DISSERTATION

Titel der Dissertation

Rhythm, patterns and the primate roots of language evolution

Verfasser

Andrea Ravignani

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2014

To my little sister,
and my great parents.

Acknowledgments

I am grateful to my lead supervisor, Prof. W. Tecumseh Fitch, for believing in me from the beginning, employing me on his grant and willing to supervise my experimental research even though I had an extremely theoretical background, for giving me much freedom to explore the research topics I liked the most and to travel to conferences and other labs, for teaching me the importance of brevity and clarity when disseminating science, for gradually challenging me on my weaknesses as a scientist and a person.

I am thankful to my external supervisor, Prof. Simon Kirby, for accepting to be involved in my PhD and inviting me to Edinburgh for a semester, for all his efforts which made my Scottish experience so unique, for showing me the importance of modeling, for his lovely lectures and great advice during our meetings, for showing me how science and art can combine in unique ways. It has been a real privilege.

Thank you to Bruno Gingras and Nina Stobbe, for being my additional, unofficial advisors, teaching me so much about experimental design, reading my manuscripts, and encouraging me during difficult moments, and Kenny Smith and Jelle Zuidema,

Contents

1 | Introduction

1.1 Language, Music and Animal Communication: A Comparative Perspective

When listening to human speech, two straightforward characteristics are readily perceivable. Speech is composed of discrete elements, and it unfolds over time. These general features are arguably shared with, among others, human music and at least some non-human animal communication systems. However, while language and music critically rely on both *temporal patterning* and *complex relations* among constituent acoustic elements, the same does not seem to hold for most non-human primate communication systems (ten Cate & Okanoya, 2012). Nonetheless, the systematic lack of a particular structural feature in the signal *production* repertoire of an organism does not necessarily imply the organism's inability to *perceive* that structural feature in external stimuli. This consideration is particularly important to understand the evolution of the cognitive ability to process language and music. Human cognition shares a number of psychological capacities with other primate species by common ancestry; similarly, language and music processing might rest, at least partially, on such shared capabilities, which evolved millions of years ago.

Proximate questions about human abilities to process linguistic and musical structures can be answered through experimental tests. Human adults, infants and non-human animals are tested on a variety of psychological tasks, and their behavioral or physiological responses are used as a proxy to measure an underlying cognitive ability. Comparing different species' performance on similar tasks (Fitch, Huber, & Bugnyar, 2010) can provide insights regarding human cognitive skills at different evolutionary stages.

Ultimate explanations for the emergence of language and music can be obtained by adopting approaches from (evolutionary) dynamical systems (Nowak & Krakauer, 1999; Plotkin & Nowak, 2000; Sigmund, 1993) and agent-based modeling (Kirby, 2001; Miranda, Kirby, & Todd, 2003; Mitchell, 2009; Todd, 2000). In evolutionary biology, mathematical models are built to test and compare alternative evolutionary scenarios. The approach adopted is a good example of inverse problems: the initial conditions are usually unknown and need to be reconstructed, while the final modern state is observable. For the particular cases of language, music and animal communication, these initial states are ancestral forms of communication in some common primate, vertebrate or animal ancestor with humans. The final conditions could be any number of possible behavioral and cognitive systems, spanning crickets chirping in unison, humans jamming on a jazz standard, or a pair of gibbons bonding by "singing" to each other (Ravignani, Bowling, & Kirby, in press; Wallin, Merker, & Brown, 2001).

1.2   Key Questions

Analogously to paleobiologists investigating morphological changes in trilobites, many researchers in the evolution of human cognition are concerned with phylogenetic, behavioral and cultural changes of (possibly correlated) traits over time. In particular, language and music appear to be to prominent (although possibly indirect) products of human evolution (Fitch, 2006, 2010; Patel, 2010; Wallin, Merker, & Brown, 2001).

The main question underlying this dissertation is how humans acquired the cognitive abilities to process crucial structures in music and language and at what point of our evolutionary history this happened. More specifically, I tackle three sub-questions, roughly corresponding to thesis chapters:

- Even though other primates do not have language, can they nonetheless perceive key features of speech patterns?

- Do non-vocal patterns produced by other primates resemble those found in human language and music?

- What sort of evolutionary dynamics drove our abilities for timing and rhythm

production necessary for speech and music? Did group competition or cooperation play a role?

The papers composing this dissertation tackle the questions above, investigating the sense of rhythm and patterning abilities in monkeys, apes and a hypothetical pre-musical, extinct human ancestor. The methods I use encompass tools from experimental psychology, mathematical biology and computer science, while building on previous knowledge from animal behavior, neuroscience, linguistics, musicology and cognitive psychology.

1.3   Overview of Chapters

In chapter 2, I point out how dependencies between sensory elements located at an arbitrary distance from one another are a defining feature of language and music. While humans often encounter and process such dependencies, it remains unclear whether other primates have this capability. My experiment shows that the New World monkey *Saimiri sciureus* is sensitive to such dependencies and capable of discriminating stimuli containing sensory dependencies from those lacking them.

Chapter 3 is a description of an innovative video-sonification coding methodology I developed and used to code experimental data from a number of preferential-looking experiments (among others, those in chapter 2)

Chapter 4 consists of a commentary to a position paper by Ackermann, Hage and Ziegler (henceforth AHZ) in the journal *Behavioral and Brain Sciences* (in press). AHZ provide an intriguing account of the evolution of neural underpinnings for language. They hypothesize a two-stage model, leading from monkey-like emotional vocalizations in early human ancestors, to the full-fledged neural capacities for speech present in modern humans. The first stage would see increased neural control over laryngeal muscles. The second stage would build upon the first and see the development of higher cortical functions, possibly leading to complex syntax and phonology. Crucially, the authors claim that the first stage (laryngeal control) might have occurred under social pressures for group synchronization. AHZ suggest that the inability of chimpanzees to move in synchrony with each other or external stimuli would constitute evidence supporting the acquisition of laryngeal control after the divergence

between human and chimpanzee lineages. However, I argue that this conclusion is based on lack of, rather than negative, evidence. Moreover, I point out that this scenario is much more consistent with hypotheses on the origins of human musicality (Darwin, 1871; Fitch, 2009; Geissmann, 2000; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Kirschner & Tomasello, 2009, 2010; Merker, 2000; Merker, Madison, & Eckerdal, 2009). In particular, AHZ's scenario assumes a strong relationship between vocal learning and general cognitive timing abilities (especially for rhythm and synchrony). The hypothesis that vocal learning is a prerequisite for auditory-motor entrainment, originally suggested by Patel (2006), has some empirical support, but has also been recently undermined by negative evidence. Sea lions, flexible learners although incapable of vocal mimicry, can reliably synchronize to a pulse at different tempi (Cook, Rouse, Wilson, & Reichmuth, 2013). Hence, a speechless animal species exhibits a behavioral capability identical to that postulated by AHZ as a by-product of our ancestors' path towards language.

Three relevant conclusions follow from my commentary. First of all, AHZ need to rethink and sharpen their evolutionary account for the emergence of speech. As vocal learning appears to be neither necessary nor sufficient for auditory-motor entrainment across species, one might need to postulate additional constraints for the emergence of speech apart from social coordination. Second, I stress how language and music are connected and that their research fields must inform each other (Fitch, 2009; Kirby, 2009). In particular, studying language and music in close connection can be decisive to understand their evolutionary history and order of emergence. Finally, I emphasize the importance of collecting experimental data on rhythmic abilities in non-human animals. In particular, there is a need to test non-vocal learners and apes in rhythmic tasks.

Chapter 5 describes two devices I developed to enable controlled rhythmic experiments in chimpanzees in semi-captive setups. Notably, running experiments with these devices will allow replacing the lack of evidence on chimpanzees' entrainment, used by AHZ to support their thesis, with solid data. While this paper is mostly concerned with the methodology, its last sections are devoted to pilot data from chimpanzees and corresponding ethograms.

A number of scholars have suggested how social pressures might have led to proto-

musical psychological abilities in our ancestors (Fitch, 2009; Geissmann, 2000; Hagen & Bryant, 2003; Hagen & Hammerstein, 2009; Kirschner & Tomasello, 2009, 2010; Merker, 2000; Merker et al., 2009). In particular, sociality and cooperation are emphasized among evolutionary pressures leading to an ability to synchronize movements or signals with peers. Chapter 6 suggests how selfishness can equally well lead to group synchrony and appearance of coordination (Greenfield & Roizen, 1993). A number of bird, frog and cricket species have been studied for their ability to chorus: individuals adjust the timing of their calls depending on the others, which can lead to synchrony and rhythmic acoustic patterns. In my paper, I adapt a mathematical framework previously developed by W.D. Hamilton in the context of spatial aggregation (the "selfish herd" theory; Hamilton, 1971), by employing it to model chorusing behavior in time, and possibly connect it to music origins. I suggest how each feature of Hamilton's original aggregation model has a counterpart in animal chorusing behavior. Therefore I provide a rigorous framework to investigate chorusing and synchronizing agents. I suggest how this framework might be not only relevant to animal behavior alone, but also significant to reconstruct the evolutionary origins of human rhythm in language and music.

The paper I formally comment on (Herbers, 2013) is an overview of many interesting contributions to a special *Biology Letters* issue published in occasion of the 50th anniversary of Hamilton's inclusive fitness theory. Although my comment does not relate to that part of Hamilton's work, it still follows Herbers' lead in celebrating Hamilton's genius.

Finally, chapter 7 provides (i) a general conclusion, connecting all chapters and broadening their scope, and (ii) descriptions of how the research presented here contributes to the research fields concerned and science in general.

References

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California Sea Lion (*Zalophus californianus*) Can Keep the Beat: Motor Entrainment to Rhythmic Auditory Stimuli in a Non Vocal Mimic. *Journal of Comparative Psychology* (127), 1-16.

Darwin, C. (1871). *The descent of man and selection in relation to sex*: John Murray;

London, UK.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173-215.

Fitch, W. T. (2009). The biology and evolution of rhythm: Unraveling a paradox *Language and Music as Cognitive Systems*: Oxford University Press. Oxford, UK.

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

Fitch, W. T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron*, 65(6), 795-814.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. *The origins of music*, 103-123.

Greenfield, M. D., & Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature*, 364(6438), 618-620.

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, 14(1), 21-51.

Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae*, 13(2 suppl), 291-320.

Hamilton, W. D. (1971). Geometry for the selfish herd. *Journal of theoretical Biology*, 31(2), 295-311.

Herbers, J. M. (2013). 50 Years on: the legacy of William Donald Hamilton. *Biology Letters*, 9(6), 20130792.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2), 102-110.

Kirby, S. (2009). Darwin's musical protolanguage: an increasingly compelling picture *Language and Music as Cognitive Systems* (p. 96): Oxford University Press. Oxford, UK.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of experimental child psychology*, 102(3), 299-314.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5), 354-364.

Merker, B. (2000). Synchronous chorusing and the origins of music. *Musicae Scientiae*, 3(1 suppl), 59-73.

Merker, B., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in

human rhythmic entrainment. *Cortex*, 45(1), 4-17.

Miranda, E. R., Kirby, S., & Todd, P. (2003). On computational models of the evolution of music: From the origins of musical taste to the emergence of grammars. *Contemporary Music Review*, 22(3), 91-111.

Mitchell, M. (2009). *Complexity: A guided tour*: Oxford University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14), 8028-8033.

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

Patel, A. D. (2010). *Music, language, and the brain*: Oxford university press.

Plotkin, J. B., & Nowak, M. A. (2000). Language evolution and information theory. *Journal of Theoretical Biology*, 205(1), 147-159.

Ravignani, A., Bowling, D., & Kirby, S. (in press). *The psychology of biological clocks: a new framework for the evolution of rhythm*. Paper presented at The 10th International Conference on the Evolution of Language, Vienna, Austria. Singapore: World Scientific Publishing Co.

Sigmund, K. (1993). *Games of life: exploration in ecology, evolution and behavior*: Oxford, UK: Oxford University Press.

ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1984-1994.

Todd, P. (2000). Simulating the evolution of musical behavior. *The origins of music*, 361-388.

Wallin, N. L., Merker, B., & Brown, S. (2001). *The origins of music*: MIT press.

2 | Pitch Patterns

Action at a distance: dependency sensitivity in a New World primate

Andrea Ravignani, Ruth-Sophie Sonnweber, Nina Stobbe and W. Tecumseh Fitch

Department of Cognitive Biology, University of Vienna, Althanstrasse, 14, Vienna 1090, Austria

Sensitivity to dependencies (correspondences between distant items) in sensory stimuli plays a crucial role in human music and language. Here, we show that squirrel monkeys (*Saimiri sciureus*) can detect abstract, non-adjacent dependencies in auditory stimuli. Monkeys discriminated between tone sequences containing a dependency and those lacking it, and generalized to previously unheard pitch classes and novel dependency distances. This constitutes the first pattern learning study where artificial stimuli were designed with the species' communication system in mind. These results suggest that the ability to recognize dependencies represents a capability that had already evolved in humans' last common ancestor with squirrel monkeys, and perhaps before.

1. Introduction

Human language relies on several basic and indispensable cognitive skills, including the detection of relationships or 'dependencies' between stimuli that are non-contiguous in space or time. Dependency sensitivity, defined here as the ability to recognize that two non-contiguous sensory items are related (e.g. belong to the same perceptual class), is part of everyday sensory experience and crucial for many aspects of human cognition [1–3].

The perceived 'musicality' of some languages results from how syllable types are combined to form words. In Turkish, for instance, the plural of a noun is formed by adding a suffix to its singular form. Crucially, the suffix's vowel must belong to the same *acoustic class* as the noun's last vowel, hence establishing an abstract dependency (not between specific items). Hungarian, like Turkish, also exhibits such 'vowel harmony'. In Hungarian, the first and last vowels depend on each other but they can be separated by several neutral syllables, thus exhibiting an arbitrary-distance dependency between non-adjacent elements.

Dependencies that are *both* abstract (applying to classes of elements) and occur at variable distance are essential in productively open systems like language and music. The evolutionary origins, e.g. in primates, of the cognitive ability to detect dependencies are unknown. Human infants already possess the capacity to track non-adjacent dependencies in natural language [3]. In 'artificial languages', dependencies between non-adjacent elements are particularly easy to detect if occurring between perceptually similar elements [2,4] or at the edges of stimuli [5].

Previous comparative animal research has demonstrated awareness of dependencies *either* occurring at a *fixed distance* [6,7] or between *specific items* [5]. Detection of abstract dependencies at arbitrary variable distances (crucially beyond one intervening element, already shown in [4,7]) has never been demonstrated before in a non-human animal (though see [8] for initial hints). The current study tested the hypothesis that a non-human primate species could detect abstract, non-adjacent dependencies in acoustic stimuli, even when dependencies occurred over an arbitrary variable number of intervening sounds.

19
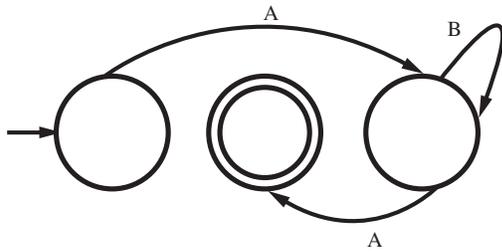
Figure 1. Finite state machine generating and recognizing AB^nA strings. Every transition (arrows) from one state to another (circles) produces a new element of the string (A or B). Any sequence of transitions beginning in the (leftmost) start state (denoted with an arrow) and finalizing in the accept state (denoted by concentric circles) produces a string containing a dependency.

We used formal language theory as a precise mathematical framework to characterize string complexity [9,10]. The formal language used to generate stimuli [11], AB^nA (not employed in empirical research before) captures a single arbitrary-distance dependency between similar elements at its edges (figure 1). AB^nA characterizes strings with one A at the beginning, one A at the end, and n repetitions of B in between. Any other combination of As and Bs violates this rule. Notably, this pattern captures aspects of naturally occurring linguistic phenomena (as seen for Hungarian), while taking into account edge and perceptual similarity effects in designing the stimuli [4,5].

2. Material and methods

(a) Subjects and experimental procedure

Six group-housed squirrel monkeys (*Saimiri sciuresu*) were individually trained over eight months to enter a sound booth voluntarily. The experiment used a habituation–discrimination paradigm [5], consisting of exposure to habituation stimuli and individual testing using novel stimuli. If able to perceive the relation of dependency between the first and last A elements, monkeys should react differently when tested with sounds obeying, versus those violating, the rule.

Animals were habituated to 360 stimuli (two sessions over 2 days), played in random order to all individuals simultaneously (80 min total).

During the test phase, individual monkeys entered the acoustic booth and sat on a perch. One experimenter inside the booth (wearing headphones playing custom-generated masking music to prevent unconscious cuing) fed insects to the subject *between* playbacks.

(b) Stimuli description

Exposure and test stimuli were generated following the AB^nA rule. As and Bs were mapped to two pure sine wave tone classes, high (H) and low (L), consisting of 44 elements each. 'Low' tone frequencies were randomly and uniformly sampled from an interval centred at 2 kHz; 'High' tones had mean frequency 11 kHz (interval endpoints: $\pm10\%$ of mean; duration: 225 ± 15 ms). For the habituation, As were matched to the low category and Bs to the high category, $(n = 1,\ldots,3)$. Thus, monkeys were habituated to a set of three patterns: LHL, LHHL and LHHHL.

The frequency classes here were chosen because: (i) squirrel monkeys are equally sensitive to sounds in these frequency ranges [12], (ii) durations and frequencies of species specific vocalizations exist in these ranges [13] and (iii) pure tones avoid potential confounds involved in using recorded monkey calls, where reactions might be elicited by the meaning attached to calls, rather than patterns formed from them. Furthermore, inspired by the time-domain characteristics of squirrel monkeys' vocalizations [13], the tones composing our stimuli are markedly shorter than the units employed in previous similar experiments. To our knowledge, this is the first animal pattern perception experiment using pure tone stimuli specifically tailored to a particular species' communication system.

(c) Video coding and data analysis

We eliminated the possibility of coder bias with three concurrent coding strategies: (i) reactions were videotaped and coded by multiple raters, who were (ii) otherwise not involved in this experiment, and blind to the hypothesis being tested and (iii) completely blind to which stimulus was played [14], to ensure that no bias could affect coding decisions. Our method [14] involves replacing the original audio of the experiment with sinewave placeholders, ruling out knowledge of which stimulus was played.

Three colleagues annotated head turns towards the loudspeaker of 45° or more. Before video coding started, we established the criterion that only head turns starting after stimulus onset and within 7 s from the playback onset (four times the duration of the longest stimulus) would be extracted from the annotations and further analysed. The average index of concordance [15] was 0.875 (calculated on 24 trials unused in this study).

Data analysis was performed in SPSS and STATA. Parametric tests were used after testing for normality (Shapiro–Wilk) and homoskedasticity (Levene) ($n = 6$ or $n = 4$, all p-values ≥ 0.27).

(d) Test 1

Test 1 investigated whether squirrel monkeys (i) acquired the dependency rule, showing different reactions between stimuli obeying or violating it, (ii) generalized the rule over new instantiations of sound patterns and (iii) generalized to dependencies between low sounds separated by a previously unheard number of intervening high sounds (extensions).

Half the stimuli for test 1 were *consistent* with the exposure rule (C_1, index indicating test 1) and half represented *violations* (V_1) of the dependency rule (table 1). Consistent stimuli either followed the same overall pattern and length as habituation stimuli, but involved novel tone combinations (the particular tones composing each pattern were re-sampled anew from their respective pitch classes) or contained a previously unheard number of intervening low tones, generalizing the rule by induction over n.

(e) Test 2: meta-generalization

Before this test, no novel habituation stimuli were presented. The only difference between test 2 and test 1 was that the mapping between low and high tones was inverted, so that in test 2 As corresponded to high tones and Bs to low frequencies (e.g. HLH).

A monkey succeeding at test 2 should perceive a habituation stimulus like LHHL and a test stimulus, like HLLLH as belonging to the same class, while regard a sound such as HLLL as a violation to the original rule LH^nL.

3. Results

For each monkey, $PR(V_1)$ was greater than or equal to $PR(C_1)$ (PR = percentage of reactions), with $PR(C_1) = 60.4\%$ and $PR(V_1) = 77.1\%$ (s.d.: 18.4 both). Overall, $PR(V_1)$ differed significantly from $PR(C_1)$ (figure 2; paired t-test, $n = 6$, $t = 3.16$, $p = 0.025$). Responses did not differ between stimuli missing the first or last low tone ($n = 6$, $t = 0.54$, $p = 0.611$; see electronic supplementary material, S1).

Table 1. Experimental patterns. Breakdown of stimuli type by class and subclass, and number (specified when greater than 1) of different exemplars the monkeys were exposed to during the habituation and the tests.

stimulus class	subclass	Test 1	Test 2
habituation		LHL (60), LH^2L (120), LH^3L (180)	
consistent	repetition	LHL, LH^2L, LH^3L (2)	HLH, HL^2H, HL^3H (2)
	extension	LH^4L (2), LH^5L (2)	HL^4H (2), HL^5H (2)
violation	missing first	HL, H^2L, H^3L, H^4L	LH, L^2H, L^3H, L^4H
	missing last	LH, LH^2, LH^3, LH^4	HL, HL^2, HL^3, HL^4

In test 2, the monkeys did not show any difference between $PR(C_2)$ and $PR(V_2)$. (paired t-test, $n = 4$, $t = 1.98$, $p = 0.141$.) Taking test 2 after test 1 might have generated order effects (monkeys could have habituated to two violation stimuli, HL and LH, presented in test 1, see electronic supplementary material, S1). In fact, a paired t-test, comparing $PR(C_2)$ to $PR(V'_1)$ (novel violations) showed a significant effect of stimulus type on response (figure 2, $n = 4$, $t = 4.64$, $p = 0.019$), suggesting a generalization from LH^nL to HL^nH.

We ran a repeated measures ANOVA involving test type (test 1 versus test 2) and grammaticality (violation versus consistent). Reactions to LH and HL were also excluded in test 1, to maintain a one-to-one correspondence between stimuli across tests. We found an effect of grammaticality (2×2 ANOVA, $n = 4$, $F = 23.14$, $p = 0.017$); but no effect of test type ($F = 0.06$, $p = 0.822$) and no interactions ($F = 0.27$, $p = 0.638$).

Figure 2. Histograms for percentage reactions in test 1 (left, $n = 6$) and test 2 (right, $n = 4$). The average percentage consistent (white) and violation (grey) trials that elicited a reaction are displayed in each case (mean $\pm$ s.e.m.). For test 2, reactions to novel violations (see Results) are shown.

4. Discussion

Squirrel monkeys consistently recognized and generalized the pattern AB^nA at different levels, showing sensitivity to arbitrary-distance dependencies.

Test 1 showed that our subjects effectively generalized the specific pattern beyond specific pitches or stimulus lengths. Rather than matching specific pitches, the monkeys attended to *relations* between sound categories when discriminating between stimuli containing or lacking a dependency. Together, both tests suggest that generalization to a higher level of abstraction, featuring previously unseen combinations of elements, occurred based solely on specific instantiations of the sound classes heard during the exposure. We were able to rule out some alternative, lower level explanations through our design and additional tests (e.g. monkeys do not attend exclusively to one of the stimulus' edges, see electronic supplementary material, S1): testing primates in an operant setup could help exclude additional simpler discrimination strategies.

Previous animal research has dealt mainly with dependencies occurring at a fixed distance: namely, at no more than one element apart. The formal language AB^nA we used has relatively low computational complexity (finite state, strictly three-local [9]), but nonetheless possesses adequate representational power to capture dependencies between elements at arbitrary distance. In fact, the presence of sensory dependencies and grammar complexity can be orthogonal questions. Previous experiments whose stimuli included the AB^nA substring do not provide evidence of dependency processing: super-grammars featuring AB^nA can be mastered (significantly) without processing dependencies, and vice versa.
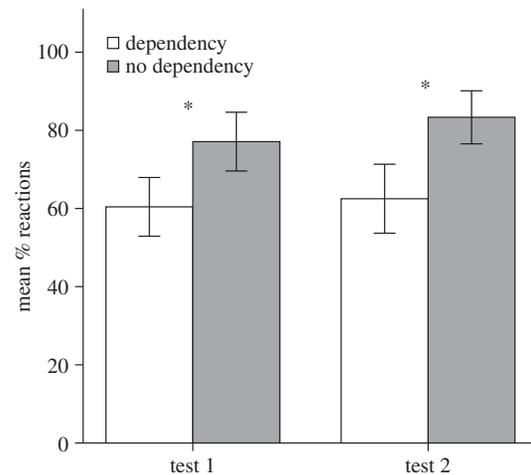
Pattern perception experiments aim to test cognitive abilities involving high-level properties of the patterns, rather than basic acoustic perception skills or semantic biases [10]. Many previous studies used human speech syllables, which may not be salient to all animal species. Pilot work with patterns made up of human syllables indicated a lack of discrimination between stimuli classes: our short high-frequency tone units might have enhanced performance.

Squirrel monkeys are sensitive to abstract dependencies of different lengths and can generalize to new lengths and auditory parameters of the stimuli. Human and squirrel monkey lineages diverged at least 36 Ma [16], and our findings suggest that dependency sensitivity was present in these primate ancestors. If so, most living apes and monkeys should exhibit this ability, which need not be evolutionarily related to communication and vocal flexibility, but could be a by-product of other cognitive abilities.

Despite its value in both language and music, dependency sensitivity apparently did not evolve specifically for use in these cognitive systems. Although no squirrel monkey will probably ever speak a human language, these monkeys possess the cognitive potential to recognize the rule generating plurals of Turkish nouns, or many other linguistic phenomena.

References

1. Gebhart AL, Newport EL, Aslin RN. 2009 Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* **16**, 486–490. (doi:10.3758/PBR.16.3.486)

2. Newport EL, Aslin RN. 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* **48**, 127–162. (doi:10.1016/S0010-0285(03)00128-2)

3. van Heugten M, Shi R. 2010 Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *J. Acoust. Soc. Am.* **128**, EL223–EL228. (doi:10.1121/1.348 6197)

4. Creel SC, Newport EL, Aslin RN. 2004 Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **30**, 1119–1130. (doi:10.1037/0278-7393.30.5.1119)

5. Endress AD, Carden S, Versace E, Hauser MD. 2010 The apes' edge: positional learning in chimpanzees and humans. *Anim. Cogn.* **13**, 483–495. (doi:10.1007/s10071-009-0299-8)

6. Murphy RA, Mondragón E, Murphy VA. 2008 Rule learning by rats. *Science* **319**, 1849–1851. (doi:10.1126/science.1151564)

7. Newport EL, Hauser MD, Spaepen G, Aslin RN. 2004 Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* **49**, 85–117. (doi:10.1016/j.cogpsych.2003.12.002)

8. Versace E. 2008 Processing regularities without language. PhD thesis, Università degli studi di Trieste, Italy.

9. Jäger G, Rogers J. 2012 Formal language theory: refining the Chomsky hierarchy. *Phil. Trans. R. Soc. B* **367**, 1956–1970. (doi:10.1098/rstb.2012.0077)

10. ten Cate C, Okanoya K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B* **367**, 1984–1994. (doi:10.1098/rstb.2012.0055)

11. Chomsky N. 1956 Three models for the description of language. *IRE Trans. Inform. Theory* **2**, 113–124. (doi:10.1109/TIT.1956.1056813)

12. Wienicke A, Häusler U, Jürgens U. 2001 Auditory frequency discrimination in the squirrel monkey. *J. Comp. Physiol. A* **187**, 189–195. (doi:10.1007/s003590100189)

13. Biben M, Symmes D, Masataka N. 1986 Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys (*Saimiri sciureus*). *Behaviour* **98**, 259–273. (doi:10.1163/15685 3986X00991)

14. Ravignani A, Fitch WT. 2012 Sonification of experimental parameters as a new method for efficient coding of behavior. In *Proceedings of Measuring Behavior*, 8th International Conference on Methods and Techniques in Behavioral Research, Utrecht, The Netherlands, August 28–31 2012 (eds AJ Spink, F Grieco, OE Krips, LWS Loijens, LPJJ Noldus, PH Zimmerman), pp. 376–379. Utrecht, The Netherlands: Noldus Information Technology.

15. Martin P, Bateson P. 1993 *Measuring behaviour: an introductory guide*. Cambridge, UK: Cambridge University Press.

16. Glazko GV, Nei M. 2003 Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434. (doi:10.1093/molbev/msg050)

Action at a Distance: Dependency Sensitivity in a New World Primate

Andrea Ravignani, Ruth-Sophie Sonnweber, Nina Stobbe, W. Tecumseh Fitch

COMPLETE METHODS

Subjects. We tested six squirrel monkeys (*Saimiri sciureus*), three males and three females, aged between 2 and 6 years. These primates were group-housed (with two more conspecifics who did not take part in the experiments) in an indoor enclosure (90 m^3) at the University of Vienna. They were fed twice a day and had *ad libitum* access to water. The animals took part in the experiments voluntarily and were free to leave the experiment or training session at any time. Positive reinforcement only was used for all training. Two monkeys, A and T, only completed the first test. For T the test chamber door could not be entirely closed during the experiment, hence he was the last individual to take Test 1. All experimental procedures were non-invasive and carried out in accordance with the relevant Austrian legislation.

Stimuli Description. A.R. and W.T.F wrote custom Python (www.python.org) software to create stimuli, process data and code videos. The stimuli were generated by concatenating pure sine wave tones belonging to two pitch classes. *A*s and *B*s were initially mapped to two tone classes, high (H) and low (L), consisting of 44 elements in each set. Tone frequencies in class L

1

were randomly and uniformly sampled from an interval centred at 2kHz; class H tones had mean frequency 11kHz. In both cases, the endpoints of the intervals were located at ±10% of the mean. The duration of each tone was randomly sampled between 210 and 240 ms. Amplitudes were randomly normalized to a peak amplitude varying between 40 and 60% of maximum. Pure tones were concatenated using trapezoidal envelopes of 250 ms each, thus providing equal durations to stimuli with same number of tones and variable-length pauses. Concatenating the variable-length tones within a fixed length window of 250 ms granted a variable-length silence (average of 50ms, uniformly distributed over a 50±30 ms interval). Silence between tones was introduced to ease segmentation of elements composing a stimulus. Additionally, segmentation was insured by the chosen frequency range. This was chosen by extrapolation from the just noticeable difference thresholds in squirrel monkeys' auditory perception found by Wienicke et al. [1].

Tones were then arranged following the AB^nA rule: for the habituation, As were matched to the L category and Bs to the H category, with n = 1, 2 or 3. The specific tone chosen within a class was again randomly determined.

Monkeys were thus habituated to a set of three patterns: LHL, LHHL and LHHHL. These are the shortest possible strings that can be generated by the AB^+A rule. There were 360 *habituation* stimuli; they were unevenly divided in proportions 60:120:180 to control for bigram transition probabilities to the extent possible here (see below).

Controlling for transition probabilities. We aimed to prevent discrimination

based on superficial, statistical properties inferable from only parts of the stimulus (substrings) [2]. A simple heuristic could involve tracking the occurrences of bigrams during the habituation, and using only this information later, when discriminating between classes. Two is the length of the shortest substring in the habituation stimuli; hence a bigram strategy (as opposed to looking at longer n-grams) is the only one applicable to the entire stimulus set. Thus, a first order Markov chain could be used as a shortcut to discriminate between stimuli types. For example, violation stimuli missing the final A might elicit a reaction due to the low transitional probability of BB relative to the (highly probable) BA bigram.

By exposing our subjects to habituation stimuli with lengths of uneven proportions (AB^3A were three times more frequent than ABA), we avoided this potential confound. With our proportions of exposure stimuli, the probability of hearing an A after a B is equal to that of hearing a B followed by a B, both being equal to 0.5.

Experimental setup: Habituation. Sounds were played using iTunes (habituation) and the custom software Playback 3.0 (tests) running on a MacBook Pro laptop computer, connected to one of two loudspeakers, JBL Control 2P, through an audio mixer (Mackie 802-VLZ3). One of the loudspeakers was installed outside the testing booth, 50cm from the enclosure and broadcasted stimuli during the habituation phase. Habituation stimuli were played to the entire group. During habituation playbacks, the minimum registered sound pressure level was 61dB (relative to 39 dB for

3

silence) within the monkey enclosure. Sound pressure levels were measured using an SL-400 Voltcraft SPL meter (A-weighted).

Experimental setup: Tests. To avoid any prior habituation to test stimuli presented to other group members, subjects were individually tested in an acoustic booth (250 Series Mini Sound Shelter, Industrial Acoustics Company: Inner dimensions in meters: 0.60x0.86x1.68) located inside the larger monkey enclosure. A loudspeaker was installed in the upper part of the acoustic booth and hidden for the monkeys´ view by a piece of fabric. At a lower level inside the booth, 89 cm from the floor and 14 cm away from the wall, a perch was mounted (width: 24 cm), running orthogonally to the direction of the loudspeaker. This ensured that the experimental subject, tested while sitting on it, faced away from the sound source, located approximately 50 cm behind its head. For test playbacks, the minimum registered sound pressure level was 50dB (vs. 29dB for silence) within the acoustic booth.

Experimental Procedure. Monkeys were individually trained, over a period of 8 months, to enter the sound booth voluntarily and to sit calmly on a perch while the doors of the booth remained closed. During this time, one of the experimenters stayed in the sound booth with the individual and rewarded it with food for desirable behaviour. Sounds unrelated to those used as stimuli in the experiment were played to them occasionally. After the subjects were well familiarized to the acoustic isolation from the group and sat calmly, actual experiments took place.

4

The experiment followed a habituation-discrimination paradigm, split into two main phases: (1) the exposure of the entire group to the habituation stimuli and (2) the individual testing using novel stimuli. During the exposure phase, animals were habituated to the stimuli during 2 sessions distributed over 2 days. In each session, the habituation stimuli were played in random order to all individuals in the group simultaneously. The inter-onset-interval between stimuli was 5 seconds, yielding a total of 80 minutes of exposure. This is a longer exposure than what normally used in similar experiments, chosen because long pre-experimental exposure to the stimuli appear to increase the likelihood of efficient recognition (see, for instance, [3]).

The test phase took place over a period of 3 days. Before each test session, a "refresher" exposure (at least 30 minutes) was again broadcasted to all animals. Subsequently, individual monkeys were encouraged to voluntarily enter the acoustic booth using food rewards. Once a monkey was sitting calmly on the perch facing away from the loudspeaker, the door was closed. One of the experimenters (E1), sitting inside the booth, fed insects (*Zophobas morio* and *Galleria mellonella*), a preferred treat, to the subject between trials. E1 continuously listened via over-ear headphones to loud music, whose high frequency range was artificially enhanced in order to mask the stimuli, to prevent any stimulus-specific cueing by E1. A second experimenter, in charge of the playbacks, informed E1 of upcoming trials via an in-ear walkie-talkie system at which point, E1 stopped feeding the monkey for at least 5 sec.

Test 1 followed both habituation phases (testing started within 3 hours from the second habituation). For each individual monkey, and independently from

5

the other subjects, Test 2 followed Test 1. Monkeys took the second test within 48 hours from the first, depending on the individual willingness to participate. When unable to complete an entire session (as soon as an animal showed intention to leave the booth, the door would be opened and the experiment interrupted), test was later resumed (which happened in 3 cases).

Video Coding and Data Analysis. Experiments were digitally recorded at 29.97 frames per second. The digitized video of each experiment (consisting of several trials) was trimmed (iMovie 9.0; Apple Computer). To ensure coding blind to experimental conditions, all original audio tracks were removed and replaced with "auditory placeholders", signalling the beginning and the end of each trial [4].

Three colleagues, experienced in video coding, but not involved in running the experiment, were given instructions to annotate head turns towards the loudspeaker of 45 degrees or more. Since coders annotated entire experiments and not single trials, the number of consistent and violation trials coded by one person was the same, ensuring that results for whole tests could not be due to different coding styles or biased scoring. Before blind-coding all data, a total of 24 trials (equivalent to 15% of the analyzed trials), collected after Test 2 and unused in this study, were coded by subsets of two people to assess inter-individual agreement. This resulted in an average index of concordance [5] of 0.875.

Videos were coded using ELAN [6, 7] and annotations subsequently exported. Custom written Python software automatically extracted the information

6

corresponding to all tests, subjects and coders. Before video coding started, we established two constraints regulating which annotations would be analyzed: only head turns starting (a) after stimulus onset and (b) within 7 seconds from the playback onset, corresponding to four times the duration of the longest tone stimulus (and twice the duration of the longest syllable stimulus, see section Pilot Experiments with Human-spoken Syllables) were extracted using a custom Python script.

Data analysis was performed in SPSS 19 and STATA 11 on the percentage of stimuli per category that elicited a reaction. Parametric tests were only used after testing for normality (Shapiro-Wilk test, n=6 or n=4, all p-values ≥ 0.27) and homoskedasticity (Levene test, n=6 or n=4, all p-values ≥ 0.32) assumptions.

ADDITIONAL RESULTS AND DISCUSSION

Additional Results and Discussion: Ruling Out Simple Heuristics.
Previous artificial grammar learning research has shown how positive results may arise even when the subjects tested do not master the intended pattern [8, 9]. Experimental subjects would hence use a less cognitive-demanding shortcut to provide answers in an operant conditioning setup. This issue can also arise when using a habituation-discrimination paradigm, as in our case. In our experiments, animals are assumed to react always and only to the joint *presence of the first and last elements* if these belong to the same class, although alternative detection strategies might exist. In particular, the animals could accept or reject stimuli by comparing the edges of each stimulus

7

(naively using logical OR or XOR operators) in three different ways: (a) accept all stimuli that either begin or end with a low tone, (b) at a group level, either accept all stimuli that begin with a low tone or accept all stimuli that end with a low tone, or (c) at an individual level, either accept stimuli starting with a low tone or accept all stimuli ending with a low tone.

If monkeys applied strategy (a), consistent and violation stimuli would be all either rejected or accepted, by virtue of either starting with a low tone, ending with a low tone, or both. Paired samples t-tests (see results in the main text), however, do not support this hypothesis: there is a significant difference between consistent and violation stimuli, hence the monkeys pay attention to more than the mere occurrence of one low tone, independently of its location and context.

If monkeys applied strategy (b), the rate of response to violation stimuli missing the first low tone should overall differ from the reaction rate to stimuli missing the last low tone. A paired samples t-test on the percentage responses to those two stimulus categories revealed no significant difference in reactions between LH* and H*L stimuli (Test 1, n=6, t=.54, p=0.611; SPSS). This result does not support the hypothesis that squirrel monkeys have a species-specific bias or group tendency towards attending either at the first or at the last element of the stimuli in order to discriminate between consistent and violation trials.

Finally, if our subjects applied an edge rule at an individual level, each individual would show different reaction rates between violation stimuli missing the first low tone and stimuli missing the last low tone. A Fisher's

8

exact test, however, revealed no association between violation type (beginning vs. end) and reaction (head turn vs. no reaction) at an individual level (Test 1, all p>0.42; calculated using a custom Python script: www.python.org).

It should be noted that tests for hypotheses (b) and (c) do not constitute evidence per se, rather lack of support for a hypothesis. In other words, it is possible that a larger sample size would lead to rejection of the null hypotheses in question. However, the available data supports the hypothesis that our subjects are attending both at the beginning and the end of the stimuli when showing a behavioural response.

Additional Results and Discussion: Test 2. The results of the ANOVA described in the main paper are robust to sampling conditional to novelty. They remained unaltered when the number of elements in each stimulus class was counterbalanced by excluding reactions to the two shortest "consistent" stimuli (ABA and ABBA). We found an effect of stimulus class (2x2 ANOVA, n=4, F=13.5, p=0.035) but no effect of test type (F=0.02, p=0.909) and no interaction (F=0.16, p=0.718). All monkeys took test 1 before test 2. This could create some order effects, such as general decrease in responsiveness over time, habituation to some of the "violation" stimuli, etc. Such order effects, however, could only influence the results either by providing false negatives (failure to detect a cognitive ability nonetheless present in the monkeys) or decreasing the effect size (similar reaction percentages to consistent and violation stimuli). The existence of a significance difference in

9

31

Test 2 only when the previously heard violations are discarded seems to support this conclusion. We are not able, instead, to imagine a scenario by which exposure to violation and consistent stimuli in Test 1 could produce a false positive in Test 2. That is, it seems unlikely that positive results in Test 2 might be completely explained by the test order instead that by monkeys' cognitive skills.

Additional Discussion: Stimulus Novelty and Pitch Transitions. While the specific tones composing the test stimuli are not novel with respect to the frequencies heard during the familiarization phase, the tone combination is completely novel. This is granted by the high number of individual tones contained in a category. For instance, given a test stimulus composed by 5 tones, the probability that it was also broadcasted during habituation by chance is approximately 0.000001. That is, such thing would happen once every one million experiments.

The novelty of the stimuli, crucially, lies in the particular combination of elements belonging to classes, rather than in the novelty of the specific tones within a class. Our design attempts to mirror abstract patterning in language and music. In Western tonal music (rock, pop, classical, etc.), for instance, new exemplars are formed by combining well-established elements (notes) to form novel patterns (melodies, chord progressions, etc.). These melodies, in turn, contribute to re-defining categories (e.g. modulations lead the very same chord to serve radically different functions at two different points in a musical

10

piece; even though the physical characteristics of a chord do not vary over time, its perception does).

Our tests are also robust to the eventuality that the monkeys attended exclusively to pitch transitions between items from different classes rather than the items themselves. Low tones (L) separated by a series of high tones (H) could be perceived and conceptualized as pitch leaps (P=LH or HL) separated by a series of transitions between similar frequencies (S=HH or LL). If so, our statistical analysis would provide inference on the perception of a $PS^{n-1}P$ pattern rather than the LH^nL. However, these two patterns (and their violations) are conceptually identical, leaving our results and conclusions unaltered.

Pilot Experiments with Human-spoken Syllables. Prior to the experiments reported here, the same individual monkeys were habituated and tested in a similar paradigm using human syllables (uttered by a woman for A and a man for B) as categories members. These experiments, matched in stimuli type and length, provided no evidence of discrimination between stimuli types. A Shapiro-Wilk test on the difference between percentage reactions to violations and consistent stimuli (n=6, p=0.035) made us reject the normality assumption, and neither a sign test for related samples (n=6, p=1.0), nor a Wilcoxon Signed Rank test (n=6, p=0.65) showed evidence of any discrimination between stimuli classes. Because non-parametric tests usually have less statistical power than their parametric counterparts, this lack of evidence might be due to the use of a different statistical test, but a parametric

11

test also provided no evidence for discrimination (Paired samples t-test: n=6, p=0.69).

References

1. Wienicke A., Häusler U., Jürgens U. 2001 Auditory frequency discrimination in the squirrel monkey. *J Comp Physiol A* **187**(3), 189-195.

2. Knowlton B.J., Squire L.R. 1994 The information acquired during artificial grammar learning. *J Exp Psychol Learn Mem Cogn* **20**(1), 79.

3. Gebhart A.L., Newport E.L., Aslin R.N. 2009 Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon Bull Rev* **16**(3), 486-490.

4. Ravignani A., Fitch W.T. 2012 Sonification of experimental parameters as a new method for efficient coding of behavior. In Proceedings of Measuring Behavior 2012 (ed. A.J. Spink et al.), Utrecht, The Netherlands. pp. 376-379.

5. Martin P., Bateson P. 1993 *Measuring behaviour: an introductory guide*, Cambridge University Press.

6. Lausberg H., Sloetjes H. 2009 Coding gestural behavior with the NEUROGES-ELAN system. *Beh Res Meth* **41**(3), 841-849.

7. 2013 Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands.

8. ten Cate C., Okanoya K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos Trans R Soc Lond B* **367**(1598), 1984-1994.

12

9.   van Heijningen C.A., de Visser J., Zuidema W., ten Cate C. 2009 Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci USA* **106**(48), 20538-20543.

3 | Data Sonification

Sonification of Experimental Parameters as a New Method for Efficient Coding of Behavior

Andrea Ravignani[1], W. Tecumseh Fitch[2]

Department of Cognitive Biology, University of Vienna, Vienna, Austria.

[1]andrea.ravignani@univie.ac.at, [2]tecumseh.fitch@univie.ac.at

Abstract

Cognitive research is often focused on experimental condition-driven reactions. Ethological studies frequently rely on the observation of naturally occurring specific behaviors. In both cases, subjects are filmed during the study, so that afterwards behaviors can be coded on video. Coding should typically be blind to experimental conditions, but often requires more information than that present on video. We introduce a method for blind-coding of behavioral videos that takes care of both issues via three main innovations. First, of particular significance for playback studies, it allows creation of a "soundtrack" of the study, that is, a track composed of synthesized sounds representing different aspects of the experimental conditions, or other events, over time. Second, it facilitates coding behavior using this audio track, together with the possibly muted original video. This enables coding blindly to conditions as required, but not ignoring other relevant events. Third, our method makes use of freely available, multi-platform software, including scripts we developed.

Keywords. Sonification, blind coding, observational data, scoring behavior, experimental condition, behavioral observation, animal behavior, playback experiment, Python, ELAN.

Introduction

Experimental research in behavior and cognition often follows a typical methodological pattern. While a subject is exposed to several conditions, its reactions are videotaped. Afterwards, the researcher scores these data by blind coding specific behaviors from the video.

We developed a method that allows for an easy and more ergonomic video coding of behaviors and reactions, particularly suited for playback studies. While the person coding the video is busy with her visual task, she may need further information other than that visible in, or audible from, the original recording. Similarly, when a video is to be coded blindly, and its audio track must therefore be silenced, the coder nonetheless needs temporal guidance on what specific behaviors to code and at what point in time. In our approach, such information is broadcasted to the coder by the occurrence of specific sounds, played synchronously with the video.

Methods

The crucial innovation in our approach is to have an audio track where all aspects of the experiment relevant to video coding are mapped to sound (that is to say, "sonified"). Here we outline the step-by-step procedure that can be used to achieve this auditory-aided video coding (see Figure 1). The emphasis is on the methodology and the conceptual approach used, rather than the specific pieces of software employed. Therefore, in the rest of the section, the general description of each step is followed by a short description of the specific software solution we chose. Our choices, however, do not preclude alternative solutions.

1. While filming, the experimenter keeps a log of the events that, not being present on video, she would still like to include at the time of coding. Alternatively, the experimenter may use experimental software that automatically keeps a log with time and type of experimental conditions. For example, on Mac OS X platforms, Playback 3.0 [1] offers such a possibility for playback experiments (see reference [1] for download instructions), and its output are plain txt files. In any case, a log can be thought of as a list of different events whose time of occurrence is specified.

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

376

39
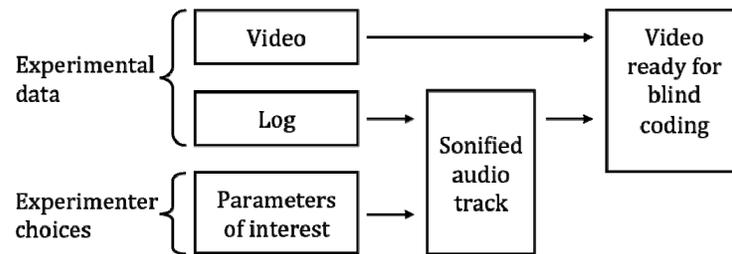
Figure 1. General procedure from experimental data to video annotation. The log of events collected during the study, together with the sounds the researcher decides to associate with particular parameters of interest, gives the sonified soundtrack of the experiment. This gets combined with the video and it is ready to be coded.

2. The experimenter decides what sound to associate with which type of event. Our first choice, dictated by perceptual saliency, has fallen on pure tones that differ by more than one octave in frequency. However, alternative solutions are easily implementable.

3. The log of events is transformed into an audio track, where each type of sound has a different meaning. Therefore the experiment has an "auditory counterpart" associated, spanning its entire duration in real time (see Figure 2). This is accomplished using a Python (Python 2.x, www.python.org) script we wrote, whose source code is available from the first author upon request. Python is a high-level, multi-platform programming language, built-in in Mac OS X and also available for download to Windows and Linux users. In particular, our script only makes use of libraries already included in Python, therefore avoiding additional installations.

   What follows is the pseudocode on which our sonification script is based, enabling interested researchers to create an alternative version of the script using their preferred programming language.

- Create a silent audio file spanning the entire duration of the experiment.

- Create a different sound for each relevant event.

- For each time step in the audio file.

   o If an event requiring sonification is present.

      ▪ Insert the corresponding sound in the audio file.

4. Audio and video are easily synchronized by manually trimming the onset of the first event in the video. For the case of playback experiments, this is equivalent to trimming the video up to the first stimulus onset. Synchronization accuracy depends on the quality and accurateness of the event log (see point 1). In studies requiring high synchronization accuracy, researchers should opt for a computer-based log-keeping system. An upper bound for the precision of the synchronization, on the other hand, is determined by the video frame rate. Trimming can be accomplished using any video editing software. iMovie (included in Mac OS X, our choice for testing the system), Windows Movie Maker (built-in in Windows) and Apple QuickTime (both platforms) are, for instance, uncomplicated, free solutions.

5. Videos can be coded using a wide array of available software. We tested our system using the freely available video annotation software ELAN [2-3], which runs on Windows, Mac OS X and Linux platforms (see reference [2] for download). In ELAN, when simultaneously importing audio and video files, their onsets get automatically aligned. Therefore, once the video has been trimmed (see point 4), the audio-video synchronization is straightforward and automatic.

6. An additional Python script (also available on request) can be used to check whether the behaviors and reactions coded satisfy the conditions originally imposed by the researcher and dictated by the audio

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman   377

40

track. For instance, if behaviors are supposed to be coded within a given time window, the script can trim, discard or leave unaltered the annotations that exceed that time span, depending on the needs of the study. Additionally, when data from more than one coder is present, our script calculates the inter-rater agreement, both as a naïve agreement coefficient (joint probability) and using Cohen's kappa statistic [4].

In the next sections, we outline two possible research contexts where our method is particularly suitable.

Playback Experiments

In a playback experiment, different categories of sounds are played to an infant or a non-human animal. The subject's differential reactions among stimuli types are used to investigate aspects of her auditory cognition, such as discrimination capabilities, individual recognition, generalization within classes, etc.

While coding the video recording of the experiment, the researcher needs to know when a stimulus was played, so to look for reactions to code. However, to avoid biasing the coding, she must be blind to the specific sound played. Therefore, the original audio track is removed from the video and substituted with placeholder sounds. These (pure sine waves at different frequencies) signal when the playback started, when the coder can actually begin coding reactions and when no coding is allowed anymore.

This method is currently used at the University of Vienna (together with ELAN [2-3]) to code behavior in cognitive experiments in several primate species.

Observations of Naturally Occurring Behavior

Researchers in ethology usually focus on naturally occurring behaviors. Either observations are coded in real time or a focal subject (or a group of individuals) is filmed and the behavior coded at a later time. In the second case, several environmental and social phenomena may escape the camera, but are recorded by the researcher at the site. Such events may include, for instance, fights among other members of an animal group, extraneous human interactions, appearance of a predator, etc. Our method offers a way of integrating, at the time of coding, this manually recorded information with the video.

Conclusion

While coding specific behaviors from a video, two types of information are important. On the one hand, there are sounds that were recorded in the original video but may need to be concealed from the blind coder. On the other hand, there may be information not present in the video that is nonetheless essential while coding. The method
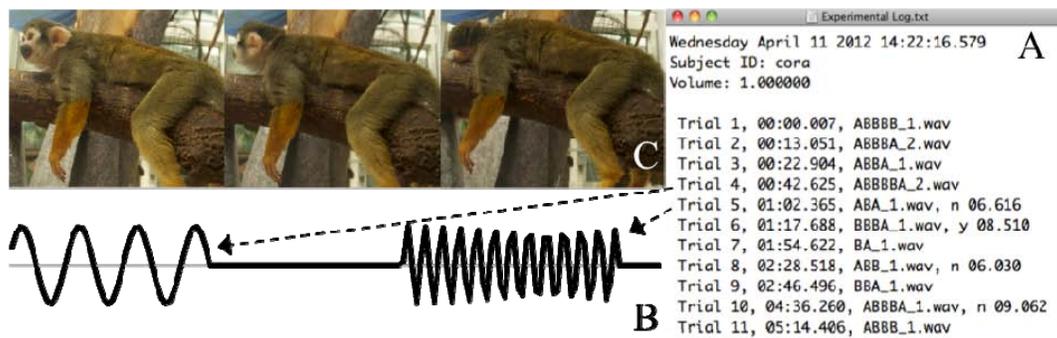
Figure 2. Sketch of the sonification procedure. Using a Python script, each event in the experimental log (A) is transformed into one or more sounds chosen by the researcher. The resulting soundtrack (B) is joined with the original video (C) and these are played simultaneously during behavioral coding. (Squirrel monkey pictures by Markus Boeckle)

we present here offers a simple solution to both issues at once. Information relevant to the experiment is sonified, that is, experiment-relevant events are transformed into sounds. This stream of sounds is then substituted to, or flanked by, the sound track of the original video. This allows coding blind to experimental conditions while keeping track of other events relevant to the study. We provided two practical examples of the many possible applications that this new method has in the study of human and non-human animal behavior and cognition. Finally, we show how the entire procedure, from performing the experiment to annotating behaviors, can be performed using freely available software.

Acknowledgments

References

1. Jones, R., Fitch, T. (2009). Playback 3.0. Copyright 2009. Download requires git (multi-platform and freely available at http://git-scm.com/) and typing the following command: git clone git://playback.git. sourceforge.net/gitroot/playback/playback Accessed September 2011.

2. ELAN Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. http://www.lat-mpi.eu/tools/elan/ Accessed March 2012.

3. Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556-1559.

4. Cohen, J. (1960) A coefficient for agreement for nominal scales. *Education and Psychological Measurement* **20**, 37–46.

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

379

42

4 | Individual Entrainment

Vocal learning, prosody and basal ganglia: Don't underestimate their complexity

Andrea Ravignani[a,*], Mauricio Martins[a,b,*] and W. Tecumseh Fitch[a]

[a]*Department of Cognitive Biology, University of Vienna, A-1090 Vienna, Austria;* [b]*Language Research Laboratory, Lisbon Faculty of Medicine, 1649-028 Lisbon, Portugal.*

* Andrea Ravignani and Mauricio Martins contributed equally to this commentary as joint first authors.

Abstract: Ackermann et al.'s arguments need sharpening and rethinking at both mechanistic and evolutionary levels. First, Ackermann et al.'s evolutionary arguments are inconsistent with recent evidence concerning non-human animal rhythmic abilities. Second, prosodic intonation conveys much more complex linguistic information than mere emotional expression. Finally, human adults' basal ganglia have a considerably wider role in speech modulation than Ackermann et al. surmise.

While Ackermann et al.'s theory is interesting, seems plausible, and may initially appear tempting, it is based on incomplete readings of several literatures. First, it is unclear why some of their arguments should only apply to the specific instances of rhythmic and prosodic control they discuss or why they *fail* to apply in other animal species. Their model assumes that enhancement of in-group cooperation and cohesion was the main driving force for the evolution of speech via the intermediate step where vocal control and rhythm production would serve as chorusing and bonding tools. A key assumption is that speech would produce rhythmic abilities as an evolutionary by-product. This scenario is in line with some empirical observations (for reviews, see (Fitch 2012; Geissmann 2000)) and previous theoretical frameworks for the *origins of music* (Hagen & Bryant 2003; Hagen & Hammerstein 2009; Merker, 2000; Merker et al. 2009). However, when applied to *language*, Ackermann et al.'s evolutionary model does not withstand cross-species validation: Many non-human animals exhibit rhythmic behaviors while lacking speech. Before primate rhythmic abilities can be compared with humans' at all, more evidence regarding flexibility in vocalizations' temporal patterning (Fedurek et al. 2013) and motor synchronization (Hattori et al. 2013) is needed in apes (cf. (Ravignani et al. 2013).

Evidence from non-primate species also seems to undermine Ackermann et al.'s model. Two bird species, both vocal learners, have been shown to entrain to steady pulses (Hasegawa et al. 2011; Patel et al. 2009), supporting Ackermann et al.'s model and Patel's hypothesis, whereby auditory-motor entrainment skills would be evolutionary by-products of vocal learning abilities (Patel 2006). However, recent evidence suggests that vocal learning and rhythmic abilities might be dissociated. Sea lions, unlike seals, show no evidence of vocal learning (Janik & Slater 1997) but nonetheless can reliably synchronize their movements to a range of musical stimuli at different tempi (Cook et al. 2013). Humans and sea lions are both rhythmically skilled, but only humans evolved vocal learning and speech. Therefore, sea lions constitute outliers inconsistent with the prediction of Ackermann et al.'s model. This species evolved cognitive rhythmic abilities, without evolving speech. Invoking additional evolutionary forces and physiological mechanisms thus appears necessary: How can Ackermann et al.'s model be modified to avoid incorrectly predicting vocal learning in rhythmic-skilled species?

Second, Ackermann et al.'s model assumes that prosodic modulation of speech conveys mainly simple motivational-emotional information, and thus that prosody and complex speech production had separate evolutionary histories. But evidence showing a tight connection between prosody and complex linguistic functions argues against this "double pathway" theory. Prosodic contour is influenced by syntactic constituent structure, semantic relations, phonological rhythm, pragmatic considerations, as well as by the length, complexity, and predictability of linguistic material (Wagner & Watson 2010). Furthermore, prosodic cues are used in childhood during acquisition of words (Christophe et al. 2008) and grammatical constructions (Männel et al. 2013), and in adulthood for syntactic processing (Christophe et al. 2008; Kjelgaard & Speer 1999; Langus et al. 2012; Wagner 2010) and word recognition (Cutler et al. 1997).

Contra Ackermann et al., such complex linguistic modulation of prosody seems to be a prerequisite for the acquisition and use of language, and this process is likely to be influenced by cognitive mechanisms specially modified in the human lineage. Comparative research on syntax precursors favors this hypothesis: The ability to assemble sequences of sounds into *hierarchical* patterns might be either human-specific or very poorly developed in other species (Conway & Christiansen 2001; ten Cate & Okanoya 2012). Hence, developmental and comparative evidence point to a more complex cognitive integration of prosody and speech than allowed by the dual-pathway proposal of Ackermann et al. The challenge for Ackermann et al.'s theory is thus to account for the modulation of prosody by human-specific cognitive functions (e.g., syntax), which are clearly not evolutionary homologues of primate emotional vocalizations controlled by the anterior cingulate cortex.

Finally, Ackermann et al. propose an ontogenetic pathway in which: (1) basal ganglia (BG) are important to generate integrated templates of orofacial and laryngeal movements during childhood, but (2) in adulthood can be retrieved from cortical areas because these motor templates become well trained. Later in ontogeny, BG would mostly subserve the modulation of emotional prosody, and not the coordination of speech production. These claims are not supported by currently available empirical data. For instance, Ackermann et al. cite Parkinson's disease (PD) data to support their claims that, in adults, BG lesions only impair emotional prosody. In fact, PD patients with normal

cognitive functioning are more impaired in semantic fluency tasks than in phonetic fluency (Henry & Crawford 2004). Additionally, contra Ackermann et al., BG subserve complex syntactic and semantic processing in adults, with empirical findings consistent across PD (Dominey & Inui 2009; Henry & Crawford 2004; Lewis et al. 1998), BG lesion (Kotz et al. 2003; Teichmann et al. 2008; Ullman et al. 1997), and neuroimaging research (Friederici & Kotz 2003). These data suggest that in adults the BG support multiple functions relevant to spoken language, not just simple emotional prosodic modulation.

Furthermore, contrary to the developmental pathway proposed by Ackermann et al., the acquisition of novel syntactic structures in adults depends on the medial temporal cortex, and the retrieval of syntactic templates *after* thorough learning mostly recruits the BG and perisylvian structures (Ullman 2004). This evidence shows that, contra Ackermann et al., BG are active in the retrieval of over-learned procedures. Ackermann et al. thus need to propose alternative explanations to reconcile child and adult data concerning the function of BG.

In conclusion, to make their model robust, Ackermann et al. must modify and refine their evolutionary and mechanistic explanations, and clarify which assumptions are necessary and which are sufficient for their explanatory framework to hold. Is their model robust enough to stand up to the clear, strong relationship between prosody and complex linguistic functions? How can Ackermann et al.'s model account for the complex functions of BG in adulthood? If in-group cohesion had to be achieved, why was precise vocal control specifically selected for, rather than general non-vocal rhythmic abilities? These and other questions need to be addressed if Ackermann et al.'s model is to become convincing.

Acknowledgements

References

Christophe, A., Millotte, S., Bernal, S. & Lidz, J. (2008) Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1–2):61–75. doi: 10.1177/00238309080510010501 [AR]

Conway, C. M. & Christiansen, M. H. (2001) Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5(12):539–46. [AR]

Cook, P., Rouse, A., Wilson, M. & Reichmuth, C. (2013) A california sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology* 127(4):412–27. [AR]

Cutler, A., Oahan, D. & van Donselaar, W. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40(2):141–201. [AR]

Dominey, P. F. & Inui, T. (2009) Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex* 45(8):1012–18. doi: 10.1016/j.cortex.2009.03.007 [AR]

Fedurek, P., Schel, A. M. & Slocombe, K. E. (2013) The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology* 67(11):1781-1789. [AR]

Fitch, W. T. (2012) The biology and evolution of rhythm: Unraveling a paradox. In: *Language and music as cognitive systems,* ed. P. Rebuschat, M. Rohrmeier, J. A. Hawkins & I. Cross, pp. 73–95. Oxford University Press. [AR]

Friederici, A. D. & Kotz, S. A. (2003) The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* 20:S8–S17. doi: 10.1016/S1053-8119(03)00522-6 [AR]

Geissmann, T. (2000) Gibbon songs and human music from an evolutionary perspective. In: *The origins of music,* ed. N. L. Wallin, B. Merker & S. Brown, pp. 103–23. Massachusetts Institute of Technology. [AR]

Hagen, E. H. & Bryant, G. A. (2003) Music and dance as a coalition signaling system. *Human Nature* 14(1):21–51. [AR]

Hagen, E. H. & Hammerstein, P. (2009) Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Musicae Scientiae* 13(2 suppl): 291–320. [AR]

Hasegawa, A., Okanoya, K., Hasegawa, T. & Seki, Y. (2011) Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports* 1(120):1-8. [AR]

Hattori, Y., Tomonaga, M. & Matsuzawa, T. (2013) Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific reports* 3(1566):1-6. [AR]

Henry, J. D. & Crawford, J. R. (2004) Verbal fluency deficits in Parkinson's disease: A meta-analysis. *Journal of the International Neuropsychological Society* 10:608–22. [AR]

Janik, V. M. & Slater, P. J. B., eds. (1997) *Vocal learning in mammals, vol. 26.* Academic Press. Elsevier Inc. [AR]

Kjelgaard, M. M. & Speer, S. R. (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40:153–94. [AR]

Kotz, S. A., Frisch, S., Cramon, S. Y. & Friederici, A. D. (2003) Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society* 9:1053–60. [AR]

Langus, A., Marchetto, E., Bion, R. A. H. & Nespor, M. (2012) Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language* 66(1):285–306. doi: 10.1016/j.jml.2011.09.004 [AR]

Lewis, F. M., Lapointe, L. L., Murdoch, B. E. & Chenery, H. J. (1998) Language impairment in Parkinson's disease. *Aphasiology* 12(3):193–206. doi: 10.1080/02687039808249446 [AR]

Männel, C., Schipke, C. S. & Friederici, A. D. (2013) The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience* 5:86–94. doi: 10.1016/j.dcn.2013.01.003 [AR]

Merker, B. (2000) Synchronous chorusing and the origins of music. *Musicae Scientiae* 3(1 suppl):59–73. [AR]

Merker, B., Madison, G. & Eckerdal, P. (2009) On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45(1):4–17. [AR]

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

Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I.. (2009) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology* 19(10):827–30. [AR]

Ravignani, A., Olivera, V. M., Gingras, B., Hofer, R., Hernández, C. R., Sonnweber, R.-S. & Fitch, W. T. (2013) Primate drum kit: A system for studying acoustic pattern production by non-human primates using acceleration and strain sensors. *Sensors* 13(8):9790–9820. [AR]

Teichmann, M., Gaura, V., Demonet, J. F., Supiot, F., Delliaux, M., Verny, C., Renou, P., Remy, P. & Bachoud-Levi, A. C. (2008) Language processing within the striatum: Evidence from a pet correlation study in huntington's disease. *Brain* 131(4):1046–56. doi: 10.1093/brain/awn036 [AR]

ten Cate, C. & Okanoya, K. (2012) Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1598):1984–94. [AR]

Ullman, M. T. (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1–2):231–70. [AR]

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J. & Pinker, S. (1997) A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience* 9(2):266–76. [AR]

Wagner, M. (2010) Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory* 28(1):183–237. doi: 10.1007/s11049-009-9086-0 [AR]

Wagner, M. & Watson, D. G. (2010) Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7–9):905–45. doi: 10.1080/01690961003589492 [AR]

5 | Ape Drumming

Article

Primate Drum Kit: A System for Studying Acoustic Pattern Production by Non-Human Primates Using Acceleration and Strain Sensors

Andrea Ravignani [1,*]**, Vicente Matellán Olivera** [2]**, Bruno Gingras** [1]**, Riccardo Hofer** [1]**, Carlos Rodríguez Hernández** [2]**, Ruth-Sophie Sonnweber** [1] **and W. Tecumseh Fitch** [1]

[1] Department of Cognitive Biology, University of Vienna, Althanstrasse, 14, Vienna A-1090, Austria;
E-Mails: bruno.gingras@univie.ac.at (B.G.); riccardo.hofer@univie.ac.at (R.H.);
ruth-sophie.sonnweber@univie.ac.at (R.S.); tecumseh.fitch@univie.ac.at (T.F.)

[2] Escuela de Ingeniería Industrial e Informática, Universidad de León, 24071 León, Spain;
E-Mails: vicente.matellan@unileon.es (V.M.); carlosrhrh@hotmail.com (C.R.)

* Author to whom correspondence should be addressed: E-Mail: andrea.ravignani@univie.ac.at;
Tel.: +43-1-4277-76161; Fax: +43-1-4277-9761.

Abstract: The possibility of achieving experimentally controlled, non-vocal acoustic production in non-human primates is a key step to enable the testing of a number of hypotheses on primate behavior and cognition. However, no device or solution is currently available, with the use of sensors in non-human animals being almost exclusively devoted to applications in food industry and animal surveillance. Specifically, no device exists which simultaneously allows: (i) spontaneous production of sound or music by non-human animals via object manipulation, (ii) systematical recording of data sensed from these movements, (iii) the possibility to alter the acoustic feedback properties of the object using remote control. We present two prototypes we developed for application with chimpanzees (*Pan troglodytes*) which, while fulfilling the aforementioned requirements, allow to arbitrarily associate sounds to physical object movements. The prototypes differ in sensing technology, costs, intended use and construction requirements. One prototype uses four piezoelectric elements embedded between layers of Plexiglas and foam. Strain data is sent to a computer running Python through an Arduino board. A second prototype consists in a modified Wii Remote contained in a gum toy. Acceleration data is sent via Bluetooth to a

computer running Max/MSP. We successfully pilot tested the first device with a group of chimpanzees. We foresee using these devices for a range of cognitive experiments.

1. Introduction

1.1. The Need for Sensing Technologies in Cognitive Research

Humans, from their birth onwards, are characterized by a strong tendency to explore the space surrounding them [1]. An important part of this exploratory behavior consists in the feedback obtained, generally of acoustic and tactile nature [1,2]. Recently, increase in computational power and low cost sensing devices encouraged the construction of projects linking movement and sound: *movement sensing* can trigger *auditory feedback* in several possible ways [3–6]. Especially in the fields of human cognitive science and artistic performance, scientists and artists have pioneered an array of approaches differing in sensing technologies and purposes [5,7,8]. The accelerometer contained in the Nintendo Wii Remote or Microsoft Kinect's infrared sensors, for instance, can provide a cheap, consumer-level sensing option [6–10]. Freely downloadable libraries are constantly developed to interface these sensors with computers [9]. Finally, visual programming languages like Max/MSP are an accessible possibility for project developers with little or no programming experience [4,5,11].

In parallel, sensing in the animal kingdom has exclusively focused on applications related to food industry and measuring animal behavior [12–14]. The emerging field of animal cognition, however, requires more advanced sensing possibilities than what is currently implemented. An increasing number of laboratories are training non-human animals to the use of touch screens in order to easily collect the animals' responses [15–18]. However, cognitive scientists working with animals are becoming more and more interested in the *production abilities* of their study species. Inspired by the cognitive gap between perception and production in humans, researchers wonder whether a similar divide can be found in other animal species. Non-human primates are among the species most likely to equal many human cognitive abilities. Chimpanzees, our closest living relatives, have already shown human-like skills in several cognitive domains [17,19]. Among chimpanzees, high-ranking males are known for their dominance displays, often consisting in striking objects in order to generate loud, prominent sounds [20]. Similarly, chimpanzees' playful behavior may involve sound production.

Chimpanzees have been reported to perceive some features of music [21], though whether they would be able to actively produce musical patterns is a topic of debate and conjectures [22]. Most evidence on rhythmic production in chimpanzees is based on observational studies and anecdotes, and no well-controlled data recording has been accomplished due, among others, to the absence of adequate technological resources [20,23]. Similarly, multimodal entrainment to a musical beat relies on an auditory-motor brain circuit and is thought to be only present in humans, a few mammalian species and some bird species [24]. Rigorously testing the presence of rhythmic abilities in

chimpanzees will be fundamental to understand what is unique in human cognition [23,25]. This, in turn, can only be achieved with adequate sensing and computational tools.

1.2. State of the Art

Our prototypes are the first attempt at bridging two strands of research that have been, to our knowledge, completely disconnected until now. On the one hand, electronic sensory devices are used in humans for scientific and expressive purposes, be it in cognitive research [7], data analysis [26] or music performance [27]. On the other hand, sensors are broadly used in non-human animals, though almost exclusively to monitor behavioral patterns and welfare [12–14].

Over the last decades, an impressive amount of research has been devoted to the sonification of data sets or sensory readings collected in real time [26]. Mapping sensory readings to sounds can have a number of applications, such as accurate tracking of vital functions during surgery [28], augmented human interaction with objects [29], and facilitation in learning new movements [2,4]. In particular, sensing for sound manipulation purposes has proven critical in developing new musical interfaces and means of expression [5,27,30,31]. Sensory technologies and computational power have enabled the development of virtual musical instruments and augmented reality environments [6,10,11,32]. Unfortunately, almost all auditory interaction technologies developed until now have focused on humans as potential users. In some cases, one could use human-specific sensors for animal purposes [19]. However, usability and species-appropriateness of the technologies used should be taken into account [33–36]: inappropriate interfaces could just be unused by, or not endure the strength of, the particular animal species.

Most technologies and sensory networks developed until now for animals have been concerned with appropriateness and species-specificity of the devices. One main area of this research deals with monitoring behavior of groups of animals [37–41]. Sensor networks are also used to monitor health and welfare in non-human animals using readings of body temperature and other physiological variables [42–47]. Quite recently, however, researchers in human-computer interaction have started to discuss the importance of developing paradigms for (non-human) animal-computer interaction [34–36]. Specific interfaces have been built so that cows [48], hamsters [49] and dogs [50,51] could interact with a computer, either aurally or visually. Apart from these few cases, the field of animal-computer interaction is just beginning to develop, and promises contributions to sustainability, cognition and adaptability of technologies to specific needs of some human groups, such as pre-linguistic infants or elderly adults [36]. In particular, the field of animal cognition has proved its strong interest in developing technological testing paradigms adapted to particular species and tasks [15,16,18,52,53].

1.3. Filling a Technological Gap

This paper describes two innovative prototypes we specifically developed for sensing movements in chimpanzees and mapping them to specific sounds in real time [23]. To our knowledge, this is the first attempt at building this type of technological tool specifically designed for primate research.

The devices we present here fill the aforementioned technological gap in animal cognition by taking advantage of chimpanzees' spontaneous behavior. We developed two prototypes whose manipulation by chimpanzees produces sounds, which can be altered depending on the specific experiment taking

place. Our prototypes rely on two main principles: *movement sensing* and *sonification*. Acceleration and strain sensors are embedded into manipulable objects. A computer receives the measurements, processes the data and sonifies it. According to Kramer *et al.*, "Sonification is defined as the use of nonspeech audio to convey information. More specifically, sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [54]. Combining the sensing with the sonification principle enables a real time mapping between movements and sounds. Therefore, one can alter the physical-acoustic characteristics of an object by assigning particular, experimentally manipulated sounds to its displacement.

This paper is organized as follows. We explain the need of animal sonification devices for cognitive research. We give an overview of the current technologies available and mention a number of desiderata such devices should have. We illustrate the overall design features of our prototypes and then describe each of them in detail.

2. Description of the Problem

2.1. Overview

The use of sensors in non-human animals is almost exclusively devoted to livestock and animal surveillance applications. The constantly growing field of animal cognition, however, requires ever finer and more specific sensing possibilities. In particular, a number of important hypotheses in primate cognition deal with the connection between movement and auditory perception [22,23,25]. In order to show one of these, rhythmic production abilities in chimpanzees, we developed two prototype devices specifically adapted to this species (Figure 1) [23,55].

Figure 1. Chimpanzee manipulating a Kong Toy, constituting the outer shell of one of our prototypes.
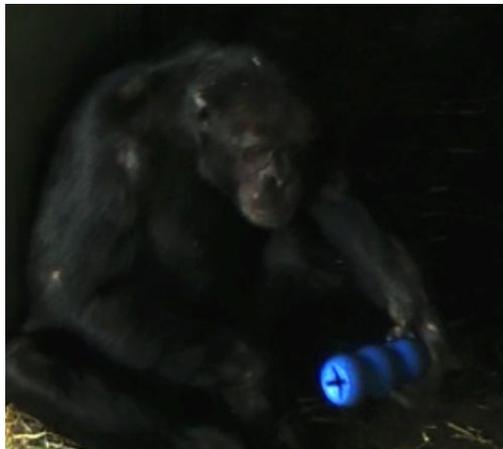
Each prototype presented here is composed of a sensing and a feedback unit, and allows to arbitrarily associate sounds to physical movements. The sensing unit sends acceleration or strain data to a computer, which converts them into sound and plays it in real time. Our prototypes, apart from constituting a first attempt at animal-computer cognitive interaction, will allow rigorous testing of a

number of hypotheses in primate cognition [22,23,25]. Moreover, we plan to make the software freely available, so that interested researchers can benefit from our findings.

2.2. Desiderata and Constraints

The devices we developed satisfy a number of desiderata, plus a crucial tradeoff between ergonomics relative to chimpanzee behavior, and efficiency of the device when used to conduct scientific experiments (see Table 1).

More specifically, desiderata and constraints imposed on the construction of the device affect both the sensing and the computation components of the project. The sensing part has to be resistant, modular, low-voltage, weather resistant, hence making it suitable for safe primate work in general. Moreover, it should be interesting for the primate, possibly inexpensive and easy to connect and configure. The software part requires quick data processing; therefore the computations performed must be moderate in number and nature, so as to limit the computational load.

Table 1. The first column shows critical requirements and whether these are satisfied by objects/surfaces naturally present in the animals' enclosure (second column), commercially available electronic drum sets (third column) and our prototypes (last column).

Requirements	Object/wall in enclosure	Commercial electronic drums	Our prototypes
Sound production when manipulated	✓	✓	✓
Precise data recording	✗	✓	✓
Physical resistance	✓	✗	✓
Weather resistance	✓	✗	✓
Low-voltage	✓	✓	✓
Modular	✗	✗	✓
Spurring interaction	✓	✗	✓

2.3. Requirements: Sensing

Chimpanzees are capable of exerting strong force by hitting and tossing objects. Their muscular strength has been reported to be several times that of a human being [56]. Chimpanzees are supposed to have free access to the device and expected to vigorously manipulate it. Hence, the sensing part of the device has to be resistant to strain and extreme shocks while remaining sensitive to the physical variations of interest. It may be that the animal succeeds nonetheless in breaking the device. Bearing in mind this possibility, the voltage, which the animal may come in contact with, should be kept to a bare minimum. This would avoid eventual electrocution of the animal and short-circuiting the electronics in case of moist environment.

In the eventuality of damage or strong pulling, the device needs to be modular. If a part of the device is damaged or removed, it is desirable that only the damaged component be replaced. Similarly, if the pulling force exerted on the sensor is strong enough for the entire device to be carried away, only a module should detach, leaving the main part of the device undamaged. Considering all these eventualities, the modules composing the device need to be inexpensive and easy to find and replace. Damage and destruction of parts are concrete possibilities.

Chimpanzees are usually housed with access to outdoor enclosures. The device has to withstand, in principle, any kind of meteorological condition. This means resistance both to long periods of direct sunlight and heat, and to rain and cold weather. Other desiderata concern the ease of preparing the device before any measure can be taken or an experiment run. The scientist should be able to easily and quickly activate the device, connect it to a computer and start recording data.

Finally, in order to obtain any data at all, the device has to incite the primate to manipulation and play. Hence its physical properties need to be such that, before receiving any auditory feedback provoked by a manipulation, the chimpanzees will be spurred to approach the device and play with it.

2.4. Requirements: Processing

The software should handle a possibly continuous stream of data. Critically, the processing part of the device needs to concurrently handle the storage, processing and playback of the very same signal. Moreover, the mapping of physical variations in the sensor to sounds must be flexible and easily adjustable. There are types of sounds that chimpanzees are most interested in and which can be used to induce interest in manipulating the object [21,57]. Similarly, testing specific hypotheses requires particular sounds and inter-onset intervals. Therefore the researcher needs to be able to adjust the acoustic properties of the feedback as fast as possible. Finally, object manipulation and auditory feedback should be perceived as causally related. Therefore, all computations must occur as close as possible to real time.

2.5. Requirements: Usability and Experience

The device should also satisfy a number of requirements from the chimpanzee's perspective as a user. From the point of view of *usability*, the movement to sound mapping should be relatively straightforward, spurring the animal to interact. Hence, typical movements chimpanzees may perform must be accounted for, such as the strong hits during pant-hoot displays or finger poking while exploring objects. The amount of learning needed to use the device should be minimal. This will likely be granted by chimpanzees' neophilia and exploratory attitude [58].

Regarding the chimpanzees' *experience*, the device should be physically reactive and acoustically satisfactory. If chimpanzees push the device, it should bounce back, hence spurring the animals to push it once again. When chimpanzees are displaying and hitting on it, it should respond with loud sounds.

Crucially, using the device should grant a sense of satisfaction and reward by itself, to ensure that the chimpanzees will return to it over time. In a number of species vocal communication correlates with the regulation of the dopaminergic system, which is in turn connected with reward and satisfaction [59,60]. In different bird species the perception and production of songs are tightly linked with reward systems [61]. Similarly, when humans listen to music reward pathways are active [62]. To our knowledge, no research has been done on sound preference in chimpanzees. Hence, the prototype should be able to play a range of sounds, so that it is possible to heuristically adjust the movement-sound mapping to chimpanzees' preferences. This way, sonic interaction will be rewarding in itself and possibly serve as an enrichment for captive animals.

2.6. Experimental Requirements

From the point of view of the experimenter, two requirements seem particularly important. The software should be flexible and expandable, so that parameters can be changed and new modules and functionalities added, depending on the particular use envisaged. Moreover, the entire apparatus should be easy to use, so that even researchers with limited technical experience can run an experimental session.

3. General Design Features

The two prototype devices we developed differ in size, connectivity and sensing technology. Our *wired prototype* consists in piezoelectric sensors embedded in a parallelepiped, connected to a computer via an Arduino board (Figure 2). Our *wireless device* is made of a hollow gum toy containing a videogame controller (Wii Remote [9], www.nintendo.com), which is connected to a Mac computer via Bluetooth (Figure 6). Each prototype is described in detail in the next two sections. Here we outline their general, common features.

Figure 2. The wired prototype. The rightmost part can be mounted on a chimpanzee enclosure. The animals come in contact with the leftmost side.

3.1. Data Collection

The general concept consists in embedding a sensor in a manipulable object. The object is designed to incite chimpanzees to interaction and play [63]. The physical characteristics and design of the sensory part are dictated by scrutinizing videos of chimpanzees playing with objects, so as to maximize the chances of interaction with our prototypes [55]. The sensors in the object record physical variations (strain or motion). Depending on the prototype, sensory data can go through a phase of preprocessing before reaching the computer. Preprocessing is done using an Arduino board (http://www.arduino.cc/), which, apart from the analog/digital conversion, is in charge of basic filtering. Once the data reaches the computer, processing is accomplished using different pieces of software. Depending on the sensor and prototype, data is elaborated using patches written in Max/MSP

(http://cycling74.com/products/max/) or scripts written in the multi-platform, freely available programming language Python (www.python.org).

3.2. Data Processing

The software processing part is responsible for four key tasks: filtering the data, extracting parameters relevant to sound production, logging specific occurrences or variations of the parameters, and playing sounds corresponding to critical parameters' states (as specified by the user).

Our basic *data filtering* method simply discards sensory readings below a specified threshold. This simple system allows nonetheless for a rough, though efficient, data classification for a low computational cost. Parameters of interest are successively extracted through concurrent computation of thresholds on sensor input and time.

The *time thresholding* allows for refractory periods in outputting the parameter. When an activation threshold is crossed and the system produces a sound, the system waits for a specified number of milliseconds before resuming its activity. This entails that, once a sound has been produced by the system, sensory readings can trigger new sounds only after a given time period. Such design choice makes it possible that only one sound gets played each time the object is manipulated, therefore avoiding misleading repetitions.

Logging variations in the parameters of interest is essential. No matter if the interactions with the device are observed or videotaped, a precise log containing the exact time and magnitude of the interaction is crucial in any scientific study. Therefore, both prototypes feature the possibility of saving data to a text file in CSV (comma separated value) format. This can be later read in Microsoft Excel or parsed by any programming language in order to extract meaningful statistics. Moreover, if the interactions are videotaped, information in the log can be later integrated with the video.

Finally, *real-time playback* is one of the key features of our prototypes. Our software is extremely flexible, when it comes both to configuration and expansion. Thresholds on parameters and settings can be altered to vary the sensitivity of the device and the properties of the sounds played. Additionally, the mapping between raw data, parameters and sound output can be changed at will by modifying few lines of code or Max objects. Once the computer has produced the sound, the signal is transmitted to an external loudspeaker and one manipulation-auditory feedback interaction is completed.

4. Wired Prototype

This prototype was designed to sonify a chimpanzee's common behavior, namely hitting on walls or objects using either the upper or lower limbs. This prototype can be hence used as a "sonically augmented" substitute for the usual resonant surface [29]. The ricochet property of this device (see *Sensory Module*) satisfies chimpanzees' natural behavioral predispositions to hit and push objects in order to generate sound [20]. Additionally, its acoustic feedback properties can be artificially manipulated, so to sound louder than the usual surfaces chimpanzees usually hit.

The Wired device is composed of a sensory module, a custom produced board, an Arduino board and a computer with a USB interface. This device is supposed to be mounted vertically, resting towards a wall or bars of chimpanzees' enclosures.

This prototype has a number of valuable features. Most prominently, it is built with inexpensive and relatively easy to find components. Assembling the device requires only moderate craftsmanship and electronic skills and the use of a few tools (drill, saw, *etc.*). This prototype is extremely safe concerning possibilities of electrocution. No electricity is required to operate its sensing module: the only electricity present is the one produced by the piezoelectric sensors under strain and therefore has negligible voltage.

While building this device and programming its software, we deemed important for it to be able to discriminate between types of mechanical stress. There are three main ways in which an individual can interact with this prototype: by simply touching it, by tapping on it, and finally by pushing it (until provoking the displacement of the upper Plexiglas layer). We decided that pushing was the only type of strain to which sounds would be associated. This discrimination was achieved both by selecting piezoelectric sensors with specific electrical characteristics and real-time data filtering with our software (see Sections 4.1 and 4.5).

4.1. Wired Sensory Module

The sensory module is a parallelepiped with dimensions 420 mm × 300 mm × 76 mm (volume: 9.576 L). The parallelepiped is made of two layers of Plexiglas (dimensions: 420 mm × 300 mm × 8 mm each) encasing three layers of solid foam (dimensions: 420 mm × 300 mm × 20 mm each). 40 mm away from each corner, four screws are passed through the Plexiglas and foam, and fixed with a die at one of their ends. Longer screws enable the parallelepiped to be fixed to a wall or any vertical surface required for acquiring data. The screws serve the function of holding the module together and, combined with the foam, confer it a ricochet property once the Plexiglas is pushed. It is essential for the module to rebound once pressed, so that the chimpanzee will be spurred to further explore and push it.

Four piezoelectric sensors are installed on the inner side of one of the Plexiglas layers with their plain side in direct contact with the Plexiglas surface. The location of each sensor is established so to induce a centroidal Voronoi tessellation on the Plexiglas surface [64]. This grants an optimal position of the piezoelectric sensors relative to the amount of sensors used and the data acquired. More specifically, each sensor lies on the center of mass of one of four tiles, thus minimizing the sum of distances between any point on the tile and the sensor.

The model of piezoelectric sensor we use was chosen heuristically after comparing and benchmarking a number of heterogeneous alternatives. Sensors were installed in the unit and exposed to three main strain categories: *touch*, *tap* and *thrust*. Among all piezoelectric elements tested, the Multicomp ABT-441-RC (Outer diameter: 27 mm, Resonant frequency: 4.2 kHz, leaded) and the Multicomp ABT-448-90-RC (Outer diameter: 35 mm, Resonant frequency: 2.9 kHz, leaded) proved to be the most appropriate for our purposes. In fact, using the language Processing (www.processing.org) to graph different parameters of each sensor under each condition, the Multicomp sensors, unlike their alternatives, showed categorically different responses depending on the type of strain. Both Multicomp elements, in fact, registered strain maxima that were higher for thrust than for tap, and higher for tap than for touch. This feature made them perfect candidates for filtering low strain manipulations by imposing a software threshold on strain maxima. All other things being equal given our requirements, we opted for employing the ABT-441-RC, its cost being about half of the ABT-448-90-RC.

4.2. Connections

Each piezoelectric sensor is connected through two unipolar cables (length: 1250 mm, outer diameter: 1.5 mm) to a custom-made board (see Figure 3 for schematics). Apart from the 8 inputs from the sensors (4 positives and 4 grounds), the board features resistors (resistance 1 MOhm), one for each piezo. Moreover, an LED (light-emitting diode) is assigned to each sensor, making it possible to check its operational state.

The custom board maps the positive pole of each piezoelectric sensor to an analog input slot on an Arduino board. All the ground signals collected on the custom board are mapped to a ground input on the Arduino. Similarly, each LED on the board connects to a different digital output on the Arduino at one end and a common ground at the other end.

Figure 3. Schematics of the custom Arduino board built for the wired prototype.
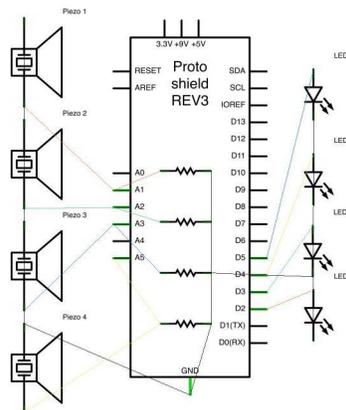
4.3. Arduino: Features

This prototype makes use of an Arduino board for several purposes. First of all, a form of analog to digital conversion is required before sending data to a computer. The Arduino board accomplishes this by obtaining a reading of the four piezoelectric sensors. If readings are above a given threshold data are processed further, otherwise they are ignored. An additional Arduino script enables to light a LED indicating that readings above a given threshold were received from a particular piezoelectric element. This ensures the possibility of checking the correct functioning of all sensors. This is particularly important for the flawless functioning of the device, as a failure in one of the sensors could pass unnoticed and be masked by the readings of the remaining sensors. Finally, the Arduino board is in charge of data preprocessing. That is, part of the integer computations required by the device are distributed and performed by the Arduino.

The particular model of Arduino used in this project is an Arduino Uno, whose clock frequency is 16 Mhz. The baud rate, equal to 28,800, was established heuristically in order to find a good tradeoff between rate of serial transmission and possible delay due to overload. As the numerical values sent by the Arduino rarely exceed 2500, the chosen baud rate grants a lower bound on transmission rate of 1 sample per millisecond, which we deem adequate for our purposes.

4.4. Evaluation and Calibration of the Sensors

Our wired prototype sonifies strain provoked by sudden thrust movements. Choices on the combination of materials used (Plexiglas, foam and four piezoelectric elements) and the movements to be sensed by the piezos needed to be adapted to our particular animal sensing context. It was therefore necessary to calibrate the device. Every sensor's voltage is represented in the Arduino as an integer between 0 and 1023. Given that the Arduino adds up the readings received by the four sensors, the sum of all readings can theoretically be between 0 and 4092. As the simple gentle touching of the device produces readings up to 500, during calibration we imposed this number as the threshold below which Arduino would ignore readings.

The sensory module was put horizontally on the floor. A calibration weight of 5 kg was dropped on the module from different heights and the resulting data was recorded by a computer and further analyzed. The weight was dropped at heights of 20, 40, 60, 80, 100, 120 140, 160 and 180 cm, 5 times for each height, giving a total of 45 trials.

Every impact produced several actual readings plus those provoked by the rebound. The data was hence filtered to discriminate between trials and rebounds following a trial using a custom Python script. The script retained only readings which satisfied the following conditions on onsets and offsets. As trials were separated by at least 2 s their onset could be easily discriminated. Trial offsets were established by discarding all readings' sequences separated by more than 500 ms from the previous one. Hence, 63 out of a total of 855 readings were automatically dropped from the data set because they clearly corresponded to ricochet stress, being separated by more than 500 ms from the previous reading. The average length of each trial was 17.6 samples (*SD* = 4.87).

Our algorithms deal only with a few samples during each interaction, as waiting for a large number of samples would provoke a delay in playback. Hence we further analyzed only up to the first 10 samples per trial.

The impact force of the weight on the module is approximately linear in the height of fall. Similarly, the current produced by one piezoelectric element is linear in the force of strain. Therefore, we hypothesized the sum of the readings from the four sensors to be approximately linear in the distance of fall. We therefore averaged within-trial readings and entered them in a linear regression model (statistical analysis was done in Stata 11.0). The model was significant [n = 45, F(1,43) = 57.35, p < 0.001] and could explain about half of the variance (Adjusted R-squared = 0.56). The fall height could significantly predict the sum of readings (t = 7.57, p < 0.001). The intercept was also statistically significant (t = 27.09, p < 0.001). The equation for the resulting strain is:

$$r = 1208.29 + 3.00 * d + e \tag{1}$$

where r is the average reading, d is the fall distance in cm and e is the error. This result confirms our hypothesis about the existence of a linear relation between impact force/fall height and Arduino's readings (see Figure 4).

Figure 4. Each data point represents a sensor reading on the y-axis dependent on different levels of potential energy (just before the object starts falling) or, equivalently, kinetic energy (just before the impact) on the x-axis. Energy is plotted as multiples of acceleration of gravity g, and ranges between 1 g and 9 g. The line corresponds to the data's best fit using a linear regression model. The grey area is the 95% confidence interval.
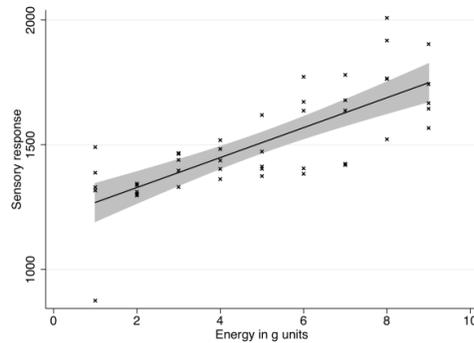
As real time playback is an important issue in our device, we decided to distribute computations between Python and Arduino. While Python is in charge of playback and logging, Arduino performs most of the threshold checks. To discriminate between *thrust* and *hit* types of strain, the Arduino implements a system of two thresholds separated by a delay. One of the main design challenges of this prototype was to make it selectively responsive to strain forces that would displace the outer Plexiglas inwards (*thrust*), as opposed to superficial strikes (*hit*). In order to systematically discriminate hits from thrusts using only the sum of the sensory readings, we systematically analyzed the times series of stress over time under the two conditions. The Arduino was set to send only samples above 1000 and a maximum of one sample every 5 ms. With a custom written Python script, we analyzed 59 repetitions of "hit" interactions and 59 repetitions of "thrust" interactions over time. For each type, the time series of each repetition and condition were aligned by onset and averaged over a time window of 5 ms (corresponding to the maximum time-resolution allowed by our Arduino settings). The resulting means and confidence intervals for both series are displayed in Figure 5.

Figure 5. Mean and confidence interval of the time series of readings (y-axis) provoked by hit (superficial strikes, dashed lines) and thrust (strikes displacing the surface, unbroken lines) strain.
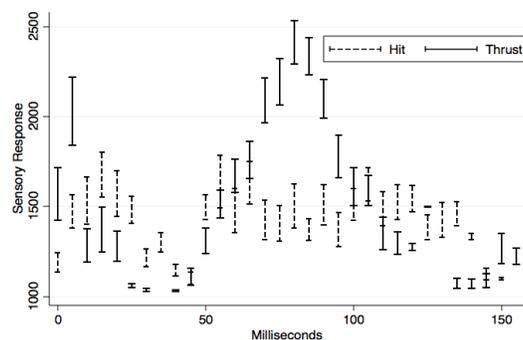
Noticeable differences between series are (i) within the first 40 ms, when thrust is first associated with higher sensory readings, then with lower ones than hit, and (ii) between 70 and 95 ms, when readings associated with thrust are well above those recorded in the "hit" condition. In the first case, a local maximum for the "thrust" condition is reached at 5 ms. Moreover, at 5 ms the average reading for thrust is significantly higher than for hit (t-test for paired independent samples adjusted for unequal variances: $t = -3.61$, $p < 0.01$, Satterthwaite's degrees of freedom = 48.15).

To summarize, we found that a reliable way of distinguishing thrust from hit strain is to (a) further process all readings above a sensory response of 1000, (b) among these, discard all movements that have not exceeded a 1765 threshold after 5 ms (corresponding to the lower 95% confidence interval of the "Thrust" condition).

Taking this into account, we programmed the Arduino as follows. At every iteration, Arduino computes the sum of four concurrent readings and checks whether the current reading exceeds a given threshold (an "if" statement probing an inequality). If the current reading is large enough, Arduino waits 5 ms, reads again the four sensors and computes their sum. This second sum is then checked against a second threshold and is sent through the serial port only if it is larger than this threshold. After that, a refractory period of 300 ms before any further reading is imposed for two main reasons. First, it is advisable not to overload the serial port buffer. Second, stopping the sampling and transmission to Python is a straightforward way to avoid having one single movement elicit multiple sounds.

The computations performed in the Arduino can be seen as a form of filtering unimportant data and noise. This way, irrelevant data is not sent to the computer and processed further. Moreover, having Arduino perform these simple integer calculations constitutes a form of distributing tasks and having the computer available for further data handling and audio playback.

4.5. Python Software

The Arduino is connected to a computer through a USB interface. The computer runs a light Python script, composed of 7 functions spanning approximately 100 lines in total. The script is responsible for performing computations on the data, storing and sonifying them, that is, mapping them to particular sounds.

The Python script uses some internal and external modules. The datetime module is used to generate timestamps while the serial module is employed to communicate serially with the Arduino. The Pygame (www.pygame.org) external module is used to play sounds. First the script calls two Pygame *mixer* functions [mixer.init( ) and mixer.Sound( )] from the *Pygame* library to initialize the mixer and load the chosen sound. The script iterates through all the computer serial ports looking for an Arduino board. If an Arduino is found, it connects to it and starts listening with a 10 ms timeout. An important part of the script is the logging of parameters, sensory values and playback times. After establishing a serial connection, the script generates a comma separated values (csv) file and writes headers, so that all relevant sensory information can be recorded. Additionally, every time the thresholds (Table 2) and auditory parameters are modified and the script is restarted, a new header with the updated parameter information is written in a log file.

67

If contiguous sensor readings exceed critical thresholds (see previous section), the Arduino sends an integer consisting of the sum of all readings. The Python script receives and parses the serial input, ascertaining that it is a well-formed integer. This takes care of possible transmission errors, as non-integers are discarded. If a valid integer value is received, the script calls a *mixer* function [play( )] to play a pre-loaded sound. As soon as a sound is played, playback time [generated using a datetime( ) timestamp], together with volume and/or sensory reading are appended to a Python list. Once an experimental session is ended using a keystroke combination, the list is written to the csv log file.

The signal is sent from the computer soundboard to an external self-powered loudspeaker, so that the chimpanzee can hear the auditory feedback to its action. The loudspeaker is supposed to be placed outside the chimpanzees' enclosure, though directly behind the sensory module (to the extent possible, depending on regulations and conditions dictated by the animal facility).

Table 2. Parameters of the prototypes that can be adjusted by the user. Both prototypes have a customizable time threshold, *i.e.*, a period during which sensor readings cease. The wired prototype has an adjustable strain threshold for its activation and a second threshold, above which a sound is triggered. The wireless prototype features an adjustable threshold on the variation in acceleration, above which a sound is played.

	Time threshold:	Activation threshold on:	Sound production threshold on:
Wired	Present	Sum of readings	Sum of readings
Wireless	Present	Difference in acceleration	Coincides with the "Activation"

4.6. Design Choices and Early Prototypes

Although the hardware and software solutions presented above may appear arbitrary, it should be noted that, in both cases, we opted for these solutions after testing different prototypes and comparing performances. The sensing module went through three different early prototypes, which featured diverse sensors contained in the same Plexiglas-foam structure presented here.

Wired prototype 1 simply featured a consumer-level piezoelectric element. This solution was quickly abandoned, as the sensory module could not systematically detect the type of stress we intended to measure.

Wired prototype 2 featured a 3-axes accelerometer. Accelerations corresponding to different stress types were compared and plotted over time. Even though thrusting and knocking elicited similar acceleration peaks in the data, their statistical properties over time markedly differed. In particular, the variance was distinctively higher for knocking stress, making this quantity a good candidate for discriminating between stress types. While developing the software, however, we found that discrimination on the basis of variance required a time window that was too wide for our solution to give auditory feedback in due time. The possibility of using an acceleration sensor was retained and piloted with a different method of detecting the nature of incoming stress.

With wired prototype 3 we piloted a hybrid between acceleration and strain sensors. Both a 3-axes accelerometer and a piezoelectric element were connected to the Arduino and their data sent to Python. As detailed above, the accelerometer was too sensitive for our purposes, sensing a number of vibrations provoked by simply knocking on the Plexiglas surface. Combining both readings would

allow us to ignore acceleration data when strain data corresponding to knocking was present. Acceleration readings would otherwise be used to calculate the actual speed of displacement of the Plexiglas surface. Unfortunately, good data could be obtained with this method only at the cost of extensive computation, undermining one of the main requirements of the tool, namely speed.

5. Wireless Prototype

This prototype was designed to sonify another common behavior found in non-human primates, that is, manipulating and playing with mobile objects. In particular, several chimpanzees have been reported to actively manipulate the Kong toy (www.kongcompany.com) [55], which is the type of toy that we use here. Using this device, one can associate loud, unusual sounds to object shaking and throwing. This can also facilitate acoustic production for individuals that do not usually hit surfaces and walls (see Wired Prototype).

The Wireless device is composed of a sensory module, namely a modified Wii Remote controller embedded in a Kong toy, and a computer with a Bluetooth interface (Figure 6). This prototype has several advantages. First of all, chimpanzees enjoy playing with objects and toys. Second, assembling this device requires less construction and electronic work than the wired prototype. All required components are available to buy online or in shops. Apart from a modification we operated on the Wii Remote in order to facilitate and speed up its interfacing with the computer, the only further work required for this prototype is software development in Max. Finally, depending on the foreseen application, a major advantage of this prototype resides in its wireless communication system. The sensory manipulable unit is in fact physically independent from, and connected via Bluetooth to, the computer.

Figure 6. The wireless prototype. The cable, connecting our external custom synchronization button to the Wii Remote, can be hidden inside the Kong toy once the synchronization is accomplished.

5.1. Sensory Module: Kong

The Wii Remote is embedded in a gum toy (Kong Genius Mike), normally used as enrichment for dogs. The toy has approximately the shape of a cylinder (diameter: 82 mm, length: 180 mm) and is hollow, allowing the Wii Remote to be tightly fitted inside. Using Genius Mike as the outer part of our prototype has a number of advantages. First of all, its shape and physical characteristics make it an interesting object for non-human animals to play with. Chimpanzees have been reported to spend time playing and exploring Kong toys [55]. Second, the Kong is slightly deformable though resistant, spurring the animals' interest in manipulating it while ensuring protection for the Wii Remote located inside. Finally, the Kong is among the most well-known and widely used enrichments for dogs. This makes it an option easily available to buy and replace in case of damage.

5.2. Sensory Module: Wii Remote

The core of this wireless prototype consists in a mildly modified Wii Remote controller. We decided to base this prototype on a consumer-level product for financial and practical reasons. The Wii Remote combines a 3-axes accelerometer, a battery pack and a Bluetooth transmitter for an affordable price. After doing a preliminary feasibility study, exploring the possibility of building an equivalent unit out of basic components (Arduino board, analog accelerometer, wireless transmitter, *etc.*), employing the Wii Remote turned out to be a cheaper and less laborious solution than its alternative.

An additional advantage of using a Wii Remote is the amount of literature and programming libraries already available for this device [9]. The Wii Remote is one of the most common, consumer-level integrated sensory devices, and it has already been used in a number of projects dealing with augmented reality, robotics, music, cognitive psychology, *etc.* [7,8,10]. Considering its broad usage in a variety of scientific, medical and artistic projects, its advantages and disadvantages have long been pointed out and improvements were implemented in later versions [3,5,6,8,10,32,33].

Nintendo does not officially provide Wii Remote's technical specifications. However, given the interest generated by this device in different scientific communities, most specifications have been reverse-engineered. Here, we report some data provided by Lee [9]. The three-axis accelerometer is an ADXL330 (Analog Devices), granting a ±3 g sensitivity range. Data is updated every 10 ms at 8 bit per axis and transmitted through a Bluetooth connection using a Broadcom 2042 chip.

Our prototype features two minor modifications, which simplify battery charge and connectivity functions. The disposable batteries were replaced with NiMH 3A batteries, rechargeable through an ordinary USB port. This allows the controller's batteries to be recharged without extracting the Wii Remote from the Kong toy. Moreover, the internal connection for the button "SYNC" was disabled and substituted for a switch located outside the main body of the Wii Remote. As this button is normally used to synchronize the device with different software, placing it outside the main body grants an easy, painless setup every time the sensory unit is prepared for use.

5.3. Max/MSP Software

The software part for this prototype is programmed in Max 6. Max/MSP is a "visual programming environment", where objects are connected, both visually and functionally, in order to form programs

(patches and subpatches). Instead of typing lines of code, algorithms are constructed by combining elements in a graphical user interface, making Max an accessible option for people with little or no programming experience. Max is widely used among artists for audio and video live performance. As its architecture is optimized for real-time signal processing, Max is an ideal option for our project.

The Max patch we use here is composed of five subpatches, responsible for data acquisition, processing, parameter extraction, sonification and logging.

Raw data acquisition is handled by the patch disis.aka.wiiremote (http://ico.bukvic.net/) developed by Ivica Ico Bukvic, based on aka.wiiremote (http://www.iamas.ac.jp/~aka/max/) by Masayuki Akamatsu. This patch enables Max to connect to the controller and to acquire the raw acceleration data in three dimensions.

The second subpatch is responsible for *data processing*. First of all, for each axis, the patch computes the running difference between consecutive accelerations over time (delta). These values are then used to calculate the Euclidean norm. These two computations are equivalent to calculating the change in acceleration over time (*jerk*) in the three-dimensional space. We focused on change in acceleration, rather than on absolute acceleration values, for practical matters. As this prototype is wireless, the chimpanzee will realistically walk while carrying it. Thus, periods of walking could be interspersed with bouts of object shaking, and our device needs to be able to discriminate between them. By only taking into account jerk, the fairly constant speed of walking will be ignored, while more abrupt variations in acceleration will be processed further. This is intended as a first filter on the data, towards the extraction of shaking and jerking movements. It is worth noticing that, as only jerk is used, there is no need to take into account the acceleration of gravity when processing data further [3,8].

Our third subpatch deals with *parameter extraction*. Its input is the computed jerk value, received every 10 ms while its output is a Boolean (specifically, a "bang" in Max syntax). This subpatch tests its input against two thresholds. A first "activation" threshold ensures that the jerk is above a critical value, in which case a "bang" is forwarded. Both the activation threshold and the incoming jerk value can be graphically displayed in real time (using a subpatch adapted from disis.aka.wiiremote). This enables visual inspection of the incoming signal and adjustment of the threshold according to the current magnitude of the parameter (see Figure 7). Setting a threshold slightly lower than the peak jerk value associated with the movement of interest can make the patch partially predictive [3,8].

Once activated, the bang reaches a timer and, after a comparison with the current time value, the timer is restarted. Verifying whether enough time has elapsed from last activation outputs a Boolean, which is sent to the sonification patch. The timer's characteristics ensure two desirable properties of our device: (i) no sound can be played within a delta time from the previous sound and (ii) if several activations are received in a short time interval, only the first one is sonified, with the timer being continuously reset. Property (i) prevents the device from producing two sounds when only one movement occurred while (ii) ensures that the production of different sound units is separated by periods of no or little movement.

The *sonification* part is kept to a minimum and is in charge of loading and playing chosen sounds. It is, however, possible to modify and expand it. Features computed in the data processing patch (such as acceleration magnitude and direction) could be sent to the sonification patch, so as to map particular movements to specific sonic features. In the simplest case, the magnitude of the acceleration could be mapped to the volume of the sound being played. An additional movement counter could for instance

be added, so as to map movements within a cluster to different sounds depending on their order of occurrence. Moreover, Max/MSP is designed to enable a fast implementation of acoustic effects, such as echo, delay, *etc.*, which we foresee as a particularly useful feature in our future experiments.

The last patch is responsible for *logging the data*. It receives the parameter extraction patch's output and writes it to a text file. It associates a time stamp, with 10 ms precision, to every input value and immediately writes both pieces of information as a new line to a .txt or .csv file. Additional data, such as the name of the sound played or acoustic parameters can be automatically added to the lines of text.

In the current version of the prototype, the sound is broadcasted through an external self-powered loudspeaker. This should be located outside the animals' enclosure and connected to the computer through cables. For small enough chimpanzee enclosures, this solution can be adequate. Even though the Wii Remote contains a loudspeaker, its low sound quality makes it unsuitable to play rich, structured sounds. We are currently working on a solution to send the audio signal back to the sensory module via radio (to avoid overloading the Bluetooth transmission) and to play sounds using an additional loudspeaker that would be contained in the Kong toy.

5.4. Testing the Prototype: Preliminary Results

We tested the operational range of the sensory unit indoors. No decrease in, or loss of, signal was reported, even at a distance greater than 20 m with several walls interposed. Using the prototype in an environment where the frequency band is shared with other devices could result in a decrease in the transmission range. However, this should not be a concern in animal facilities, where most radio devices—walkie talkies used by animal keepers and personnel—transmit at lower frequencies than Bluetooth ones. Unlike the wired prototype, which consisted of a new combination of technologies, the wireless prototype was not specifically calibrated. Given that the parameter used to produce sounds is based on the change in acceleration over time (see Section 5.3), any discordance between sensory readings and real accelerations, as long as it is a linear, systematic error, will not affect the computed jerk values and therefore will have no effect on sound production [65]. Hence, calibration was deemed unnecessary for our purposes.

As a preliminary feasibility test, we compared the readings obtained at rest to those obtained while walking with the device or when shaking it (Figure 7). In graphs a-c, accelerations along the 3 axes are plotted over time. The red line depicts accelerations along the x-axis (Wii Remote's coordinate system), the green line corresponds to the y-axis and the blue line refers to the z-axis. (Referring to the position of the Wii Remote in Figure 6, these axes correspond to upwards-downwards accelerations for x, leftwards-rightwards for y and towards/away from the observer for z.) In graphs d-f, jerk values obtained over time are depicted (red line), together with a sound activation threshold (green line).

The first column shows readings while the sensory module is at rest. Minimal variations in acceleration (top, Figure 7a) are not present in the corresponding jerk plot (bottom, Figure 7d).

The second column depicts 3-dimensional accelerations (Figure 7b) and jerk (Figure 7e) when walking with the device (human being). Accelerations are present (top) and possibly biased, the most obvious example being the blue line shifted vertically due to the acceleration of gravity. The jerk (bottom), however, is minimal and well below the depicted threshold.

The third column shows accelerations (Figure 7c) and jerk (Figure 7f) while shaking the device three times in succession (human being): here three potential activation events occurred. In order to be sonified, however, each potential event needs to pass the time activation threshold (see Section 5.3).

This preliminary test suggests that jerk can indeed be used to discriminate between movements associated with walking and movements associated with shaking the device. Indeed, an appropriate sound activation threshold can be selected such that walking does not trigger sound production but shaking the device leads to sound production. Chimpanzees show individual differences relative to gait, and their locomotion alternates between bipedal and quadrupedal [66]. Hence, the particular value of the sound activation threshold will have to be adjusted to the specific individual taking the experiment.

Figure 7. Acceleration readings (top row) from the wireless prototype and computed jerk values (bottom row) under three different conditions (columns).
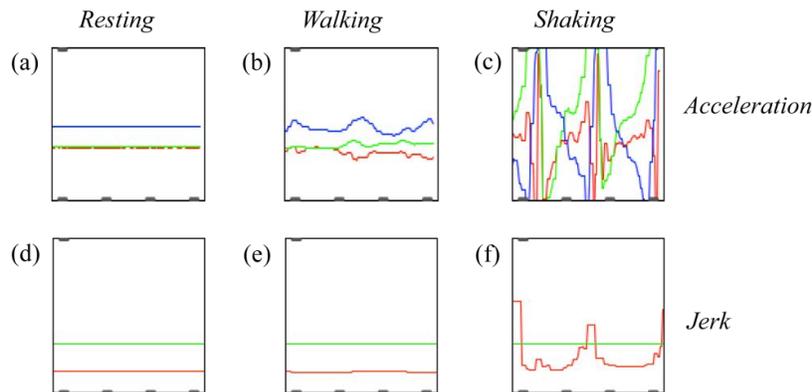
6. Pilot Work with Chimpanzees

A crucial step for validating our prototypes was to test them with non-human primates. Here we describe pilot work done with a group of chimpanzees at Edinburgh Zoo, Scotland, UK. We decided to focus on piloting the wired prototype only. This is because the vastness of the chimpanzee outdoor enclosure and the species-adequate structure of the indoor enclosure pods did not allow us to follow a focal chimpanzee transporting the wireless device. This entailed the risk that both, visual and radio contact with the device might have been lost repeatedly.

6.1. Methods

6.1.1. Study Species and Study Site

Budongo Trail, located at Edinburgh Zoo, houses 18 chimpanzees (10 females and 8 males) socially in an 1,832 m^2 outdoor enclosure and three interconnected $12 \times 12 \times 14$ m indoor enclosures [67]. The animals have *ad libitum* access to water and are fed four times a day. Research with the chimpanzees is strictly non-invasive. The animals take part in experiments voluntarily and can leave them at any moment. Before taking place, this pilot study was approved by the scientific board of Budongo Trail,

and did not entail any ethical concerns. The nature of the study was observational and no invasive methods were applied.

6.1.2. Device Setup

The wired prototype was piloted with the Budongo chimpanzees. We slightly modified the prototype to adapt it to the specific testing location (See Figure 8). The inner Plexiglas layer, containing no sensors, was replaced by a thicker (15 mm) concave one (to protect the sides of the device). The outer layer's dimensions (width and length) were increased by 36 mm each, while the Plexiglas' thickness was left unchanged. Plugs were added to the connection cables so they could be passed through multiple enclosure elements. Sounds were broadcasted using an active loudspeaker (JBL Control 2P) placed approximately 1 m behind the device.

During the pilot session described below, in order to maximize the likelihood of interaction with the device, the Arduino thresholds were lowered, so that even a delicate touch would elicit a sound. The device was installed and fixed on a wire mesh in one of the core chimpanzees' whereabouts, which granted that all individuals could have access to the device. Chimpanzees could access this area from two sides on two different floors.

Figure 8. Wired prototype, after slight modifications to fit the pilot testing location. A thicker inner Plexiglas layer granted better resistance once screwed to the chimpanzees' enclosure. Moreover, as the device had to be placed in the enclosure in its entirety, "protection walls" were built to avoid chimpanzees' contact with the solid foam.
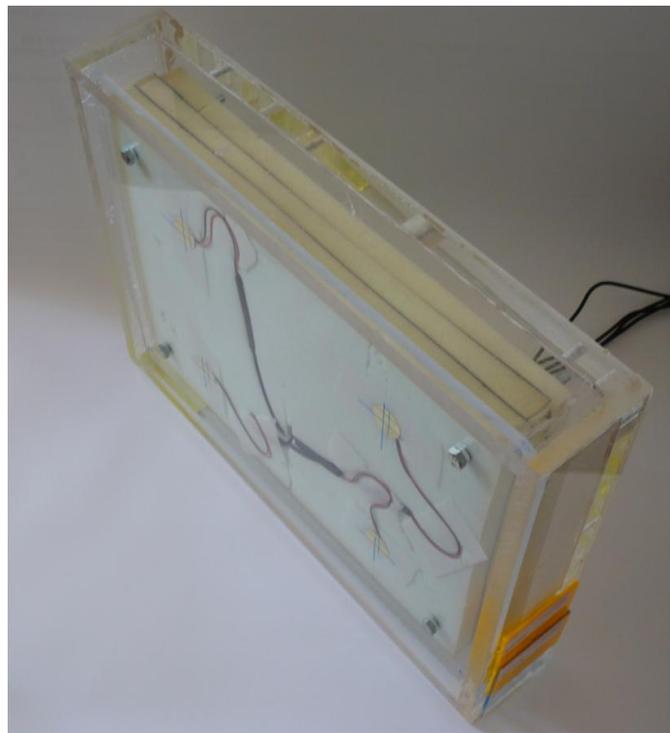
6.2. Pilot Session and Data Collection

After installing the device in the animals' enclosure, chimpanzees were allowed to enter and interact with the prototype. Given the exploratory nature of this pilot experiment, we decided to keep the sound intensity and type mostly constant (snare drum). However, during brief periods of time, the sound mapping was altered, in order to explore and probe future research directions. Throughout the approximately 33 min long session, three different types of sounds were associated to chimpanzees' manipulations: snare drum (69.5% of the time), Hi-hat drum (1.5%), and human spoken syllables (17.5%). The remaining 11.5% of the time, either no sound was associated to manipulations or no chimpanzee interacted with the device. The pilot session was recorded on video from which behavioral and interaction patterns were coded [68,69] using ELAN [70–72]. The following behaviors towards the device were coded for:

- approaches to and recessions from the device (within a 2 m distance),
- smelling, licking, looking, gently touching, grabbing and climbing,
- using additional tools (such as sticks) to interact,
- poking with finger or hand,
- pushing with shoulder, head, hands, feet or other body parts,
- hitting with hand or arm,
- pant-hoot display, including kicking or punching.

A custom Python script was used to extract the data from ELAN [70–72] annotations and fill in Table 3.

Table 3. Duration of proximity to the wired device and frequency of interaction (by interaction type) performed by 12 individual chimpanzees.

Name	Proximity (s)	Poking	Tool	Push	Hit	Display
C1	211	12	2	0	0	0
C2	140	3	0	0	0	0
C3	122	19	0	4	3	0
C4	211	32	3	75	0	0
C5	32	1	0	0	0	0
C6	68	2	0	0	0	2
C7	517	1	0	0	0	0
C8	288	31	5	35	51	2
C9	458	20	10	10	0	0
C10	6	1	0	0	0	0
C11	243	25	1	90	0	0
C12	20	2	0	3	0	0
Total	2,316	149	21	217	54	4

6.3. Results

A total of 12 individuals (reported as C1 to C12 in Table 3) voluntarily approached the device and interacted with it. Each of these chimpanzees spent between 6 s and 9 min in proximity of the device. The overall time spent by the group in proximity of the prototype exceeded 38 min: at several points in

time, two (or more) individuals were in proximity of the device. All 12 individuals who approached the device also showed exploratory behavior.

At a group level of analysis, the behavior most frequently recorded was pushing (Table 3). Five individuals pushed the prototype eliciting sounds. Two individuals (C4 with 75 and C11 with 90 pushes) in particular produced sounds by pushing. The second most observed behavior was poking. All 12 chimpanzees who interacted with the device showed this behavior. Two individuals hit the device, one of them 51 times. Five out of the 12 chimpanzees interacting with the device used a tool (such as a stick) to poke on the surface or between gaps of the prototype. Two chimpanzees used the device for dominance displays, each of them twice.

The coded behaviors can be divided into three main behavioral categories: (i) explorative behaviors (poking, tool use) [73,74], (ii) play behaviors (pushing and hitting) [75,76], and (iii) dominance behaviors (displays with hits or kicks) [75]. Applying this categorization, the behavioral observations indicate that the chimpanzees used the device foremost for play, followed by explorative behavior, and then dominance behaviors.

6.4. Discussion

Overall, the pilot session showed that the wired prototype fulfills its intended purpose. The prototype withstood vigorous manipulation, including kicks and punches. While exploring the device, chimpanzees stood on it without any damage to the unit. This suggests that both the prototype itself and our solution for fixing it to animal enclosures are stable and resistant.

Interaction rates with the device were high both in the number of participating individuals, and in the total amount of interactions observed. The majority of the chimpanzees (12 out of 18 individuals) showed some sort of interaction with the device within only 33 min of access. This suggests that the wired prototype represents an object of interest for our study species. This is a crucial prerequisite for the conduction of further experiments on acoustic pattern production in chimpanzees.

One of the most common ways of interacting with the prototype was to poke the outer Plexiglas layer. More specifically, all chimpanzees touched the sound-producing surface of the device with their hands or fingers. This can be interpreted as explorative behavior related to getting acquainted to the novel object. Further *explorative behavior*, such as looking at, licking, or touching the device, was observed at high rates. Several individuals also used tools to further interact with the device. In the long term however, as the chimpanzees familiarize themselves with the device, we predict a decrease in explorative behaviors. Hence, we argue that no experimental design should relate to these kinds of manipulations of the device.

Pushing the outer layer was the most frequent interaction method recorded overall, though only used by six chimpanzees. Hitting the prototype was a relatively infrequent behavior, found only in two individuals. This can be partly due to the novelty effect of the device. As explorative behaviors are expected to decrease over time, other behaviors such as hitting and pushing are predicted to increase. Pushing and hitting behaviors can be interpreted as *play behavior*. This interpretation is supported by the fact that the chimpanzees interacting with the prototype often showed species-specific play face (open mouth and teeth shown) [75]. For instance, individual C4 climbed on the device, held himself onto the wire mesh and started a series of pushes with his feet. The pushing was accompanied by a

facial expression typically associated to play behavior in chimpanzees. Play face was also observed when individual C8 sat in front of the device, hitting it multiple times in a row.

Play behavior can be a good starting point to build upon when designing experiments, since: (1) the amount of playful interactions with the device can be expected to increase over time, (2) using the device in a playful manner suggests that a positive self-rewarding mechanism not only increases the total amount of interactions, but also enlarges the behavioral repertoire (*i.e*., how the device is used). Experimentally manipulating sounds that are played back to the study subjects can further enhance the chimpanzees' interest and positively feed back to the exhibition of play behavior. In this respect we see also a positive potential welfare aspect for captive chimpanzees. Using the device as an enrichment tool might eventually increase the animals' wellbeing in captive environments.

Two individuals used the device to *display*, performing a series of loud vocalizations (pant-hoots), which culminated in hitting the device with upper or lower limbs. Only male chimpanzees showed this type of behavior. Considering the relatively short access time (33 min), this frequency of displays can be considered high. Allowing chimpanzees open access to such a device would likely result in its usage for such dominance displays. Hence we argue that this behavioral category offers another area of interest for future experiments.

7. Sonification: Possible Criteria and Future Directions

The device was used by a large proportion of chimpanzees and in a variety of ways. This suggests that its shape and physical characteristics will be an asset in engaging chimpanzee participants to take part to future experiments. In light of the pilot session described above and the available literature on chimpanzees' behavior, we can now put forward possibilities of movement-sound mappings, which can be implemented in our prototypes in the near future.

The sound associated to manipulations was mostly that of a snare drum at constant volume. While one of the individuals was producing the snare sound, we artificially changed the mapping to a (less prominent) hi-hat sound. This elicited an unusual vocalization, comparable to an expression of disappointment. Mapping to human-spoken syllables, instead, did not appear to elicit particular reactions. Even though these only constitute anecdotic instances, we believe that manipulation of sound mappings could have a lot of potential. Future experiments, controlling for and counterbalancing the type and proportion of sounds associated with manipulation, will be able to inform us about chimpanzees' preferences for movement-sound and emotional state-sound mappings.

During the pilot, interactions of different strength were shown. Displays, from the point of view of strain, were harder than play or explorative behavior. These two types of (soft and hard) manipulations should however be regarded as extremes of a continuum, rather than clear-cut categories. Taking into account chimpanzees' biology, displays are usually associated with increased levels of arousal. This consideration should be included when designing sonification paradigms.

Both the type of interaction and the context in which it is performed should inform the movement-sound mapping. The wired prototype can discriminate *hit* from *trust* movements. The former manipulations are often associated with exploratory and play behavior while the latter with pant-hoot displays. Hence, a first possibility could be to play the same sound in the two conditions, while assigning a higher playback intensity level to the thrust condition. Alternatively, or in addition to the intensity

manipulation, different sounds could be assigned to the various conditions. The spectral properties of the sounds should reflect this distinction: sounds rich in low frequencies, with short attack time, rough and possibly noisy (mimicking spectral characteristics of natural chimpanzee drumming during pant-hoot displays) could be assigned to the thrust condition. In order to generate a contrast, sounds with opposite characteristic could be assigned to exploratory manipulations.

Other possibilities could be to only manipulate spectral characteristics when chimpanzees play with the device, and sound intensity while chimpanzees perform dominance displays. While sound novelty and variety may be more appropriate when chimpanzees are in a playful mood, unexpectedly high or low intensity could elicit interesting reactions during dominance displays.

As mentioned above, we did not have the opportunity to pilot the wireless prototype with chimpanzees. However, we could observe chimpanzees playing with the gum toy used as outer shell for the prototype. Extrapolating from the pilot session, a continuum of sensed states could be mapped to sound. Assuming that the chimpanzees' strength in manipulating the device will increase concurrently with their arousal levels, the sound should vary accordingly. Mapping higher *jerk* values to increased sound intensity and sound roughness would be the first straightforward possibility.

All these options would have the advantage of: (i) starting a positive feedback loop between chimpanzees' arousal and sound intensity, and (ii) not scaring away explorative individuals with incommensurately loud or potent sounds. Once this kind of mapping is in place, however, we could systematically invert the mapping, so to violate the animals' expectations and record the resulting behavior.

We recognize that this is a modest range of possibilities in comparison to human sonification projects. However, this is also the first attested attempt to connect machine interaction, auditory cognition and sonification in a non-human primate.

8. Conclusions

Over the last decade, music performance and cognitive experiments have been increasingly relying on the use of human-computer interaction technologies [7,27,30]. Though many devices have been specifically produced and optimized for human use, the field of animal cognition still lacks the development of enough ad-hoc technological tools [35,49,50]. In particular, although many possibilities exist for sonifying human movements, none has been developed in the field of non-human primate research.

Here we present two prototypes that we developed for mapping chimpanzees' object manipulation to sounds, using acceleration and strain sensors. Our prototypes will serve to test crucial hypotheses on primate auditory cognition, musicality and evolutionary development [22,23,25,77]. New devices had to be created, as tools currently available for human research lack a number of features necessary when working with chimpanzees. In particular, the sensory units of our prototypes are resistant to robust handling, modular, low-voltage and relatively cheap [56]. These qualities, foreseeing a vigorous manipulation by chimpanzees, avoid the destruction of the entire device, prevent electrocution, and allow for easy replacement of parts.

The software part of our prototypes has a number of interesting qualities, which make it particularly suitable for auditory experiments. First of all, the simplicity of our algorithms grants a short delay between sensing and auditory feedback. Second, acoustic parameters can be quickly adapted to

maximize the primates' interest in the device. Finally, all relevant sensory data can be logged in a computer to enable more efficient data analysis.

Our two prototypes satisfy a number of desiderata though differ in several respects. Our wired prototype is supposed to be mounted on a vertical surface and employs cables to connect to a Mac computer via an Arduino board. It is designed to acoustically respond to thrust and has a ricochet property, which encourages such movement. Four piezoelectric elements inside the prototype are responsible for strain data acquisition. This prototype has the advantage of being particularly cheap to develop, though it requires some electronic and construction work.

Our wireless prototype consists in a modified Wii Remote embedded in a gum toy and connected to a Mac computer via Bluetooth. The data from the acceleration sensor in the Wii Remote is sonified in case of shaking and jerking movements of a given magnitude. This prototype has the advantage of being wireless and requiring little development work, as its components can be easily bought and assembled.

These prototypes are a first step towards testing a range of hypotheses on the human specificity of rhythmic abilities among primates. In light of the pilot study presented here, we intend to further adjust our designs to chimpanzees' and experimental needs. Successively we intend to test, among others: preference for movement-sound mappings, multimodal entrainment to a steady pulse, discrimination and preference for different metric and grouping structures, and rhythmic imitation capabilities.

The fact that the wired prototype was used for dominance displays suggests that the prototypes may enable researchers to tap into chimpanzees' auditory cognition and spontaneous social behavior. Jane Goodall reports the story of a chimpanzee in the wild who, through the fortuitous discovery of a resonating barrel, climbed up the dominance hierarchy by being the loudest drummer in the party [78]. Goodall's report hence hints at the importance of sonic interactions in chimpanzees' social life. Together with our pilot data, it also suggests that future experiments on non-vocal sound production in chimpanzees can be designed without losing ecological validity.

The devices presented here are noteworthy for several reasons. To our knowledge, they constitute the first attempt at building a sensory device specifically designed for chimpanzee-computer cognitive interaction. Their development was markedly interdisciplinary, benefitting from knowledge both in human-computer interaction and animal cognition. Finally, these devices will serve for several crucial experiments dealing with human-chimpanzee comparative cognition. These experiments will allow to substitute conjectures and key hypotheses on the purported uniqueness of human musicality and language abilities with solid empirical data [25,56,77].

Author Contributions

Andrea Ravignani and W. Tecumseh Fitch conceived the idea of a primate-specific device, Andrea Ravignani, Bruno Gingras and W. Tecumseh Fitch further developed the concept, Andrea Ravignani, Riccardo Hofer and Carlos Rodríguez contributed to the construction of the prototypes, Vicente Matellán, Bruno Gingras and W. Tecumseh Fitch supervised the construction and Vicente Matellán, Riccardo Hofer and Carlos Rodríguez advised on software, available technologies and calibration, Andrea Ravignani and Ruth-Sophie Sonnweber prepared and performed the pilot test with chimpanzees, Ruth-Sophie Sonnweber coded the chimpanzees' behavior from video and videotaped the interaction of chimpanzees with the Kong toy, Andrea Ravignani wrote the software, collected and analyzed the calibration data and wrote the manuscript, Andrea Ravignani, Bruno Gingras, Riccardo Hofer, Vicente Matellán, Ruth-Sophie Sonnweber and W. Tecumseh Fitch edited the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Gibson, E.J. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annu. Rev. Psychol.* **1988**, *39*, 1–42.
2. Addessi, A.; Ferrari, L. Observing Flow in Child/Music Machine Interaction. In Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research, Utrecht, The Netherlands, 28–31 August 2012; Spink, A.J., Grieco, F., Krips, O.E., Loijens, L.W.S., Noldus, L.P.J.J., Zimmerman, P.H., Eds.; Noldus Information Technology bv.: Utrecht, The Netherlands, August 2012; pp. 324–327.
3. Bencina, R.; Wilde, D.; Langley, S. Gesture-Sound Experiments: Process and Mappings. In Proceedings of The 8th International Conference on New Interfaces for Musical Expression (NIME 08), Genova, Italy, 5–7 June 2008; pp. 197–202.
4. Butler, J. Creating Pedagogical Etudes for Interactive Instruments. In Proceedings of The 8th International Conference on New Interfaces for Musical Expression (NIME 08), Genova, Italy, 5–7 June 2008; pp. 77–80.
5. Demey, M.; Leman, M.; Bossuyt, F.; Vanfleteren, J. The Musical Synchrotron: Using Wireless Motion Sensors to Study How Social Interaction Affects Synchronization with Musical Tempo. In Proceedings of The 8th International Conference on New Interfaces for Musical Expression (NIME 08), Genova, Italy, 5–7 June 2008; pp. 372–373.
6. Wong, E.L.; Yuen, W.Y.; Choy, C.S. Designing Wii Controller: A Powerful Musical Instrument in an Interactive Music Performance System. In Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, Linz, Austria, 24–26 November 2008; pp. 82–87.
7. Dale, R.; Roche, J.; Snyder, K.; McCall, R. Exploring action dynamics as an index of paired-associate learning. *PLoS One* **2008**, *3*, doi:10.1371/journal.pone.0001728.

8. Dolhansky, B.; McPherson, A.; Kim, Y.E. Designing an Expressive Virtual Percussion Instrument. In Proceedings of the 8th Sound and Music Computing Conference, Padova, Italy, 6–9 July 2011.

9. Lee, J.C. Hacking the nintendo wii remote. *IEEE Pervas. Comput.* **2008**, *7*, 39–45.

10. Miller, J.; Hammond, T. Wiiolin: A Virtual Instrument Using the Wii Remote. In Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), Sydney, Australia, 15–18 June 2010; pp. 497–500.

11. Yonezawa, T.; Clarkson, B.; Yasumura, M.; Mase, K. Context-Aware Sensor-Doll as a Music Expression Device. In Proceedings of CHI'01 extended abstracts on Human factors in computing systems, Seattle, WA, USA, 31 March–5 April 2001; pp. 307–308.

12. Ruiz-Garcia, L.; Lunadei, L.; Barreiro, P.; Robla, I. A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. *Sensors* **2009**, *9*, 4728–4750.

13. Wang, N.; Zhang, N.; Wang, M. Wireless sensors in agriculture and food industry: Recent development and future perspective. *Comput. Electron. Agric.* **2006**, *50*, 1–14.

14. Swann, D.E.; Hass, C.C.; Dalton, D.C.; Wolf, S.A. Infrared-triggered cameras for detecting wildlife: An evaluation and review. *Wildlife Soc. Bull.* **2004**, *32*, 357–365.

15. Fagot, J.; Bonté, E. Automated testing of cognitive performance in monkeys: use of a battery of computerized test systems by a troop of semi-free-ranging baboons (papio papio). *Behav. Res. Meth.* **2010**, *42*, 507–516.

16. Fagot, J.; Paleressompoulle, D. Automatic testing of cognitive performance in baboons maintained in social groups. *Behav. Res. Meth.* **2009**, *41*, 396–404.

17. Inoue, S.; Matsuzawa, T. Working memory of numerals in chimpanzees. *Curr. Biol.* **2007**, *17*, R1004–R1005.

18. Steurer, M.M.; Aust, U.; Huber, L. The vienna comparative cognition technology (VCCT): An innovative operant conditioning system for various species and experimental procedures. *Behav. Res. Meth.* **2012**, *44*, 909–918.

19. Kano, F.; Tomonaga, M. How chimpanzees look at pictures: a comparative eye-tracking study. *Proc. R. Soc. B: Biol. Sci.* **2009**, *276*, 1949–1955.

20. Arcadi, A.C.; Robert, D.; Boesch, C. Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* **1998**, *39*, 505–518.

21. Sugimoto, T.; Kobayashi, H.; Nobuyoshi, N.; Kiriyama, Y.; Takeshita, H.; Nakamura, T.; Hashiya, K. Preference for consonant music over dissonant music by an infant chimpanzee. *Primates* **2010**, *51*, 7–12.

22. Fitch, W.T. The Biology and Evolution of Rhythm: Unravelling a Paradox. In *Language and Music as Cognitive Systems*; Rebuschat, P., Rohrmeier, M., Hawkins, J.A., Cross, I., Eds.; Oxford University Press: Oxford, UK, 2011; Chapter 9, pp. 73–96.

23. Ravignani, A.; Gingras, B.; Asano, R.; Sonnweber, R.; Matellán, V.; Fitch, W.T. The Evolution of Rhythmic Cognition: New Perspectives and Technologies in Comparative Research. In Proceedings of the 35th Annual Conference of the Cognitive Science Society, Berlin, Germany, 31July–3 August 2013; Knauff, M., Pauen M., Sebanz, N. & Wachsmuth, I., Eds.; Cognitive Science Society: Austin, TX, USA, 2013; pp. 1199–1204.

24. Patel, A.D.; Iversen, J.R.; Bregman, M.R.; Schulz, I. Studying synchronization to a musical beat in nonhuman animals. *Ann.NY Acad. Sci.* **2009**, *1169*, 459–469.

25. Patel, A.D. *Music, Language, and the Brain*; Oxford University Press: New York, NY, USA, 2010; p. 520.

26. Hermann, T.; Hunt, A.; Neuhoff, J.G. *The Sonification Handbook*; Logos Verlag Berlin: Berlin, Germany, November 2011.

27. Leman, M. *Embodied Music: Cognition and Mediation Technology*; MIT Press: Cambridge, MA, USA, 2008.

28. Loeb, R.G.; Fitch, W.T. A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesth. Analg.* **2002**, *94*, 362–368.

29. Delle Monache, S.; Polotti, P.; Papetti, S.; Rocchesso, D. Sonically Augmented Found Objects. In Proceedings of 8 th International Conference on New Interfaces for Musical Expression (NIME08), Genova, Italy, 4–8 June 2008.

30. Leman, M.; Lesaffre, M.; Nijs, L.; Deweppe, A. User-oriented studies in embodied music cognition research. *Music. Sci.* **2010**, 203–223.

31. Moens, B.; van Noorden, L.; Leman, M. D-Jogger: Syncing Music with Walking. In Proceedings of 7th Sound and Music Computing Conference, Barcelona, Spain, 21–24 July 2010; pp. 451–456.

32. Bouillot, N.; Wozniewski, M.; Settel, Z.; Cooperstock, J.R. A Mobile Wireless Augmented Guitar. In Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME08), Genova, Italy, 4–8 June 2008.

33. Kiefer, C.; Collins, N.; Fitzpatrick, G. HCI Methodology for Evaluating Musical Controllers: A Case Study. In Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME08), Genova, Italy, 4–8 June 2008.

34. McGrath, R.E. Species-Appropriate Computer Mediated Interaction. In Proceedings of CHI'09 Extended Abstracts on Human Factors in Computing Systems, Boston, USA, 4–9 April 2009; pp. 2529–2534.

35. Mancini, C.; Lawson, S.; van der Linden, J.; Häkkilä, J.; Noz, F.; Wingrave, C.; Juhlin, O. Animal-Computer Interaction SIG. In Proceedings of the 2012 ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems Extended Abstracts, Austin, TX, USA, 5–10 May 2012; pp. 1233–1236.

36. Mancini, C. Animal-computer interaction: A manifesto. *Interactions* **2011**, *18*, 69–73.

37. Barbari, M.; Conti, L.; Simonini, S. Spatial Identification of Animals in Different Breeding Systems to Monitor Behavior. In Proceedings of Livestock Environment VIII, Iguassu Falls, Brazil, 31 August–4 September 2008.

38. Nadimi, E.; Sogaard, H.; Bak, T.; Oudshoorn, F.W. ZigBee-based wireless sensor networks for monitoring animal presence and pasture time in a strip of new grass. *Comput. Electron. Agric.* **2008**, *61*, 79–87.

39. Nadimi, E.; Sogaard, H.T.; Bak, T. ZigBee-based wireless sensor networks for classifying the behaviour of a herd of animals using classification Trees. *Biosyst. Eng.* **2008**, *100*, 167–176.

40. Pereira, D.P.; Dias, W.; Braga, M.; Barreto, R.; Figueiredo, C.M.C.S.; Brilhante, V.N. Model to Integration of RFID into Wireless Sensor Network for Tracking and Monitoring Animals. In Proceedings of 11th IEEE International Conference on Computational Science and Engineering (CSE'08), Washington, DC, USA, 16–18 July 2008; pp. 125–131.

41. Frost, A.; Schofield, C.; Beaulah, S.; Mottram, T.; Lines, J.; Wathes, C. A review of livestock monitoring and the need for integrated systems. *Comput. Electron. Agric.* **1997**, *17*, 139–159.

42. Cugnasca, C.; Saraiva, A.; Nääs, I.D.A.; Moura, D.d.; Ceschini, G. Ad Hoc Wireless Sensor Networks Applied to Animal Welfare Research. In *Central Theme, Technology for All: Sharing the Knowledge for Development*, Proceedings of the International Conference of Agricultural Engineering, XXXVII Brazilian Congress of Agricultural Engineering, International Livestock Environment Symposium-ILES VIII, Iguassu Falls City, Brazil, 31 August–4 September 2008.

43. Darr, M.; Zhao, L. A Wireless Data Acquisition System for Monitoring Temperature Variations in Swine Barns. In *Central Theme, Technology for All: Sharing the Knowledge for Development*, Proceedings of the International Conference of Agricultural Engineering, XXXVII Brazilian Congress of Agricultural Engineering, International Livestock Environment Symposium-ILES VIII, Iguassu Falls City, Brazil, 31 August–4 September 2008.

44. Ipema, A.; Goense, D.; Hogewerf, P.; Houwers, H.; van Roest, H. Pilot study to monitor body temperature of dairy cows with a rumen bolus. *Comput. Electron. Agric.* **2008**, *64*, 49–52.

45. Marsh, J.; Gates, R.; Day, G.; Aiken, G.; Wilkerson, E. Assessment of an Injectable RFID Temperature Sensor for Indication of Horse Well-Being. In Proceedings of ASABE Annual International Meeting, Providence, RI, USA, 29 June–2 July 2008.

46. Mayer, K.; Ellis, K.; Taylor, K. Cattle Health Monitoring Using Wireless Sensor Networks. In Proceedings of IASTED International Conference on Communication and Computer Networks, Cambridge, MA, USA, 8–10 November 2004.

47. Nagl, L.; Schmitz, R.; Warren, S.; Hildreth, T.; Erickson, H.; Andresen, D. Wearable Sensor System for Wireless State-of-Health Determination in Cattle. In Proceedings of the 25th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, Cancún, Mexico, 17–21 September 2003; pp. 3012–3015.

48. Beilharz, K.; Jakovich, J.; Ferguson, S. Hyper-Shaku (Border-Crossing): Towards the Multi-Modal Gesture-Controlled Hyper-Instrument. In Proceedings of the 2006 Conference on New Interfaces for Musical Expression, Centre Pompidou Paris, France, 4–8 June 2006; pp. 352–357.

49. Tan, R.T.K.C.; Cheok, A.D.; Peiris, R.; Todorovic, V.; Loi, H.C.; Loh, C.W.; Nguyen, D.T.K.; Sen, J.Y.P.; Yio, E.Z.; Derek, T.B.S. Metazoa ludens: Mixed reality interactions and play for small pets and humans. *Leonardo* **2008**, *41*, 308–309.

50. Savage, J.; Sanchez-Guzman, R.; Mayol-Cuevas, W.; Arce, L.; Hernandez, A.; Brier, L.; Martinez, F.; Velazquez, A.; Lopez, G. Animal-Machine Interfaces. In Proceedings of the Fourth International Symposium on Wearable Computers, Atlanta, GA, USA, 18–21 October 2000; pp. 191–192.

51. Mancini, C.; van der Linden, J.; Bryan, J.; Stuart, A. Exploring Interspecies Sensemaking: Dog Tracking Semiotics and Multispecies Ethnography. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 143–152.

52. Gazes, R.P.; Brown, E.K.; Basile, B.M.; Hampton, R.R. Automated cognitive testing of monkeys in social groups yields results comparable to individual laboratory-based testing. *Anim. Cogn.* **2012**, *16*, 1–14.

53. Taffoni, F.; Vespignani, M.; Formica, D.; Cavallo, G.; di Sorrentino, E.P.; Sabbatini, G.; Truppa, V.; Mirolli, M.; Baldassarre, G.; Visalberghi, E. A mechatronic platform for behavioral analysis on nonhuman primates. *J. Integr. Neurosci.* **2012**, *11*, 87–101.

54. Kramer, G.; Walker, B.; Bonebright, T.; Cook, P.; Flowers, J.; Miner, N.; Neuhoff, J.; Bargar, R.; Barrass, S.; Berger, J.; *et. al. The sonification report: Status of the Field and Research Agenda*; Technical Report for the National Science Foundation by Members of the International Community for Auditory Display (ICAD): Santa Fe, NM, USA, 1999; p. 3.

55. Pruetz, J.; Bloomsmith, M. Comparing two manipulable objects as enrichment for captive chimpanzefs. *Anim. Welfare* **1992**, *1*, 127–137.

56. Bauman, J.E. Observations on the strength of the chimpanzee and its implications. *J. Mammal.* **1926**, *7*, 1–9.

57. Slocombe, K.E.; Zuberbühler, K. Food-associated calls in chimpanzees: Responses to food types or food preferences? *Anim. Behav.* **2006**, *72*, 989–999.

58. Morris, D. *Primate Ethology*; Aldine: Chicago, IL, USA, 1967.

59. Simonyan, K.; Horwitz, B.; Jarvis, E.D. Dopamine regulation of human speech and bird song: A critical review. *Brain Lang.* **2012**, *122*, 142–150.

60. Fischer, J.; Hammerschmidt, K. Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: Insights into the evolution of vocal communication. *Genes Brain Behav.* **2011**, *10*, 17–27.

61. Riters, L.V.; Stevenson, S.A. Reward and vocal production: Song-associated place preference in songbirds. *Physiol. Behav.* **2012**, *106*, 87–94.

62. Earp, S.E.; Maney, D.L. Birdsong: Is it music to their ears? *Front. Evol. Neurosci.* **2012**, *4*, doi:10.3389/fnevo.2012.00014.

63. Vick, S.J.; Anderson, J.R.; Young, R. Maracas for Macaca? Evaluation of three potential enrichment objects in two species of zoo-housed macaques. *Zoo Biol.* **2000**, *19*, 181–191.

64. Du, Q.; Faber, V.; Gunzburger, M. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Rev. Soc. Ind. Appl. Math.* **1999**, *41*, 637–676.

65. Olivares, A.; Górriz, J.; Ramŕez, J.; Olivares, G. Accurate human limb angle measurement: Sensor fusion through Kalman, least mean squares and recursive least-squares adaptive filtering. *Meas. Sci. Technol.* **2011**, *22*, doi:10.1088/0957-0233/22/2/025801.

66. Sockol, M.D.; Raichlen, D.A.; Pontzer, H. Chimpanzee locomotor energetics and the origin of human bipedalism. *Proc. Natl. Acad. Sci.* **2007**, *104*, 12265–12269.

67. Schel, A.; Rawlings, B.; Claidiere, N.; Wilke, C.; Wathan, J.; Richardson, J.; Pearson, S.; Herrelko, E.S.; Whiten, A.; Slocombe, K. Network analysis of social changes in a captive chimpanzee community following the successful integration of two adult groups. *Am. J. Primatol.* **2013**, *75*, 254–266.

68. Ravignani, A.; Fitch, W.T. Sonification of Experimental Parameters as a New Method for Efficient Coding of Behavior. In Proceedings of Measuring Behavior 2012, 8th International Conference on Methods and Techniques in Behavioral Research, Utrecht, The Netherlands, 28–31 August 2012; Spink, A.J., Grieco, F., Krips, O.E., Loijens, L.W.S., Noldus, L.P.J.J., Zimmerman, P.H., Eds.; pp. 376–379.

69. Martin, P.; Bateson, P. *Measuring Behaviour: An Introductory Guide*; Cambridge University Press: Cambridge, UK, 1993.

70. Wittenburg, P.; Brugman, H.; Russel, A.; Klassmann, A.; Sloetjes, H. ELAN: A Professional Framework for Multimodality Research. In Proceedings of Fifth International Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, 24–26 May 2006; pp. 1556–1559.

71. Lausberg, H.; Sloetjes, H. Coding gestural behavior with the NEUROGES-ELAN system. *Behav. Res. Meth.* **2009**, *41*, 841–849.

72. Max Planck Institute for Psycholinguistics. Available online: http://tla.mpi.nl/tools/tla-tools/elan/ (accessed on 29 July 2013).

73. Whiten, A.; Horner, V.; de Waal, F.B.M. Conformity to cultural norms of tool use in chimpanzees. *Nature* **2005**, *437*, 737–740.

74. Inoue-Nakamura, N.; Matsuzawa, T. Development of stone tool use by wild chimpanzees (Pan troglodytes). *J. Comp. Psychol.* **1997**, *111*, 159–173.

75. Nishida, T. *Chimpanzee Behavior in the Wild: an Audio-Visual Encyclopedia*; Springer: Berlin, Germany, 2010.

76. Maestripieri, D. *Games Primates Play: An Undercover Investigaton of the Evolution and Economics of Human Relationships*; Basic Books: New York, NY, USA, 2012.

77. Patel, A.D. Language, music, syntax and the brain. *Nat. Neurosci.* **2003**, *6*, 674–681.

78. Goodall, J. *In the Shadow of Man*; Mariner Books: New York, NY, USA, 2009.

6 | Group Chorusing

Comment

Author for correspondence:
Andrea Ravignani
e-mail: andrea.ravignani@univie.ac.at

Chronometry for the chorusing herd: Hamilton's legacy on context-dependent acoustic signalling—a comment on Herbers (2013)

Andrea Ravignani[1,2]

[1]Department of Cognitive Biology, University of Vienna, Althanstrasse 14, Vienna 1090, Austria
[2]Language Evolution and Computation Research Unit, University of Edinburgh, Edinburgh, UK

Biology Letters' special feature on Hamilton's legacy pays due tribute to a brilliant mind. Herbers [1] and the other contributors paint a compelling picture of how Hamilton's work on inclusive fitness anticipated much contemporary evolutionary thinking, although sometimes not acknowledged until much later.

A more recent, although equally cited work by Hamilton is the 'Geometry for the selfish herd' [2], an elegant mathematical description of why individuals aggregate in space. In the spirit of this special feature [1], I illustrate why Hamilton's herd model should be recognized as an early mathematical formalism applicable to unrelated, although crucial, biological phenomena. Notably, Hamilton's model of gregarious behaviour can be directly applied to the problem of context-dependent acoustic signalling as follows, with the potential to describe how interdependent individual calls combine into choruses.

Many animals communicate acoustically, often with an emphasis on signal timing, rather than other acoustic properties [3]. Synchrony and chorusing occur in insects, amphibians, birds and mammals. An overarching question is how individuals 'distribute' their calls over time and why different individuals' calls group together, leading to synchronous, alternating or phase-locked choruses [3]. Two hypotheses, suggested and tested in [4], predict clusters of calls: individuals could maximize overall sound intensity to attract females or, alternatively, individuals could call in quasi-synchrony to decrease the individual risk of predation. In both cases, individuals would tend to call close to each other, so to increase signal amplitude or alter individual conspicuousness (depending on the receiver), similarly to what happens in human applauding [5].

Suppose three frogs, A, B and C, call periodically in time, say every second, although with different relative phases (see figure 1). B and C occur within a short time interval (short *silence*). A precedes them by a long interval (long *SILENCE*). The resulting acoustic pattern is A-*SILENCE*-B-*silence*-C-*SILENCE*-A-...A can modify its conspicuousness by shortening its 'domain of silence', i.e. timing its signal so it co-occurs, on average, with others' calls. The most noise-robust, error-resistant strategy for A is to delay its call and signal exactly halfway between B and C; A calls, on average, in an 'acoustically dense' time period.

Assume, after Hamilton, that individuals A, B and C are located on a circular lily pond [2]. Instead of delaying or anticipating their signal phase/timing, they try to hide in-between other individuals. For instance, A occupies an isolated position on the pond, making it vulnerable to predators. A therefore seeks to decrease its 'domain of danger': if B and C are closer to each other than A is to any of them, A will jump and land between B and C [2].

Once formulated in these terms, it is clear how the mechanics of Hamilton's spatial predation model map one-to-one onto the acoustic signalling mechanism
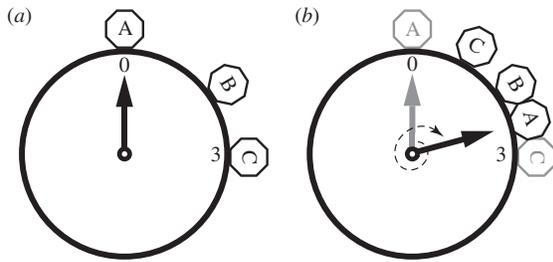
Figure 1. Clocks showing, for each individual, signalling time in two contiguous periods. Individual A signalled at 00.00, B at 02.00 and C at 03.00 (*a*). As agents choose when to call simultaneously, in the next time period (*b*), A remains silent for longer than a whole round (dashed thin arrow), postponing its call to 02.30. Similarly, C shifts its signal to 01.00. Adapted from [2].

Table 1. Comparison between parameters from the one-dimensional model in [2] and the mathematically equivalent, context-dependent signalling framework sketched here.

gregarious behaviour [2]	context-dependent signalling (present model)
circular lily pond	time period (1 unit)
moving agent (e.g. frog)	acoustic signal produced by agent
agent location	time of signal produced by agent
movement clockwise or counter-clockwise	signal delay or anticipation
distance between two agents	time elapsed between two agents' signals
'domain of danger'	'domain of silence': the amount of silence, measured in time units, preceding and following a call

sketched here (table 1). The original spatial model featured a closed, circular space. Circular metaphors (e.g. clocks) are also appropriate to represent periodic events, and the 'circular' feature in [2] enables its direct application to periodic signals over time, as required in models of chorusing.

In both cases, a general model is derived from applying the basic 'time shift' mechanism to all individuals (cf. figure 1, A delays its call and C anticipates its), and dynamically over time (figure 1*a* versus *b*). At every time period, most individuals will have either changed location or adapted their calls, making previous decisions suboptimal and spurring individuals to compensate by jumping to a better location, or shifting the phase of their upcoming call to an acoustically denser period of time. Computer simulations for the predation model showed formation of clusters of individuals [2]. By analogy, group signalling dynamics should begin with randomly occurring individual calls scattered over time and converge towards a few, high-intensity acoustic peaks (produced by several near-synchronous individuals).

An additional, deeper mathematical link connects Hamilton's model of space with dynamical processes in time. Hamilton noted that only one initial configuration, three evenly spaced frogs, will prevent aggregation [2]; decades later, the mathematical investigation of rhythm and timing in biological systems found that the same initial configuration will prevent synchronization of oscillators in time [6].

Herbers admits that one volume cannot do full justice to Hamilton's genius, anticipating how his ideas will 'influence the field over the coming 50 years' [1]. Hopefully, as I show here, Hamilton's mathematical insights will inform future research on both rhythmic processes in humans, such as language and music, and context-dependent acoustic signalling in other species.

References

1. Herbers JM. 2013 50 years on: the legacy of William Donald Hamilton. *Biol. Lett.* **9**, 20130792. (doi:10.1098/rsbl.2013.0792)

2. Hamilton WD. 1971 Geometry for the selfish herd. *J. Theor. Biol.* **31**, 295–311. (doi:10.1016/0022-5193(71)90189-5)

3. Grafe TU. 1999 A function of synchronous chorusing and a novel female preference shift in an anuran. *Proc. R. Soc. Lond. B* **266**, 2331–2336. (doi:10.1098/rspb.1999.0927)

4. Ryan MJ, Tuttle MD, Taft LK. 1981 The costs and benefits of frog chorusing behavior. *Behav. Ecol. Sociobiol.* **8**, 273–278. (doi:10.1007/BF00299526)

5. Néda Z, Ravasz E, Brechet Y, Vicsek T, Barabási AL. 2000 Self-organizing processes: the sound of many hands clapping. *Nature* **403**, 849–850. (doi:10.1038/35002660)

6. Strogatz SH, Stewart I. 1993 Coupled oscillators and biological synchronization. *Sci. Am.* **269**, 102–109. (doi:10.1038/scientificamerican1293-102)

7 | Concluding Discussion

7.1 Overview and Methodological Discussion

This thesis provides contributions to the study of the evolution of primate cognition along three main axes: *methodological, experimental* and *theoretical.*

Methodologically, I develop software and hardware solutions useful in experiments and animal enrichment. Using Python[1] and ELAN[2] (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), I introduce a novel method to blind-code videos from preferential looking experiments. This is particularly useful for playback experiments in infants and non-human animals, as it circumvents all potential sources of bias when annotating behavior from videos. Apart from this method for perception experiments, I propose a new approach to the study of acoustic production in apes. In fact, I develop two novel devices enabling non-vocal acoustic production in chimpanzees.

Experimentally, I show that other primates are both sensitive to, and in principle can be enabled to produce, structures akin to those present in language and music. Namely, I show that squirrel monkeys can process structures with dependencies. Moreover, after developing novel sound production devices, I pilot test one of them with chimpanzees: a large number of apes use the device in a variety of ways.

On a *theoretical* level, I contribute to the debate on the evolutionary pressures which gave rise to language and music. Avoiding any explicit formula, I outline a mathematical model, which can be straightforwardly applied to signaling strategies of a "pre-musical" human ancestor. Along a parallel line, I suggest how a recent theory of language evolution, to appear in the journal *Behavioral and Brain Sciences*,

---

[1]www.python.org

[2]ELAN. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. Retrieved from http://tla.mpi.nl/tools/tla-tools/elan/

can be equally well applied to the evolution of human musicality. This is particularly suggestive for the relationship between language and music, in general, and for the validity of Ackermann et al. (in press)'s evolutionary account.

7.2   Joint Conclusions

Language and music have a number of common features (Fitch, 2006), with cultural transmission and complexity among the most prominent (Kirby, 2009). This thesis, focusing on the complexity side of both, studies the evolution of the human cognitive abilities to produce and process complex structures in time (Fitch, 2013).

A theme present in most studies described here is human uniqueness in relation to evidence coming from non-human animal experiments. A general interim conclusion of my research is that most individual and group behaviors which I studied and which underlie human cognition for music and language can be found in other species.

The research presented in this thesis is consistent with, and will hopefully contribute to, three major scientific turns of the disciplines it encompasses.

First, similarly to what happened historically for the study of human behavior and psychology, a second cognitive revolution is taking place in the study of non-human animals. Animal cognition is a growing and exciting field, and experimental data and models are replacing previous conjectures and speculations.

Second, the study of language origins and evolution has especially flourished in the last decade. In particular, an interdisciplinary volume edited by Christiansen and Kirby (2003) has provided the ground and theoretical framework for decades of research to come. The present thesis builds upon and further develops the comparative (Hauser & Fitch, 2003) and social (Dunbar, 2003) approaches, with attention for issues of domain general cognition and learning (Kirby & Christiansen, 2003).

Finally, the study of music cognition is still in its infancy, and research on the origins of cognitive abilities underlying music is also just commencing. Inspired by language research, my thesis adapts similar methodologies to the study of music.

7.3 Contribution of the Thesis to the Field

After Fitch & Hauser (2004), the debate on pattern learning in non-human animals has been mostly centered on computational complexity vs. evolutionary history (Gentner, Fenn, Margoliash, & Nusbaum, 2006; Stobbe, Westphal-Fitch, Aust, & Fitch, 2012; van Heijningen, de Visser, Zuidema, & ten Cate, 2009). In other words, it might be that the ability to process structures of increasing mathematical complexity positively correlates with (evolutionary) proximity to humans, either by homology (common ancestry) or analogy (independent evolutionary histories). My paper (chapter 2) provides a fresh contribution to the debate. The pattern I test in squirrel monkeys exhibits a low computational complexity but nonetheless captures interesting linguistic phenomena. Moreover, it is the first "grammar learning" experiment in squirrel monkeys. Considering the already available knowledge of this species' nervous system, my results suggest that squirrel monkeys are a promising species for future (non-invasive) pattern learning research.

The field of animal behavior and cognition is growing at a fast pace. Cognitive research contributes to better understanding of animal welfare issues, increasing the importance of non-invasive methodologies to rigorously study animal minds. My methodological paper (chapter 3) provides a bias-free methodology to code animal reactions from videos. My method only uses freely available software and open-source Python scripts, making it accessible to every interested researcher in the field.

The theoretical article in chapter 4 constitutes a contribution to the fields of music and language evolution by integrating empirical evidence and proposing an original theoretical view on the evolution of speech and music. In particular, the journal's format encourages debate by presenting a target article, tens of short commentaries, and responses to said commentaries. The theoretical considerations I provide in chapter 4 will satisfy two foreseeable purposes. They will spur the authors of the target article to refine their evolutionary arguments, especially by making them aware of the experimental evidence on language and music origins coming from non-human animal studies. Moreover, my theoretical considerations will hopefully encourage other researchers to investigate different species' abilities to entrain and produce rhythmic patterns.

Rhythmic patterning and abilities to synchronize might have been crucial for social bonding in human evolutionary history (Fitch, 2009). In order to investigate ancestral states of this cognitive ability, the study of our closest relatives, the chimpanzees, is of utmost importance. Additionally, vocal learning abilities, which constitute a precondition for human speech, have been hypothesized to share neural circuitry for entraining to isochronous stimuli (Patel, 2006). A number of species has been tested, but the evidence for rhythmic abilities in apes is extremely limited (Hattori, Tomonaga, & Matsuzawa, 2013). Chimpanzees have been observed and studied while drumming and chorusing in the wild (Arcadi, Robert, & Boesch, 1998; Arcadi, Robert, & Mugurusi, 2004; de Waal, 1988). In zoos, their displays are limited, and this has been attributed to social factors (Kirschner, Call, & Fitch, 2009). Alternatively, chimpanzees might simply be missing the right resonant surfaces needed to (i) spur them to initiate the motor action, and (ii) convey a satisfactory sound (Ravignani et al., 2013). My research in chapter 5 introduces two sound-producing electronic devices specifically designed for chimpanzees. These were developed with the desiderata of: (i) motivating chimpanzees to interaction, (ii) enabling the production of virtually any conceivable sound, as selected by the experimenter, and (iii) allowing rigorous, direct and automatic data collection, hence preventing the need to code data from audio or video files.

Together with similar advances for automated electronic cognitive testing in the field of comparative cognition (Fagot & Bonté, 2010; Fagot & Paleressompoulle, 2009; Steurer, Aust, & Huber, 2012), I am sure that my methodology will be crucial in contributing to a technological turn in animal behavior. Additionally, my devices have been piloted with chimpanzees and will soon be used in rhythmic pattern production experiments.

The contribution provided by my mathematical commentary (chapter 6) is two-fold. First, it provides a new theoretical view on chorusing and inter-dependent group behavior in animal signaling. In particular, the framework I propose is particularly suited to link animal chorusing with musical rhythm and its origins in pre-musical hominids. Together with other published (Ravignani, Bowling, & Kirby, in press) and ongoing work, I am working towards a quantitative framework for the evolution of rhythm in primate societies. Additionally, chapter 6 shows that theoretical and

modeling research does not need to invent new tools and concepts for every new problem. On the contrary, and similarly to what happens for instance in statistical physics, old and well-established insights can be creatively applied to solve new, unrelated problems.

References

Arcadi, A. C., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, 39(4), 505-518.

Arcadi, A. C., Robert, D., & Mugurusi, F. (2004). A comparison of buttress drumming by male chimpanzees from two populations. *Primates*, 45(2), 135-139.

Christiansen, M. H., & Kirby, S. (2003). *Language evolution*: Oxford University Press.

de Waal, F. B. (1988). The communicative repertoire of captive bonobos (Pan paniscus), compared to that of chimpanzees. *Behaviour*, 183-251.

Dunbar, R. I. (2003). The origin and subsequent evolution of language *Language evolution*: Oxford University Press.

Fagot, J., & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio). *Behavior research methods*, 42(2), 507-516.

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior research methods*, 41(2), 396-404.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173-215.

Fitch, W. T. (2009). The biology and evolution of rhythm: Unraveling a paradox *Language and Music as Cognitive Systems*: Oxford University Press. Oxford, UK.

Fitch, W. T. (2013). Rhythmic Cognition in Humans and Animals: Distinguishing Meter and Pulse Perception. [Hypothesis & Theory]. *Frontiers in Systems Neuroscience*, 7.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377-380.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440(7088), 1204-1207.

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific reports*, 3.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? *Language evolution*: Oxford University Press.

Kirby, S. (2009). Darwin's musical protolanguage: an increasingly compelling picture *Language and Music as Cognitive Systems* (p. 96): Oxford University Press. Oxford, UK.

Kirby, S., & Christiansen, M. H. (2003). From language learning to language evolution *Language evolution*: Oxford University Press.

Kirschner, S., Call, J., & Fitch, W. T. (2009). *Drumming in African great apes: On the origins of instrumental music.* Poster presented at the 7th conference of the European Society for the Cognitive Sciences of Music, Jyväskylä.

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

Ravignani, A., Bowling, D., & Kirby, S. (in press). *The psychology of biological clocks: a new framework for the evolution of rhythm.* Paper presented at The 10th International Conference on the Evolution of Language, Vienna, Austria. Singapore: World Scientific Publishing Co.

Ravignani, A., Gingras, B., Asano, R., Sonnweber, R., Matellán, V., & Fitch, W. T. (2013). *The Evolution of Rhythmic Cognition: New Perspectives and Technologies in Comparative Research.* Proceedings of the 35th Annual Conference of the Cognitive Science Society, Berlin, Germany.

Steurer, M. M., Aust, U., & Huber, L. (2012). The Vienna comparative cognition technology (VCCT): An innovative operant conditioning system for various species and experimental procedures. *Behavior research methods*, 44(4), 909-918.

Stobbe, N., Westphal-Fitch, G., Aust, U., & Fitch, W. T. (2012). Visual artificial grammar learning: comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1995-2006.

van Heijningen, C. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci USA*, 106(48), 20538-20543.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006, 2006). *ELAN: a professional framework for multimodality research, Fifth International Conference on Language Resources and Evaluation.* Paper presented at the Proceedings of LREC.

8 | Appendix

8.1 Publications Contributing to the Thesis

Chapter 2

> Ravignani, A., Sonnweber, R.-S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biology Letters*, 9, 20130852.

Andrea Ravignani designed and performed the experiments, analyzed the data and was the main author of the manuscript.

Chapter 3

> Ravignani, A. & Fitch, W.T. (2012) Sonification of experimental parameters as a new method for efficient coding of behavior. *Proceedings of Measuring Behavior 2012*, 8th International Conference on Methods and Techniques in Behavioral Research (Utrecht, The Netherlands, August 28-31, 2012) Edited by A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman ISBN 978-90-74821-87-2.

Andrea Ravignani developed the methodology, wrote new Python scripts or adapted existing ones, and wrote the manuscript.

Chapter 4

> Ravignani A.*, Martins, M.* & Fitch, W. T. (in press). Vocal learning, prosody and basal ganglia: Don't underestimate their complexity. *Be-*

havioral and Brain Sciences. [*Andrea Ravignani and Mauricio Martins contributed equally to this commentary as joint first authors.]

Andrea Ravignani wrote the first section of the manuscript, concerning the evolution of vocal learning and rhythm. Although the full manuscript is attached, the first section is the only one discussed in the present thesis.

The manuscript has been accepted for publication in Behavioral and Brain Sciences on November, 21st, 2013. It will appear as a comment on the following target article:

Ackermann, H, Hage SR & Ziegler, W. (in press). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences.* Edited by: Cambridge University Press.

Full copyright has been transferred and belongs to Cambridge University Press.

Chapter 5

Ravignani, A., Olivera, V. M., Gingras, B., Hofer, R., Hernández, C. R., Sonnweber, R.-S., & Fitch, W. T. (2013). Primate Drum Kit: A System for Studying Acoustic Pattern Production by Non-Human Primates Using Acceleration and Strain Sensors. *Sensors*, 13(8), 9790-9820.

Andrea Ravignani developed the methodology, wrote the software, assembled the prototype devices, performed the calibration and pilot experiments, analyzed the calibration and experimental data and was the main author of the manuscript.

Chapter 6

Ravignani, A. (2014). Chronometry for the chorusing herd: Hamilton's legacy on context-dependent acoustic signalling - a comment on Herbers (2013). *Biology Letters*, 10(1), 20131018.

Andrea Ravignani was the sole author of this research and manuscript.

8.2   Dissemination

The research described in this thesis has been or will be presented at the following events as oral (*) or poster presentation:

- International conference on the Evolution of Language, University of Vienna, Austria, 04/2014.*

- International Conference on Timing and Time Perception, Ionian Academy, Corfu, GR, 03/2014.*

- Computational Linguistics & Psycholinguistics Seminar, University of Antwerp, BE, 01/2014.*

- Behavioral Biology Seminar, University of Leiden, The Netherlands, 10/2013.*

- Computational Linguistics Seminar, University of Amsterdam, The Netherlands, 10/2013.*

- Summer school: Music and the origins of language, Cortona, Italy, 09/2013.

- Annual Meeting of the Cognitive Science Society, Humboldt University Berlin, Germany, 08/2013.*

- International Congress of Linguists, University of Geneva, Switzerland, 07/2013.**

- Language Evolution & Computation Seminar, UK, 06/2013.*

- Stirling Behaviour and Evolution Research Group (BERG Seminar), UK, 04/2013.*

- Language Evolution & Computation Seminar, UK, 03/2013.*

- Language at Edinburgh Lunch, UK, 02/2013.

- Annual Meeting of the Scottish Primate Research Group, UK, 01/2013.*

- Language Evolution & Computation Graduate Retreat, UK, 01/2013.*

- Behavioral Biology Seminar, University of Leiden, The Netherlands, 09/2012.*

- Computational Linguistics Seminar, University of Amsterdam, The Netherlands, 09/2012.*

- Measuring Behavior 2012, Utrecht University, The Netherlands, 08/2012.*

- International Summer School in Systematic Musicology, European Commission Erasmus Intensive Programme, University of Jyvaskyla, Finland, 08/2011.

8.3   Curriculum Vitae

Education

- PhD candidate

  Dept. of Cognitive Biology, *University of Vienna* (2011-present).

  Advisors: Prof. W. Tecumseh Fitch (Cognitive Biology, Vienna), Prof. Simon Kirby (Language Evolution and Computation Research Unit, Edinburgh).

- MSc, Mathematics, Informatics and Applications

  Université Paris 1, Panthéon-Sorbonne (2010)

- MSc *Erasmus Mundus*, Economic Analysis

  Universitat Autònoma de Barcelona (2010)

- MSc, Economics

  Universitá Ca' Foscari di Venezia ("equipollenza a Laurea Magistrale", 2010)

- BSc, Mathematics

  Universitá di Roma La Sapienza (2007)

Exchanges and Research Visits

- Research visit, *Computational Linguistics & Psycholinguistics Center, University of Antwerp*, BE (01-02/2014)

- Research visit, *Center for the Study of Rationality, The Hebrew University of Jerusalem*, IL (11-12/2013)

- Research visit, *Institute for Logic, Language and Computation, University of Amsterdam*, NL (10/2013)

- Chimpanzee project, *Budongo Trail, Royal Zoological Society of Scotland*, UK (06-07/2013)

- Visiting Postgraduate Research Student, *Language Evolution and Computation Research Unit, University of Edinburgh*, UK (01-06/2013)

- Exchange visit, *Robotics Dept., Universidad de León*, ES (10-12/2012)

- Chimpanzee project, *The University of Texas*, Austin, US (10-11/2011)

- Graduate courses, *Universität Bielefeld*, DE (02-07/2009)

- Exchange, graduate and undergraduate courses, *Universidad Complutense de Madrid*, ES (09/2005-09/2006)

Additional Publications (peer-reviewed)

- Ravignani, A., Bowling, D. & Kirby, S. (in press) The psychology of biological clocks: a new framework for the evolution of rhythm. *The Evolution of Language: Proceedings of the 10th international conference.* Edited by: Erica A Cartmill, Seán Roberts, Heidi Lyn and Hannah Cornish. Publisher: Singapore: World Scientific Publishing Co.

- Martins, M., Raju, A. & Ravignani, A. (in press) Evaluating the role of quantitative modeling in language evolution. *The Past, Present and Future of Language Evolution Research.* ISBN: 978-4-9906340-1-8.

- Ravignani, A.; Gingras, B.; Asano, R.; Sonnweber, R.; Matellán, V.; Fitch, W.T. (2013) The Evolution of Rhythmic Cognition: New Perspectives and Technologies in Comparative Research. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Ed: Cognitive Science Society (pp. 1199-1204).

8.4   Zusammenfassung

Menschliche Sprache und Musik unterscheiden sich von der Kommunikation nicht-menschlicher Tiere in mehreren strukturellen Aspekten. Eine wichtige kognitive Voraussetzung für Sprachwahrnehmung und -produktion ist die Fähigkeit Muster und Mustersystematik zu verstehen und produzieren zu können. Ich verwende einen vergleichenden tierpsychologischen Ansatz um die Evolution kognitiver Fähigkeiten, die im Zusammenhang mit Rhythmus und Mustersequenzierung stehen, zu erforschen. Experimentelle Daten von Menschenaffen und Affen - gewonnen durch Einsatz innovativer Methoden - werden mit mathematischen Modellen kombiniert. So werden theoretische Erkenntnisse in experimentelle Ergebnisse integriert. Zunächst zeige ich, dass schon eine Spezies von Neuweltaffen die kognitive Fähigkeit besitzt sensorische Abhängigkeiten in musikalisch-auditiven Reizen zu verarbeiten. Dieses Ergebnis zeigt, dass diese psychologische Fähigkeit schon in unserem sprachlosen letzten gemeinsamen Vorfahren mit Totenkopfaffen vorhanden war und erst später in der menschlichen Evolution für Sprache und die Verarbeitung musikalischer Struktur herangezogen wurde. Musterproduktions- und Musterwahrnehmungsfähigkeiten beim Menschen unterscheiden sich und befinden sich nicht notwendigerweise in gegenseitiger Abhängigkeit. Somit kann vermutet werden, dass diese Kluft auch in anderen Arten bestehen könnte. Um dies zu untersuchen habe ich ein "Primate Drum Kit" entwickelt. Mit diesem Gerät können Schimpansen spontan nicht-vokale, rhythmische Muster produzieren. Meine Methode wird dazu erlauben, Produktionsmuster an der Grenze der sprachlichen Phonologie und musikalischen Rhythmus bei Schimpansen zu testen. Ich diskutiere wie vokales Lernen - eine essentielle Voraussetzung für die menschliche Sprache - in Zusammenhang mit rhythmischen Fähigkeiten steht. Basierend auf Tierstudien und Modellen kann argumentiert werden, dass vokale und rhythmische, nicht-vokale Produktion beim Menschen unter evolutionärem sozialen Druck entstanden sind. Schliesslich erforsche ich Rhythmus auf Gruppenebene mittels eines mathematischen Modells für zeitabhängige Signalisierungen. Dieses Modell, welches ursprünglich zur Untersuchung von Synchronie bei Grillen und der Aggregation von Fröschen erstellt wurde, kann für die Evolution von Rhythmus in Hominiden angepasst werden. Dies wird erlauben die Konsistenz der Hypothesen zur sozialen

Herkunft von Sprache und Musik zu untersuchen. Die empirischen Daten, Methodologien und theoretischen Erkenntnisse dieser Dissertation werden zu laufenden Untersuchungen über die Ursprünge und die Entwicklung von Musik und Sprache beitragen.

8.5 Abstract

Human language and music differ from non-human animal communication in several structural respects. Accordingly, cognitive abilities for patterning are crucial in enabling humans to perceive and produce language and music. Here I adopt a comparative approach to study the evolution of human cognitive abilities for rhythm and pattern sequencing. Experimental data from apes and monkeys - tested using novel methodological paradigms - are integrated with mathematical modeling. First, I show that a New World primate is capable of processing sensory dependencies in music-like auditory stimuli. This finding suggests that this psychological ability was present in our language-less last common ancestor and, later in human evolution, was employed for language and musical structure processing. Second, as pattern production and perception abilities have been shown to differ in humans, the same divide could exist in other species. The ape-specific devices I developed will allow chimpanzees to spontaneously produce non-vocal, possibly rhythmic acoustic patterns. This method will serve to test pattern production at the boundary of linguistic phonology and musical rhythm in chimpanzees, and to explore the production-perception divide in our closest living relatives. Third, I discuss how vocal learning in non-human animals, a necessary precondition for human speech, relates to rhythmic abilities. In particular, both vocal and non-vocal rhythmic acoustic production could have arisen in humans under social evolutionary pressures. Finally, to investigate rhythm at a group level, I introduce a mathematical framework of time-dependent signaling. This model, originally developed to investigate synchrony in crickets and spatial aggregation in frogs, can be adapted to hominid evolutionary history, so to assess consistency of hypotheses on the social origins of language and music. The empirical evidence, methodological tools and theoretical insights provided in this thesis will significantly contribute to ongoing research on the origins and evolution of music and language.