10.1177/102986490100500102

Slone, L. K. Abney, D. H., Smith, L. B. & Yu, C. (2017). *The Burstiness of Parents' Utterances: Classification of Vocal Temporal Structure during Parent-Child Naturalistic Interaction.* Poster presented at the International Conference on Interdisciplinary Advances in Statistical Learning, Bilbao, Spain.

Smith, L. B., Jayaraman, S., Clerkin, E. M., & Yu, C. (2018). The developing body creates a curriculum for infant statistical learning. *Trends in Cognitive Sciences*, 1–37. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., & Slone, L. K. (2017). A Developmental Approach to Machine Learning? *Frontiers in Psychology*, *8*, 2124. https://doi.org/10.3389/fpsyg.2017.02124

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251-258. https://doi.org/10.1016/j.tics.2014.02007

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343-348. https://doi.org/10.1016/S1364-6613(03)00156-6

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: the dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. http://doi.org/10.1111/j.1467-7687.2009.00947.x

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407-419. https://doi.org/10.1080/15248372.2014.933430

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565. https://doi.org/10.2307/1127555

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PLoS ONE*, *8*(11), e80646, 1-12. http://doi.org/10.1371/journal.pone.0080646

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., Suskind, E., Leininger, L., Goldin-Meadow, S., & Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*, *43*(2), 366-406. https://doi.org/10.1017/S0305000915000033

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), e12456. http://doi.org/10.1111/desc.12456

Thiessen, E. D. (2011). When variability matters more than meaning: The effect of lexical forms on use of phonemic contrasts. *Developmental Psychology*, *47*(5), 1448–1458. http://doi.org/10.1037/a0024439.

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*, 855. https://doi.org/10.3389/fpsyg.2013.00855

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*(4), 885–913. http://doi.org/10.1037/0033-295X.107.4.885

Tomasello, M. (2010). The social-pragmatic theory of word learning. *Pragmatics*, *10*(4), 401–413. https://doi.org/10.1075/prag.10.4.01tom

Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development*, *19*(1), 83-92. https://doi.org/10.1016/S0163-6383(96)90046-6

Trainor, L. J., & Hannon, E. E. (2013). Musical Development. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 1–75). London, UK: Elsevier Inc. https//dx.doi.org/10.1016/B978-0-12-381460-0.00011-0

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*(3), 289–296. https://doi.org/10.1111/j.1467-7687.2004.00348.x

Trehub, S. E. (1989). Infants' perception of musical sequences: Implications for language acquisition. *Journal of Speech-Language Pathology and Audiology*, 13(3), 3-11.

Trehub, S. E. (2006). Infants as musical connoisseurs. In G. McPherson (Ed.) *The child as musician: A handbook of musical development* (pp. 33-49). Oxford: Oxford University Press.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, *55*(3), 821–830.

Trehub, S. E., & Degé, F. (2015). Reflections on infants as musical connoisseurs. *The child as musician: A handbook of musical development*, *2*, 31-55.

Trehub, S. E., Hannon, E. E., & Schachner, A. (2010). Perspectives on music and affect in the early years. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion*. Oxford University Press.

Trehub, S. E., & Schellenberg, E. G. (1995). Music: Its relevance to infants. *Child Development*, *11*, 1–24.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development*, *58*(3), 741–749.

Trehub, S. E., & Trainor, L. J., (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research, 12,* 43– 77.

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental Psychology*, *33*(3), 500–507. https://doi.org/10.1037/0012-1649.33.3.500

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126-156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2013). That's more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, *23*(2), 105–122. http://doi.org/10.1002/icd.1824

U.S. Census Bureau, 2016 American Community Survey, 5-Year Estimates. https://www.census.gov/data/developers/data-sets/acs-5year.html

Valerio, W. H., Reynolds, A. M., Morgan, G. B., & McNair, A. A. (2012). Construct Validity of the Children's Music-Related Behavior Questionnaire. *Journal of Research in Music Education*, *60*(2), 186–200. http://doi.org/10.1177/0022429412444450.

Vazquez, A., Oliveira, J. G., Dezsö, Z., Goh, K.-II., Kondor, I., & Barabási, A.-L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E*, *73*(3), 435. https://doi.org/10.1103/PhysRevE.73.036127

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*(1), 71–86. https://doi.org/10.1016/0749-596X(88)90049-6

VanDam, M., & Silbert, N. H. (2016). Fidelity of automatic speech processing for adult and child talker classifications. *PloS One, 11*(8), e0160588. https://doi.org/10.1371/journal.pone.0160588

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in speech and language, 37*(2), 128-142. https://doi.org/10.1055/s-0036-1580745

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. http://doi.org/10.1111/j.1467-8624.2012.01781.x

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. http://doi.org/10.1016/j.cognition.2008.07.013

Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2006). Infants' memory for musical performances. *Developmental Science*, *9*(6), 583–589. https://doi.org/10.1111/j.1467-7687.2006.00536.x

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. http://doi.org/10.1016/j.patcog.2005.01.025

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. http://doi.org/10.1177/0956797613488145.

Weiss, M. W., Schellenberg, E. G., & Trehub, S. E. (2017). Generality of the Memory Advantage for Vocal Melodies. *Music Perception*, *34*(3), 313–318. http://doi.org/10.1525/mp.2017.34.3.313

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the Way She Sings: Enhanced Memory for Vocal Melodies. *Psychological Science*, *23*(10), 1074–1078. http://doi.org/10.1177/0956797612442552

Weiss, M. W., Schellenberg, E. G., Trehub, S. E., & Dawber, E.J. (2015). Enhanced processing of vocal melodies in childhood. *Developmental Psychology, 51*(3), 370-377. http://doi.org/10.1037/a0038784

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils Dilate for Vocal or Familiar Music. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1061-1065. http://doi.org/10.1037/xhp0000226

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Williams, K. E., Barrett, M. S., Welch, G. F., Abad, V., & Broughton, M. (2015). Associations between early shared music activities in the home and later child outcomes: Findings from the Longitudinal Study of Australian Children. *Early Childhood Research Quarterly, 31*, 113-124. https://doi.org/10.1016/j.ecresq.2015.01.004

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559).

Xu, D., Yapanel, U., & Gray, S. (2009). Reliability of the LENA Language Environment Analysis System in young children's natural home environment. *LENA Foundation*.

Young, S. (2008). Lullaby light shows: everyday musical experience among under-two-year-olds. *International Journal of Music Education*, 26(1), 33–46. doi:10.1177/0255761407085648.

Young, S., & Gillen, J. (2007). Toward a revised understanding of young children's musical activities: Reflections from the "Day in the Life" project. *Current Musicology*, *84*(1), 7–27.

Young, S., Street, A., & Davies, E. (2007). The Music One-to-One project: Developing approaches to music with parents and under-two-year-olds. *European Early Childhood Education Research Journal*, *15*(2), 253–267. https://doi.org/10.1080/13502930701321675

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414-420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*(1), 1-22. https://doi.org/10.3758/s13423-013-0443-y

Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 731–741. http://doi.org/10.1037/a0015005

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, *33*, 251–256. https://doi.org/10.1080/00221309.1945.10544509