

Fachbereich Philosophie und Geisteswissenschaften
der Freien Universität Berlin

The phylogenetic origin and mechanism of sound
symbolism - the role of action-perception circuits

Dissertation

zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

vorgelegt von

Konstantina Margiotoudi

Berlin 2020

Ersgutachter: Prof. Dr. Dr. Friedemann Pulvermüller

Zweitgutachterin: Prof. Dr. Christine Mooshammer

Tag der Disputation: 11. Dezember 2020

Table of Contents

Abstract	7
Zusammenfassung	9
List of abbreviations	12
1. General Introduction	13
1.1. Semantic theories	16
1.2. Iconicity in semiotic theories	18
1.2.1. Iconicity in spoken language	20
1.2.2. Sound symbolism	21
1.2.3. The case of “maluma-takete” mappings	22
1.3. Why does sound symbolism matter?	23
1.3.1. Language acquisition	25
1.3.2. Language evolution	29
1.4. Possible mechanisms behind sound symbolism	31
1.5. Actions: the missing element behind sound symbolism	33
1.6. Focus of the present dissertation	34
2. Testing sound symbolic mappings, pitch-shape and pitch-spatial position correspondences in a two-alternative forced choice task	37
2.1. Introduction	38
2.2. Materials and Methods	42
2.3. Data analysis	45

2.4. Results	48
2.5. Discussion	52
3. Sound symbolic congruency detection in humans but not in great apes	60
3.1. Introduction	61
3.2. Experiment 1	65
3.2.1. Materials and Methods	65
3.2.2. Data analysis	68
3.2.3. Results	69
3.3. Experiment 2	71
3.3.1. Materials and Methods	71
3.3.2. Data analysis	73
3.3.3. Results	74
3.4. Interim Discussion	75
3.5. Experiment 3	78
3.5.1. Materials and Methods	78
3.5.2. Data analysis	79
3.5.3. Results	79
3.6. Discussion	80
4. Action sound-shape congruencies explain sound symbolism	88
4.1. Introduction	89
4.2. Materials and Methods	95
4.3. Data analysis	103
4.4. Results	105
4.5. Discussion	109
4.6. Preliminary studies	120
4.6.1. Study 1	120
4.6.2. Study 2	125

5. General Discussion	135
5.1. Summary of findings	135
5.1.1. Chapter 2	135
5.1.2. Chapter 3	136
5.1.3. Chapter 4	137
5.2. Interpretation of findings	138
5.2.1. Action-perception circuits (APCs) and the arcuate fasciculus . . .	140
5.2.2. APCs: carriers of sound symbolism	142
5.2.3. Theoretical implications for language evolution	145
5.2.4. From crossmodal correspondences to sound symbolism	148
5.3. Limitations and perspectives	150
5.4. Conclusion	153
References	153
A. Appendix Chapter 2	176
B. Appendix Chapter 3	180
C. Appendix Chapter 4	185
List of publications	193
Erklärung	194

Acknowledgements

I would like to start by expressing my profound gratitude to my mentor and supervisor Prof. Friedemann Pulvermüller for his support and motivation during all this experience. Thank you very much, it was a great chance for me to work under your guidance. I would like also to thank my collaborators Dr. Manuel Bohn and Dr. Matthias Allritz for their warm welcome in Leipzig and their help in the world of R and of apes! Also special thanks to my colleagues at the brain language laboratory (BLL) at the Frei Universität for all the scientific (and non scientific) chats that we shared and for your guidance during the PhD adventure! Thank you guys! I would like to thank also the Berlin School of Mind and Brain and the Onassis Foundation for funding the present work. I cannot miss to thank my cohort 10 from the Berlin School of Mind and Brain. You are special each one of you and you turned this PhD journey to the most exotic experience I could have ever imagined! Double D! Spatial thanks to the Franco-Italian group of DLR, for all the dinners and PCs we shared! A big thank you also goes to the world of theater and to all the people we rehearsed and performed together! I would like to thank Ioanna, who has always been a great friend to me. Also big thanks to my friends Georgia and Paraskevi for all the great laughs and chats. My biggest gratitude goes to three special women in my life for all their support and love! Efcharisto Dafni, Efaki, Dora! I want also to say a thank you to Hugo for being next to me and for making everyday warmer and sweeter!

Finally, I would like to dedicate this work to my beloved Kostas and Frideriki for guarding my dreams.

Abstract

As opposed to the classic Saussurean view on the arbitrary relationship between linguistic form and meaning, non-arbitrariness is a pervasive feature in human language. Sound symbolism—namely, the intrinsic relationship between meaningless speech sounds and visual shapes—is a typical case of non-arbitrariness. A demonstration of sound symbolism is the “maluma-takete” effect, in which immanent links are observed between meaningless ‘round’ or ‘sharp’ speech sounds (e.g., maluma vs. takete) and round or sharp abstract visual shapes, respectively. An extensive amount of empirical work suggests that these mappings are shared by humans and play a distinct role in the emergence and acquisition of language. However, important questions are still pending on the origins and mechanism of sound symbolic processing. Those questions are addressed in the present work.

The first part of this dissertation focuses on the validation of sound symbolic effects in a forced choice task, and on the interaction of sound symbolism with two crossmodal mappings shared by humans. To address this question, human subjects were tested with a forced choice task on sound symbolic mappings crossed with two crossmodal audiovisual mappings (pitch-shape and pitch-spatial position). Subjects performed significantly above chance only for the sound symbolic associations but not for the other two mappings. Sound symbolic effects were replicated, while the other two crossmodal mappings involving low-level audiovisual properties, such as pitch and spatial position, did not emerge.

The second issue examined in the present dissertation are the phylogenetic origins

of sound symbolic associations. Human subjects and a group of touchscreen trained great apes were tested with a forced choice task on sound symbolic mappings. Only humans were able to process and/or infer the links between meaningless speech sounds and abstract shapes. These results reveal, for the first time, the specificity of humans' sound symbolic ability, which can be related to neurobiological findings on the distinct development and connectivity of the human language network.

The last part of the dissertation investigates whether action knowledge and knowledge of the perceptual outputs of our actions can provide a possible explanation of sound symbolic mappings. In a series of experiments, human subjects performed sound symbolic mappings, and mappings of 'round' or 'sharp' hand actions sounds with the shapes produced by these hand actions. In addition, the auditory and visual stimuli of both conditions were crossed. Subjects significantly detected congruencies for all mappings, and most importantly, a positive correlation was observed in their performances across conditions. Physical acoustic and visual similarities between the audiovisual by-products of our hand actions with the sound symbolic pseudowords and shapes show that the link between meaningless speech sounds and abstract visual shapes is found in action knowledge. From a neurobiological perspective the link between actions and the audiovisual by-products of our actions is also in accordance with distributed action-perception circuits in the human brain. Action-perception circuits, supported by the human neuroanatomical connectivity between auditory, visual, and motor cortices, and under associative learning, emerge and carry the perceptual and motor knowledge of our actions. These findings give a novel explanation for how symbolic communication is linked to our sensorimotor experiences.

To sum up, the present dissertation (i) validates the presence of sound symbolic effects in a forced choice task, (ii) shows that sound symbolic ability is specific to humans, and (iii) that action knowledge can provide the mechanistic glue of mapping meaningless speech sounds to abstract shapes. Overall, the present work contributes to a better understanding of the phylogenetic origins and mechanism of sound symbolic ability in humans.

Zusammenfassung

Im Gegensatz zur klassischen Saussureschen Ansicht über die willkürliche Beziehung zwischen sprachlicher Form und Bedeutung ist die Nichtwillkürlichkeit ein durchdringendes Merkmal der menschlichen Sprache. Lautsymbolik—nämlich die intrinsische Beziehung zwischen bedeutungslosen Sprachlauten und visuellen Formen—ist ein typischer Fall von Nichtwillkürlichkeit. Ein Beispiel für Klangsymbolik ist der “malumataketete” Effekt, bei dem immanente Verbindungen zwischen bedeutungslosen ‘runden’ oder ‘scharfen’ Sprachlauten (z.B. maluma vs. takete) und runden bzw. scharfen abstrakten visuellen Formen beobachtet werden. Umfangreiche empirische Arbeiten legen nahe, dass diese Zuordnungen von Menschen vorgenommen werden und bei der Entstehung und dem Erwerb von Sprache eine besondere Rolle spielen. Wichtige Fragen zu Ursprung und Mechanismus der Verarbeitung von Lautsymbolen sind jedoch noch offen. Diese Fragen werden in der vorliegenden Arbeit behandelt.

Der erste Teil dieser Dissertation konzentriert sich auf die Validierung von klangsymbolischen Effekten in einer Forced-Choice-Auswahlaufgabe (*erzwungene Wahl*) und auf die Interaktion von Klangsymbolik mit zwei crossmodalen Mappings, die von Menschen vorgenommen werden. Um dieser Frage nachzugehen, wurden menschliche Probanden mit einer Auswahlaufgabe mit zwei Auswahlmöglichkeiten auf klangsymbolische Zuordnungen getestet, die mit zwei crossmodalen audiovisuellen Zuordnungen (Tonhöhenform und Tonhöhen-Raum-Position) gekreuzt wurden. Die Versuchspersonen erbrachten nur bei den klangsymbolischen Assoziationen eine signifikant über dem Zufall liegende Leistung, nicht aber bei den beiden anderen Zuordnungen. Tonsymbolische Effekte wurden

repliziert, während die beiden anderen crossmodalen Zuordnungen, die audiovisuelle Eigenschaften auf niedriger Ebene wie Tonhöhe und räumliche Position beinhalteten, nicht auftraten.

Das zweite Thema, das in der vorliegenden Dissertation untersucht wird, sind die phylogenetischen Ursprünge der klangsymbolischen Assoziationen. Menschliche Versuchspersonen und eine Gruppe von Menschenaffen, die auf Touchscreens trainiert wurden, wurden mit einer Forced-Choice-Aufgabe auf klangsymbolische Zuordnungen getestet. Nur Menschen waren in der Lage, die Verbindungen zwischen bedeutungslosen Sprachlauten und abstrakten Formen zu verarbeiten und/oder abzuleiten. Diese Ergebnisse zeigen zum ersten Mal die Spezifität der lautsymbolischen Fähigkeit des Menschen, die mit neurobiologischen Erkenntnissen über die ausgeprägte Entwicklung und Konnektivität des menschlichen Sprachnetzwerks in Verbindung gebracht werden kann.

Der letzte Teil der Dissertation untersucht darüber hinaus, ob Handlungswissen und das Wissen um die Wahrnehmungsergebnisse unserer Handlungen eine mögliche Erklärung für solide symbolische Mappings liefern können. In einer Reihe von Experimenten führten menschliche Versuchspersonen klangsymbolische Mappings durch sowie Mappings von ‘runden’ oder ‘scharfen’ Handaktionen Klänge mit den durch diese Handaktionen erzeugten Formen. Darüber hinaus wurden die auditiven und visuellen Reize beider Bedingungen gekreuzt. Die Probanden stellten bei allen Zuordnungen signifikant Kongruenzen fest, und, was am wichtigsten war, es wurde eine positive Korrelation ihrer Leistungen unter allen Bedingungen beobachtet. Physikalische akustische und visuelle Ähnlichkeiten zwischen den audiovisuellen Nebenprodukten unserer Handaktionen mit den klangsymbolischen Pseudowörtern und Formen zeigen, dass die Verbindung zwischen bedeutungslosen Sprachlauten und abstrakten visuellen Formen im Handlungswissen zu finden ist. Aus neurobiologischer Sicht stimmt die Verbindung zwischen Handlungen und den audiovisuellen Nebenprodukten unserer Handlungen auch mit den verteilten Handlungs- und Wahrnehmungskreisläufen im menschlichen Gehirn überein. Aktions-Wahrnehmungsnetzwerken, die durch die neuroanatomische Konnektivität zwischen auditorischen, visuellen und motorischen kortikalen Arealen des Menschen unterstützt wer-

den, entstehen und tragen unter assoziativem Lernen das perzeptuelle und motorische Wissen unserer Handlungen. Diese Erkenntnisse geben eine neuartige Erklärung dafür, wie symbolische Kommunikation in unseren sensomotorischen Erfahrungen verknüpft ist.

Zusammenfassend lässt sich sagen, dass die vorliegende Dissertation (i) das Vorhandensein von lautsymbolischen Effekten in einer Forced-Choice-Aufgabe validiert, (ii) zeigt, dass lautsymbolische Fähigkeiten spezifisch für Menschen sind, und (iii) dass Handlungswissen den mechanistischen Klebstoff liefern kann, um bedeutungslose Sprachlaute auf abstrakte Formen abzubilden. Insgesamt trägt die vorliegende Arbeit zu einem besseren Verständnis der phylogenetischen Ursprünge und des Mechanismus der lautsymbolischen Fähigkeit des Menschen bei.

List of abbreviations

- AF.....arcuate fasciculus
APC.....action-perception circuit
CI.....confidence interval
CVCV.....consonant-vowel-consonant-vowel
F0.....fundamental frequency
FDR.....false discovery rate
GLMM.....generalized linear mixed modal
IFOF.....inferior fronto-occipital fasciculus
LRT.....likelihood ratio test
M.....mean
PSD.....power spectral density
SD.....standard deviation
SE.....standard error
SFL.....superior longitudinal fasciculus
TMS.....trascranial magnetic stimulation
VWM.....verbal working memory

1. General Introduction

Σωκράτης: Γιὰ τοῦ Ἴππωνικοῦ καὶ Ἑρμογένη, μιὰ παλιὰ παροιμία λέει ὅτι « τὰ καλὰ πράγματα εἶναι δύσκολα», ὅταν πρόκειται νὰ μάθουμε τὴ φύση τους καὶ μάλιστα ἢ σπουδὴ γιὰ τὰ ὀνόματα δὲν τυχαίνει νὰ εἶναι εὐκόλη δουλειά.

- Πλάτων, *Κρατύλος* (360 π.Χ.)

Socrates Hermogenes, son of Hipponicus, there is an ancient saying that knowledge of high things is hard to gain; and surely knowledge of names is no small matter.

- Plato, *Cratylus* (360 BC)

(translation by Harold N. Fowler)

A central question that has troubled the scientists working on the origins of human language is the emergence of linguistic symbols. Scientists are searching to understand how humans learned to associate linguistic symbols to communicate about separate pieces of information available in the complex environment they live in, in their internal states and through their experiences. Linguistic symbols in the form of sequences of speech sounds (words) traditionally convey meanings in a totally arbitrary manner (Saussure, 1959). For instance, any word can stand equally well to express the concept of a *chair*, such as the German word '*Stuhl*' or the French word '*chaise*'. How linguistic symbols can convey meaning about perceived objects or actions, and how can we explain the emergence of such a system? This problem is well described by Harnad (1990) as the "symbol grounding problem". Harnad (1990) used the thought experiment of the Chinese room to explain this problem (Searle, 1980): if an English speaker, for example, has to learn Chinese as a second language by using only a Chinese dictionary, each Chinese word will be defined by another Chinese word. Consequently, the English speaker will be in vicious circles, without being able to have access to meanings, as the words

will refer to other symbols and not to things in the world. As such, Chinese symbols will be connected to other symbols that are meaningless to the English speaker. The thought experiment of the Chinese room depicts the challenge to conceive the meaning of a symbol when it is linked to other arbitrary symbols. To be able to convey semantic meanings of symbols it is necessary that the meanings are intrinsically connected to the system, namely words should be grounded to our perceptual and motor experiences.

Iconicity both in the vocal and gestural domain can provide a solution to this symbol grounding problem. More specifically, vocal iconicity —defined as the resemblance between properties of linguistic form and meaning (Perniss and Vigliocco, 2014)—is proposed to allow for the direct links between linguistic symbols and the sensorimotor properties of the referents.

Classic theories on the origins of human vocal communication identified the importance of vocal iconicity in the form of onomatopoeia —the vocal imitation of different environmental sounds— in the emergence of spoken language in humans. An onomatopoeic example would be the reference to a cat by using the word "meow-meow", which resembles the sound that a cat produces. Darwin (1888) considered the origins of vocal communication to be in the imitation of animal and environmental sounds; hence, he suggested the potential of linguistic sounds to be “naturally” connected to their meanings. The same view was held by the “bow-wow” theories, placing imitation of auditory stimuli in the origins of human vocal communication (for an overview, see Aitchison, 2000)

Considered to be a type of vocal iconicity, there is another linguistic phenomenon, known as sound symbolism, during which meaningless speech sounds can express meaning in other sensory modalities, beyond the acoustic domain. Sound symbolism is an umbrella term that describes the non-arbitrary associations between meaningless speech sounds and sensory or other meanings (Hinton et al., 2006). The present dissertation deals with the most popular demonstration of sound symbolism, namely the "malumataketete" effect. Specifically, the gestalt psychologist Köhler (1929) first described the

phenomenon on how meaningless speech sounds (e.g., “maluma-takete”) can be perceived as ‘round’ or ‘sharp’ and express information about the roundness or the sharpness of a visual object.

The notions of the previous century saw the relation between linguistic form and meaning to be established through arbitrariness and cultural conventions. Nevertheless, a significant number of empirical studies over the last ten years have been focusing on the topic of non-arbitrariness and of sound symbolism in human language.¹ An interdisciplinary field of research has started to thoroughly investigate how linguistic sounds can have a natural, non-arbitrary connection with their meanings across different modalities, in the form of sound symbolism. That is, how meaningless speech sounds can directly convey meaning about certain perceptual and motor qualities of their referents?

The present chapter provides an overview of the theoretical and experimental work on sound symbolism and on its importance as a property of the linguistic ability. In the first part of this section, an overview of semantic theories will be presented. Given that sound symbolism is generally considered as a type of iconicity, in the next section will be given different definitions present in semiotics on what comprises an iconic signal, and hence iconicity. Thereafter, the different forms of iconicity found in spoken language are presented and it will be explained why sound symbolism cannot be captured by the current definitions of iconicity, and why it should be reconsidered as a case of iconicity. Next, some cases are described that are encompassed by the term sound symbolism, beyond the “maluma-takete” effect. After that, an overview is given of the crosslinguistic research in the field, and of the properties of pseudoword-shape mappings of the “maluma-takete” type. The following part deals with the theoretical proposals and the experimental findings that suggest how sound symbolism plays a critical role in spoken language acquisition and evolution. With respect to language acquisition, experimental

¹According to the Web of Science (ISI) from 2010-2020, there have been 438 publications on the topic of sound symbolism. In contrast, between the years 1945-2009 only 97 publications were reported on the same topic. However, this number does not correspond only to sound symbolism as described above, as sometimes is misused to refer also to works on ideophones.

studies in children and infants are presented that explain how sound symbolic ability can assist in word learning in the first years of life. As for language evolution, different arguments are described, for how sound symbolism could have assisted in the emergence of spoken language in the first humans. After that, different theories are presented that suggest possible mechanisms behind sound symbolic mappings, as well as a new theoretical model proposing action knowledge as a plausible explanation behind sound symbolic mappings. The final section describes the issues addressed by the present dissertation.

1.1. Semantic theories

A large body of linguistic inquiries has been dedicated to questions regarding the nature of linguistic meaning and on the relation between linguistic form and meaning. What are the linguistic meanings of the word ‘knife’ and the pseudoword ‘maluma’? These questions have been a debating point for centuries. In linguistics, the questions on the nature of meaning are subject of semantics (Löbner, 2013). As Charles Hockett stated, “This debate in semantics has been the source of more trouble than any other aspect of communicative behavior” (Hockett, 1959).

The debate on the origins of meaning goes back to antiquity. In Ancient Greece, we find *Cratylus*, the protagonist of the homonym dialog of Plato, arguing with Socrates that the relationship between the words and their meanings is natural/innate, namely that the structure of the words has a natural link to their meanings (Bestor, 1980). A different view comes from the Far East, with the Confucian philosopher Xun Zi (310 BC - 235 BC), suggesting that the relationship between words and their meanings is completely arbitrary, based on cultural convention, and that the sounds of the words could equally well express any meaning (as cited in Hinton et al., 2006). These two views, geographically and theoretically distant, illustrate well two different theoretical views on the relation between words and meanings.

For many years, the dominant view was in favor of an arbitrary relation between

linguistic form and meaning; that is the combination of any speech sound was believed to be able to signify anything. This traditional view is found in the writings of Ferdinand de Saussure, who described the linguistic system as a complex abstract system that consists of signs. The signs in turn have two components: the sound form (the signifier) and the meaning (the signified). This dyadic relation is established through “l’arbitraire du signe”, namely arbitrariness and convention (Saussure, 1959). Hockett also used the same term of arbitrariness as one of the 13 design features of language, which allows unlimited communication about any concept.

A different model proposed by Ogden et al. (1923) was based on a triadic and not dyadic relation between referent and sign, consisting of three components (the referent, the sign, and the thought of the referent). The relationship between the sign and the referent is established through the mediation of the mental concept, the thought of the referent. “Between the symbol and the referent there is no relevant relation, other than the indirect one, which consists in its being used by someone to stand for a referent.”

Other semantic theories focused on the aspect of concept and proposed that the concept of a word is composed of several semantic features. For example, the word ‘knife’ includes semantic features such as + tool, + metal, + object etc. According to this proposal, a concept is the result of a composition of different semantic features (Katz and Fodor, 1963). The principle of compositionality of the linguistic meaning is found in theories on modularity of human mind, which consider language to be processed in a distinct module in the human brain (Fodor, 1983). The problematic aspect of such theories is that they do not explain how these concepts are related to the real world (for a discussion, see Pulvermüller, 2018b). A solution to this problem is provided by embodied theories on language, suggesting that meaning includes representations of action or object schemas and processed in our sensory and motor systems and not to a distinct module in the human brain (Barsalou, 1999; Glenberg and Gallese, 2012; Pulvermüller, 1999, 2013a).

In sound symbolism, first, the relation between form and referent established via

arbitrariness or convention as described in the dyadic and triadic models of semantic meaning, does not apply. The relationship between the sound form (e.g., ‘maluma’) and the referent is not based on arbitrariness, as the sound form ‘maluma’ informs about the perceptual properties of the referent (e.g., round, smooth, and not sharp, edgy). Hence, the relationship between signified and signifier can not be explained by arbitrariness or convention. The mapping between ‘round’ or ‘sharp’ sounding pseudowords and abstract shapes requires a closer examination in order to understand how these links are established and how meaning is conveyed. Moreover, in contrast to the modular models of semantic theories, here an embodied view on the semantic processing of sound symbolism seems logical. Indeed, for a symbolic system to express meaning, it is important that the latter is linked to the real world through our perceptual and motor experiences with it (for a detailed discussion, see Pulvermüller, 2013b). The meaning needs to be connected to these sensorimotor experiences to solve the problem of symbol grounding (Harnad, 1990). As sound symbolic pseudowords and their sound forms trigger meaning about certain perceptual properties of the referents (e.g., round, curved object), and the link between sound form and meaning seems intuitive and non-arbitrary, embodied theories of language are most appropriate to support these mappings. Under this view, the meaning of the pseudoword ‘maluma’ would include the representations of round/curved objects and processed in the respective sensory and/or motor areas of the brain.

1.2. Iconicity in semiotic theories

Iconicity in language might be used as an umbrella term to express the natural resemblance between linguistic form and meaning, both in the vocal and in the gestural domain (Perniss and Vigliocco, 2014). In fact, there are different definitions given by semioticians to what iconicity is.

Charles Sanders Peirce, in his tripartite categorization of signs (icon, index, and symbol), described an icon as a sign that stands for something because it resembles it: “An Icon is a sign which refers to the Object that it denotes merely by virtue of characters

of its own, and which it possesses, just the same, whether any such Object actually exists or not” (Peirce, 1960). Another word used by Peirce to describe the icon is that of "likeness"; hence, the relation between an icon and object is based on the similarity between the form and referent.

According to Morris, who first used the term “iconicity”, an icon is defined based on shared properties between the referent and the sign. A sign can be iconic when it shares a collection of properties with the object it denotes. According to this definition, iconicity is a matter of degrees, with some iconic signs being more “iconic” because they share more properties with their donata than other iconic signs (Nöth, 1995).

On the other hand, Umberto Eco has questioned the general role of signs and highlighted the importance of sign-functions, calling into question previous definitions of iconicity. He developed a detailed criticism of iconicity and on the similarity/resemblance of properties shared between the sign and the referent. In fact, he suggested that the naturalness between form and meaning is a network of cultural stipulations/conventions. For instance, the drawing of a horse and the continuous line tracing of its profile are perceived as the abstract representation of a horse based on a cultural convention. Only a trained eye can perceive the profile of the horse, and this perception derives from cultural convention (Eco et al., 1976).

The different views and definitions with respect to an icon and iconicity are explained by the fact that iconicity can be expressed in different ways (e.g., gestural, vocal, drawings, etc.) and provide varied information. Despite the various definitions of iconicity, its existence is undoubtedly present in the linguistic domain, both in the gestural and vocal communication (Monaghan et al., 2014). Iconicity in the gestural domain is found in the form of iconic gestures (Kendon, 2004; McNeill, 2006), which are of particular interest in research on sign languages (Perniss et al., 2010, 2018). In the vocal domain, iconicity can typically be encoded in the lexicon through a sequence of phonemes that can express sensory or affective meanings (Blasi et al., 2019; Winter et al., 2017). For instance, a common example of vocal iconicity that expresses sensory meaning is the word "woof-woof", as it provides acoustic information and more specifically resembles

the sound of a dog barking.

1.2.1. Iconicity in spoken language

In spoken language, there are various expressions of iconicity. One of these is phonesthemes. Phonesthemes are “language-specific morpheme-like phoneme clusters that lack compositionality” (Johansson et al., 2019b). For example, "gl-" in English expresses sensory meaning for vision and light, such as the words "glitter" or "gloss" (Bergen, 2004). However, there is a debate over whether phonesthemes are truly instances of iconicity in spoken language since they are not reported crosslinguistically. Hence, their inclusion in the different categories of vocal iconicity remains questionable (Cuskley and Kirby, 2013).

Another demonstration of iconicity in the vocal domain is ideophones—namely, “marked words that depict sensory imaging” (Dingemanse, 2012). For instance, the ideophone “pata-pata”—more commonly known as a Japanese mimetic—means “to hit a flat surface with a flat object repeatedly”. This ideophone provides information about the repetition of the activity and about the sensory properties of the object and of the surface (Hamano, 1994). Ideophones have been extensively reported in different languages, such as Sub Saharan, African, and Asian languages (e.g., Dingemanse et al., 2016; Hinton et al., 2006). Undoubtedly, the most frequent case in the category of ideophones is onomatopoeia, words that imitate sounds coming from animals or the environment (Berlin and O’Neill, 1981; Dingemanse, 2018). For instance, the words *bang* or *boom* are onomatopoeic words in English depicting the sound of an explosion.

Although sound symbolism (e.g., Köhler’s example of “maluma-takete”) is considered in the literature as a type of vocal iconicity, there are few problems to this categorization. First, there is no resemblance between the pseudoword ‘maluma’ or ‘takete’ and the perceptual properties of a curved or a sharp shape correspondingly. Iconicity in terms of resemblance (Perniss and Vigliocco, 2014) is evident for onomatopoeic words such as *bang* or *boom* which exactly resemble the noise produced by an explosion. Same

for the terms ‘likeness’ (Peirce, 1960) or ‘shared properties’ (Morris, 1946) described in semiotic theories, they do not hold for sound symbolic mappings. The pseudowords ‘maluma’ or ‘takete’ neither sound like or share any properties with some abstract curved or sharp shapes. Sound symbolism is a distinct phenomenon for which the sound form does not resemble, looks/sounds like, share properties with the referent and should be reconsidered as a case of iconicity in spoken language. In order to understand how meaning is conveyed in the case of sound symbolism it is necessary to investigate the mechanism behind these mappings.

1.2.2. Sound symbolism

Sound symbolism is often used to cover all these phenomena, for which there are mappings between individual meaningless speech sounds (or sequences of speech sounds) and a range of sensory meanings (Hinton et al., 2006; Winter, 2019). In the classic categorization by Hinton et al. (2006) there are four categories of sound symbolism.² More precisely, in synaesthetic sound symbolism, acoustic phenomena represent features in other non-acoustic modalities. The phenomenon of “maluma-takete” would best fit to this category of sound symbolism by Hinton et al. (2006). Another term used to characterize these relations is also known as phonetic symbolism (Brown et al., 1955; Sapir, 1929).

Sapir (1929) was one of the first to provide preliminary evidence for the presence of sound symbolic associations between vowel and size in the vocal domain. Specifically, he showed that the vowel /i/ is associated with small objects and the vowel /a/ with large objects. These associations were evident after he presented to subjects two pseudowords /mal/ and /mil/, which both signified a table. His preliminary report revealed that the word /mal/ was selected by subjects as expressing a big table and the word /mil/

²Hinton et al. (2006) divided sound symbolism into four categories: (1) corporeal sound symbolism, which includes non-segmentable utterances tied to the physical and emotional state of the speaker, (2) imitative, which is identical to onomatopoeia, (3) synaesthetic sound symbolism, and (4) conventional sound symbolism, which represents phonesthemes.

a small one. After Sapir’s work, several other studies provided robust evidence on the sound symbolic associations between meaningless speech sounds and size (Bross, 2018; Knoeferle et al., 2017; Lockwood et al., 2016; Mondloch and Maurer, 2004; Peña et al., 2011).

In addition to meaningless speech sounds-size mappings, many studies have described the presence of non-arbitrary mappings between single phonemes (or sequences) mapped to several other sensory domains. For example, combinations of certain vowels and consonants can represent sensory meaning for tastes, such as the pseudoword ‘kiki’ being mapped better to a saury cranberry sauce than the pseudoword ‘bouba’ (Gallace et al., 2011) (see also Motoki et al., 2020; Simner et al., 2010; Spence and Ngo, 2012). Moreover, iconic meanings expressed by different phonetic features are present for various sensory and motor properties (for a review, see Blasi et al., 2016; Johansson et al., 2019b), such as motion with front vowels being associated with small or sharp movements and back vowels with round and large movements (Koppensteiner et al., 2016; Shinohara et al., 2016), speed with back vowels mapped to low than fast speeds (Cuskley, 2013), color with the vowel /i/ mapped to light colors and /u/ to dark colors (Mok et al., 2019), and texture with voiced consonants associated with rough textures and voiceless consonants to smoothness (Sakamoto and Watanabe, 2018).

1.2.3. The case of “maluma-takete” mappings

Admittedly, one of the most widely investigated sound symbolic examples—and the focus of the present dissertation—is pseudoword-shape associations. In his book *Gestalt Psychology*, Köhler (1929) described the non-arbitrary links between certain pseudowords and shapes. He mentioned that the pseudoword ‘maluma’ (or ‘baluma’) is matched better to a round/cloudy figure and the pseudoword ‘takete’ to a sharp/edgy figure. After Köhler (1929), the sound-shape associations gained widespread attention by the paper of Ramachandran and Hubbard (2001), who modified the initial “maluma-takete” example and established the new “bouba-kiki” example. In this paper the authors found

that 95% of the population reports the pseudoword ‘bouba’ to fit better to a round shape and the pseudoword ‘kiki’ to a sharp one.³ A number of studies followed and experimentally confirmed these observations on pseudoword-shape mappings. One could claim that the majority of studies focused (i) on the specific phonetic, articulatory, or acoustic features of the speech sounds associated with round or sharp shapes (for an overview, see Table 1.1), and (ii) on the crosslinguistic presence of these sound symbolic effects.

The cross-linguistic presence of pseudoword-shape associations is legitimate, with many studies reporting the effect in different languages (English: Maurer et al., 2006), (Japanese: Asano et al., 2015), (French: Fort et al., 2015), (Spanish: Pejovic and Molnar, 2017), (Kitongwe: Davis, 1961), (Himba: Bremner et al., 2013). Two failed replications of the effect in speakers of Hunjara in Papua New Guinea (Rogers and Ross, 1975) and speakers of Syuba in Nepal (Styles and Gawne, 2017) are also noteworthy.

As can be seen, sound symbolism in the form of meaningless speech sounds mapped to various perceptual and motor meanings is a broad linguistic phenomenon. Most studies are concerned with the cross-linguistic presence of these mappings, and with the range of modality features that meaningless human sounds can express. Perhaps, the most studied case of sound symbolism is the “maluma-takete” mapping.

Finally, as the topic of the present dissertation focuses on the pseudoword-shape mapping of “maluma-takete”, the term “sound symbolism” will refer to this specific mapping.

1.3. Why does sound symbolism matter?

In recent years, the growing interest in the cross-linguistic presence of sound symbolism has highlighted its importance in the functional and communicative properties of language. Since more and more evidence has accumulated for the designation of sound

³The 95% of sound symbolism detection needs to be interpreted with caution, as it was not accompanied by any statistical reports in the original paper.

References	Properties (Round vs. Sharp shapes)
Nielsen and Rendall (2013); Maurer et al. (2006); Fort et al. (2015); D’Onofrio (2014)	back rounded vowels vs. front unrounded vowels
McCormick et al. (2015); Fort et al. (2015); Nielsen and Rendall (2013)	voiced consonants vs. voiceless consonants
Knoeferle et al. (2017)	low second vs. high second formant (F2)
O’Boyle and Tarte (1980); Marks (1987)	low frequencies vs. high frequencies
Parise and Spence (2012)	sine wave vs. square wave

Table 1.1.: Various speech sound properties reported in the literature of “maluma-takete” associations. The right column indicates the properties of speech sounds associated to round vs. sharp shapes and the left columns the corresponding references.

symbolism as a linguistic universal, there are questions arising regarding its role in language acquisition and language evolution. Several studies have tried to answer these questions by testing sound symbolic mappings, particularly in human children and infants.

1.3.1. Language acquisition

Already in the 1940s, Irwin and Newland (1940) studied the sound symbolic ability of children and provided evidence that at the age of nine (and not before that), humans can match nonsense words to abstract shapes, similar to the ones presented by Köhler (1929). In the last 15 years, sound symbolism within the sphere of developmental studies and its role in language acquisition has received much attention and can be divided into two main research branches. The first is dealing with questions regarding the nature of pseudoword-shape mappings and whether these mappings are innate or emerge due to exposure to a linguistic environment. The second branch of research encompasses topics on the function of sound symbolism in language acquisition and word learning.

In respect to the first category, it is not generally agreed-upon when sound symbolic ability emerges in humans. However, there is evidence for sound symbolic sensitivity as early as 4 months of age. In a preferential looking paradigm, 4-month infants were tested on the classic “bouba-kiki” effect (Ozturk et al., 2013). Infants were presented in every trial with one shape (round vs. sharp) and one pseudoword (‘round’ vs. ‘sharp’ sounding). Infants looked longer at trials in which there was incongruency between speech sound and shape (i.e., a round shape co-presented with the pseudoword “kiki”). A recent meta-analysis of 11 published and unpublished studies, with subjects’ ages ranging from 4 to 38 months (Fort et al., 2018), suggests that sensitivity to the “maluma-takete” effect is present but moderate in early life and before the age of three. However, there is still not conclusive evidence for a specific age at which the effect emerges. In addition, this sensitivity is present first for the ‘maluma’-type pseudowords and hence for roundness, and later for the sharp category of ‘takete’-type pseudowords

and hence sharpness (Fort et al., 2018). Given the above, Fort et al. (2018) conclude that sensitivity to sound symbolism could be understood as an interplay between a biologically endowed perceptual ability of mapping acoustic properties of speech sounds to abstract visual shapes like those in “bouba-kiki mappings”, and to learned sound symbolic regularities present in the linguistic environment, with ‘round’ sounding words referring to round/curved objects and vice versa for ‘sharp’ sounding words. However there is no study to show such sound-shape associations in adults’ lexicon. As for the earlier sensitivity to round pseudoword-shape mappings, Fort et al. (2018) propose that crossmodal co-occurrences could perhaps explain this effect, since children during the first years of their lives, interact with round, smooth objects that produce soft sounds. This exposure could possibly explain their sensitivity in detecting round sound symbolic associations rather than sharp ones. Nonetheless, the authors neither explained exactly how these co-occurrences are translated to pseudoword-shape mappings nor there is any experimental study examining this scenario.

The discussion on the nature of sound symbolism is in accordance with the general debate on another phenomenon, relatively similar to sound symbolism, known as crossmodal correspondences—namely, “a compatibility effect between attributes or dimensions of a stimulus (i.e., an object or event) in different sensory modalities (be they redundant or not)” (Spence, 2011). For example, a high-pitched tone fits better to a bright stimulus, and a low-pitched tone to a dark one (Hubbard, 1996). The two main positions regarding the origins of these sensory mappings, are that they are either innate (Walker et al., 2010, 2018) or learned through crossmodal statistical regularities present in our environment (Ernst, 2007; Lewkowicz and Minar, 2014; Parise et al., 2014). An example of these natural environmental statistical regularities is found in the associations between frequency and elevation, with high-pitched sounds related to spatial elevation or rise and vice versa for low-pitched sounds. According to Parise et al. (2014) this natural statistical mapping can be explained by the statistics of natural auditory scenes, with correlations between the different noise sources in the environment and sound location in vertical space.

Sound symbolism is often discussed in the literature as a type of crossmodal association shared by humans (Spence, 2011). Nevertheless, since sound symbolism is also a linguistic phenomenon (and not only an association of low-level crossmodal features, like pitch and luminance), the nature of sound symbolism should not be entirely inferred from the nature of crossmodal correspondences.

Regarding the second branch of research, sound symbolism is proposed to bootstrap the acquisition of language and viewed as the first lexicon in the early years of life (Imai and Kita, 2014). A couple of months after birth, humans face the difficult task of mapping words to referents in the environment with most of the form-meaning relations in language being conventional and arbitrary. Sound symbolism could facilitate referentiality —namely associating correctly the word form to its meaning —as it allows for a non-arbitrary mapping between linguistic information and sensory features of the referents (Perniss et al., 2010). This property of sound symbolism would seem valid, if sound symbolism would be a case of iconicity. Assuming a relationship of resemblance between sound form and referent (e.g., ‘maluma’ resembles a round shape) could indeed facilitate the establishment of sound form-referent relationship. However, as discussed above, the characterization of sound symbolism under the resemblance terms is not appropriate. Even if resemblance is not present, it is plausible that the intuitive fit between the sound form and certain sensory qualities of the referent in the case of “maluma-takete” (e.g., front vowels fit better to sharp objects, and back vowels to large or round objects) can reduce referential ambiguity, compared to an arbitrary word form-referent mapping. Last but not least, it is still very important to investigate what allows sound form-referent links in sound symbolism, and what makes a sound express an object’s roundness or sharpness.

Beyond referential insight, sound symbolic associations are proposed to be a useful cue for novel word learning in human infants. Specifically, for pseudoword-shape “maluma-takete” mappings, 14-month Japanese infants were tested with a preferential looking paradigm on the effects of exposure to congruent versus incongruent pseudoword-shape mappings in two different groups. The results of the study showed that, regardless of

the group they were assigned (congruent vs. incongruent), infants could detect sound symbolic mappings and used this sensitivity to establish word referent associations for other instances. From that, Miyazaki et al. (2013) proposed that sound symbolism might facilitate the link between form and meaning and helps childrens' word learning. (Miyazaki et al., 2013).

Moreover, sound symbolic mappings can facilitate verb learning and the generalization of newly learned verbs to new situations when those verbs have some sound symbolic properties (Imai et al., 2008; Kantartzis et al., 2011). Recently, Kantartzis et al. (2019) showed that the semantic representations created by sound symbolic learning can be retained long-term in the memory of 3-year-old children. In this study, children learned novel verbs that were either sound symbolic matched or mismatched to different actions. The next day, the children were asked if the verbs they learned could be matched to a new scene presented to them. Sound symbolism in verbs facilitated the retention of semantic information of the newly learned verbs and could be generalized to novel situations. These verbs can facilitate the differentiation of events or states, as the sound form of the verb can inform directly about the characteristics of these events/states. These examples demonstrate that the sound form of the verbs plausibly evokes a perceptual or motor representations to the individuals that facilitates the link of meaning to referent. However, it remains unclear what these representations are and how exactly they facilitate the word form to meaning link.

It can be concluded that sound symbolic mappings can be detected in the first years of life and most likely provide a useful cue for the link between word form and referent. Ultimately, these sound symbolic advantages could provide information about the role of sound symbolism in the emergence of the precursors of human language, known as protolanguages (Kita, 2008).

1.3.2. Language evolution

The role of sound symbolism in the emergence of protolanguages has already been highlighted by Köhler (1929), who stated the following:

I take it for granted, then, that there are some similarities between the experiences we have through different sensory organs. In passing we may remark that in *primitive languages* one finds much evidence for assuming that the names of things and events often originate according to this similarity between their properties in vision or touch, and certain sounds acoustic wholes. In modern languages, it is true, most of these names have been lost.

As can be seen, Köhler (1929) recognized the importance of “maluma-takete” associations in the roots of primitive languages.

Sound symbolism as well as vocal iconicity have been proposed as possible ways for how human ancestors began to understand the power of speech sounds in expressing meaning. According to Imai and Kita (2014), sound symbolism could have brought referential insight to our ancestors as they realized that linguistic sounds can express meaning about things that surround them. Hence, sound symbolism could have contributed in dissolving the referential ambiguity problem (i.e., how to map linguistic information to objects, events, concepts) and allowed an easier way for mapping linguistic form to meaning (Perniss and Vigliocco, 2014) by the expression of the sensorimotor features of the world’s referents (Winter et al., 2017; Winter, 2019). However, here again sound symbolism is viewed as a case of iconicity and ‘resemblance’ is the key between form-referent relationship. Although, resemblance is not present in sound symbolism as in a onomatopoeia (e.g., “meow-meow” to refer to a cat by resembling the sound of the cat), the non-arbitrariness in the “maluma-takete” example might have also facilitated referential insight. To understand this function of sound symbolism it is essential to understand the mechanism behind this linguistic phenomenon.

As mentioned above, beyond sound-shape mappings, other sensory experiences can also be expressed via sound symbolism (e.g., taste, shape, or movement). In parallel, it is important to highlight that sound symbolic detection or processing is an ability shared by humans and could have facilitated their mutual communication, as they would have shared a common ground of sensorimotor experiences (Cuskley and Kirby, 2013). The importance of sensorimotor information communicated by linguistic sounds in the form of sound symbolism is also in accordance with embodied views on language, suggesting that meanings and concepts are linked in the brain’s sensory and motor systems (Barsalou, 1999; Pulvermüller, 1999). Under such a view, sound symbolic communication could have facilitated the links of a protolexicon in the sensorimotor experiences of a group of people (Cuskley and Kirby, 2013).

Another plausible mechanism of sound symbolism in language emergence is displacement in speech (i.e., to communicate about things that are not present). According to Perniss and Vigliocco (2014), iconicity, including sound symbolism, could have contributed to the conceptual representation of things—namely, the formation of concepts in our mind for object/events that are not directly present in our environment. Iconicity, and hence sound symbolism, can allow for the displacement of our sensorimotor experiences and information, which are not directly present in our environment but about which we need to communicate. Displacement is best demonstrated in onomatopoeic examples, in which linguistic sounds imitate the acoustic output produced—for instance, by an animal or by any other environmental source. On the other hand, for sound symbolism, this function is not so clear as the sound form does not resemble a shape and its roundness or sharpness and hence can not stand for it. This problem is related again to the false consideration of sound symbolism, as a case of iconicity.

Except of its potential role in referential insight and displacement, there is empirical evidence for the role of sound symbolic mappings on language emergence, which comes from studies on communicative games and iterated learning. For example, in a “charades” communication game, in which one had to communicate a meaning to their partner only by vocalizing, subjects produced several non-linguistic iconic vocalizations to express

a set of meanings. According to the authors, these results suggest that the origins of spoken languages can be found in sound symbolic processing and iconicity (Perlman and Lupyan, 2018). Notably, another iterated learning paradigm focused specifically on pseudoword-shape associations of the “maluma-takete” effect (Jones et al., 2014). In that study, subjects divided in ten generations had to learn an “alien language” based on a set of words matched to specific shapes, and then had to reproduce that language in a testing phase. A random sample of the output of a given generation was the learning material for the next generation. Across generations, words related to round shapes started to become more ‘round’ sounding, similarly to ‘maluma’ or ‘bouba’ pseudowords. These findings provide evidence for the emergence of pseudoword-shape associations in an experimental setting, and according to the authors, show how these types of mappings can assist word learning as they can be easier coded, retrieved and remembered, thus shape long-term language change.

To summarize, sound symbolism is considered important in acquisition, learning and evolution of spoken language. Sound symbolism expresses meaning related to sensorimotor experiences, and it is proposed as a facilitatory factor in the emergence of the humans’ sophisticated ability to map linguistic forms to referents. This ability, in turn, could have accelerated and boosted learning, memorization, and retrieval of a protolexicon rooted in our sensorimotor systems. Finally, it is highlighted that although treated as a case of iconicity, sound symbolism is a distinct phenomenon and thus conclusions about its role in language acquisition and evolution under the umbrella of iconicity need to be reconsidered. In order to understand exactly how and why these sound-shape mappings are linked to our sensorimotor experiences and how they could facilitate language learning, it is important to examine the mechanism behind it.

1.4. Possible mechanisms behind sound symbolism

Different theories are found in the literature of sound symbolism, trying to explain the mechanism of such a mapping. One prominent theory is that of Ramachandran

and Hubbard (2001). In this theory the mechanism behind sound-shape associations is found in our articulatory gestures. According to the "syneasthetic articulatory" account of sound symbolism, the authors claim that the movements of the tongue on the palate mimic the round or sharp patterns of abstract visual shapes. So far there is no experimental evidence supporting the link and the resemblance of tongue movements to abstract visual and sharp shapes. Moreover, this account, would not be in accordance with developmental studies, showing sound symbolic mappings to be present at a very early age (Fort et al., 2018). Based on this theory, sound symbolic ability would require precise knowledge of tongue movements of infants and young children, in order to map these movements to abstract shapes. Considering that such a knowledge is difficult even after training in human adults (Ouni, 2011), it seems that articulatory gestures can not explain sound symbolic mappings.

A different proposal by Ohala (1994), known as the frequency code theory, suggests that sound symbolism is found in statistical crossmodal co-occurrences in the environment. For example, a possible explanation for sound-size mappings (Sapir, 1929), and why association of large (small) objects are linked with segments of low (high) frequency, such as vowels having low (high) second formant (i.e., /o/ vs. /i/), is due to the statistical co-occurrence of these features in nature. Large animals vocalize in low frequencies and small animals in high frequencies. These behaviours are present due to differences in the size of the vocal apparatuses; large animals have large vocal apparatuses resulting in the production of lower frequencies and the contrary happens for smaller animals. This account, however, is limited to only sound-size mappings and not to sound-shape mappings as in the "maluma-takete" example. Here it is not clear in which cases sharp shapes co-occur often in the environment with 'sharp' sounding pseudowords or the other way around for round mappings, such that we could learn very early in life these type of mappings.

Last but not least, another proposal suggests that the mechanism of sound symbolism is found in orthography (Cuskley et al., 2017). Visual features of graphemes (roundness vs. sharpness) are showed to predict sound-shape mappings. The explanation of this

effect is that these mappings are mediated by visual mapping strategies between letters and sounds. Visual properties of letters could facilitate the mapping of meaningless speech sounds to abstract shapes. However, orthography seems difficult to provide the explanation behind sound symbolism. The crosslinguistic presence of these mappings, even in illiterate populations (Bremner et al., 2013), as well as empirical evidence for their detection early in life (for a review, see Fort et al., 2018), are not in accordance with the view that orthography can explain these mappings.

The current theories in the field of sound symbolic mappings, and specifically for sound-shape associations of “maluma-takete” type, do not offer conclusive evidence on how meaningless speech sounds can be associated to visual properties of abstract shapes. As the link between the pseudoword ‘maluma’ and a cloudy shape is not a relation of resemblance, the pseudoword ‘maluma’ does not ‘sound’ or look like a curved shape. Although previous theoretical frameworks identified that sound symbolism carries meaning about perceptual properties of an object and thus its meaning should be linked to our sensorimotor experiences (Perniss and Vigliocco, 2014), it remains a mystery how information in the auditory and visual modality come together.

1.5. Actions: the missing element behind sound symbolism

Actions and the knowledge of our actions could provide the missing link between pseudoword and shape mappings. From the first moments of our life, we are collecting sensory and motor experiences while interacting with our environment. Different movements and interactions with objects would result in different sounds and shapes produced by these movements. Learning the auditory and visual outputs of ‘round’ or ‘sharp’ movements could explain the link between ‘round’/‘sharp’ sounding pseudowords to round/sharp shapes. For instance, a sharp movement has auditory and visual by-products, such as a sharp/rough sounding sound and a sharp visual imprint. Vocal imitation of these

sounds could result in ‘round’ or ‘sharp’ sounding pseudowords, similar to ‘maluma’ and ‘takete’, which are later mapped to abstract visual shapes. An important prerequisite for the perceptual and motor learning of these actions would be the neurobiological infrastructure of the human brain. Studies using diffusion tensor imaging have shown that neuroanatomical connections between perceptual and motor cortices in the human brain (Rilling et al., 2008; Rilling, 2014) can support this type of associative learning. The role of actions in sound symbolic mappings would be in accordance with the crosslinguistic presence of “maluma-takete” mappings, as movements and their auditory and visual outputs are universal, sound symbolic mappings cannot be language specific. Finally, as perceptual and motor learning is important for this theoretical proposal, its framework would fit to developmental research on sound symbolic detection that shows evidence for improvement in sound symbolism detection with age (Fort et al., 2013). More experience and exposure to actions and to the auditory and visual by-products of these actions could enhance associative learning between auditory and visual outputs of our actions, across the human lifespan, and therefore strengthen sound symbolic ability.

1.6. Focus of the present dissertation

Despite the extent of empirical research on the topic of sound symbolism it is still unclear what is the mechanism behind these non-arbitrary mappings. The present dissertation aims to examine the mechanism behind the most popular and studied mapping in the literature of sound symbolism—namely, the “maluma-takete” effect (Köhler, 1929)—and the possibility that action knowledge is the missing link between pseudoword-shape mappings. To achieve that, a series of questions will be examined : *Can we replicate sound symbolic effects with a variety of pseudowords in a forced choice task, and what is the relation of sound symbolism with other immanent associations, which include different audiovisual properties? When did sound symbolic ability emerge in the course of evolution, and how this could be related to its mechanism? Is action knowledge important for the mechanism of sound symbolism?*

In order to explore the mechanism behind sound symbolic mappings, it is first important to validate the effect in speakers of different languages and across a set of different ‘round’ and ‘sharp’ sounding pseudowords, beyond the classic ‘maluma’ and ‘takete’ example. Moreover, apart from sound symbolism, other mappings between modality-specific signals are also present in our perceptual environment (e.g., a high-pitched tone mapped to a sharp shape and a low-pitched tone to a round shape). This broader cognitive phenomenon is known as crossmodal correspondences (Spence, 2011). A problem that arises is to understand how humans make decisions when several of these mappings are simultaneously present, and whether sound symbolism can still be detected. Chapter 2 addresses the validation of sound symbolism, and the interactions between two crossmodal mappings and sound symbolism. Parameters such as the pitch of pseudowords and shape’s position introduce two crossmodal mappings next to the classic pseudoword-shape mapping, namely pitch-spatial position and pitch-shape. Although these different mappings have been reported separately in previous studies, it is not known whether their effects are still robust or how they interact together in a forced choice task. Moreover, testing together sound symbolism and crossmodal mappings will improve our understanding on the interaction of these effects and on the properties or mechanisms they might share.

Sound symbolism is considered to be the fossil of protovocal systems in humans (Kita, 2008). Studying the phylogenetic history of sound symbolic ability is very essential to better understand the origins of this ability in humans, as well as its mechanism. Non-human primates, and specifically great apes, are the best model we have at hand to do that. The study of sound symbolism in great apes can allow us to better understand the cognitive and communicative abilities of the last common ancestor shared between humans and great apes, roughly 11 million years ago (White et al., 2009). Moreover, research on the phylogenetic origins of human language can improve our knowledge on the evolution of cognitive abilities in humans and help us understand how cognitive abilities could support symbolic ability (Zlatev et al., 2005). Chapter 3 focuses on the phylogenetic origins of sound symbolic ability, by testing, for the first time, pseudoword-

shape mappings both in humans and in a group of touchscreen trained great apes. Using the same forced choice task, both species were tested on their ability to make intuitive mappings between meaningless speech sounds and visual shapes.

Finally, despite the vast number of studies on pseudoword-shape associations and their effects on the functional and communicative language properties, there is still a lack of consensus on the mechanism of sound symbolic mapping. Chapter 4 investigates, under the hypothesis that sound symbolic mappings are linked to our sensorimotor interactions with the environment (Perniss and Vigliocco, 2014), and under the neurobiological view that meaning is grounded in the perceptual and motor systems in the human brain (Pulvermüller, 2013a), the role of action knowledge as a novel scenario explaining the mechanism of “maluma-takete” mappings. With the same forced choice task, human subjects were tested on the classic sound symbolic mappings and on mappings between ‘round’ and ‘sharp’ sounds of actions and the visual traces produced by these actions. Testing for the first time how the sound symbolic ability of humans relates to their ability to map natural action sounds to the visual products of these actions is a way to investigate whether sound symbolism is linked to our sensorimotor experiences.

To sum up, the overall goal of the present dissertation is to examine a series of fundamental issues, regarding (1) the validation of sound symbolic effects and its interaction with other crossmodal mappings, (2) the phylogenetic origin, and (3) the role of actions behind the mechanism of the most studied sound symbolic mapping, that is speech sounds mapped to abstract shapes.

2. Testing sound symbolic mappings,
pitch-shape and pitch-spatial
position correspondences in a
two-alternative forced choice task

Abstract

Sound symbolism in the form of pseudoword-shape associations, refers to mappings of meaningless speech ‘round’ (‘sharp’) sounding pseudowords to abstract curved (sharp) visual shapes. Crossmodal correspondences are another phenomenon, similar to sound symbolism. Crossmodal correspondences refer to the compatibility effect between features from different modalities, shared by humans. These correspondences can facilitate the grouping of perceptual information present in our environment. In the present study, we test in healthy humans, with a two-alternative forced choice task, sound symbolism, and the interaction of this mapping with modality-specific features present in two crossmodal correspondences. Each forced choice trial included: (1) sound symbolism, namely meaningless speech sounds ‘round’ or ‘sharp’ sounding matched to round or sharp shapes, respectively, (2) pitch-shape mapping, namely high-pitched (low-pitched) sound matched to sharp (round) shapes, (3) pitch-spatial position mapping, that is high-pitched (low-pitched) sound matched to high (low) spatial position. The results replicated the sound symbolic congruency detection effects, while the overall performance of the subjects was determined by this mapping only. Despite previous findings that reported pitch-shape and pitch-spatial position mappings separately, and in different types of tasks, our results propose that during their co-presentation in a forced choice task, only sound symbolic mappings emerged and not the other two correspondences. The perceived ‘roundness’ or ‘sharpness’ in sound symbolic pseudowords plausibly overshadowed the low-level feature of pitch related to the other two correspondences. The present findings point out the need for further investigation on the interaction of different audiovisual mappings processed by humans.

2.1. Introduction

Our environment is filled with multimodal information coming from different or same spatiotemporal directions. Parameters such as time and space help us group together

different properties emerging from the same object/event and facilitate multisensory binding (Calvert et al., 2004). In parallel, other top-down factors, such as semantic (Chen and Spence, 2010) or crossmodal congruencies can make easier the grouping of information into the same sensory event. Crossmodal correspondence is a term introduced by Spence (2011) and refers "to a compatibility effect between attributes or dimensions of a stimulus (i.e., an object or event) in different sensory modalities (be they redundant or not)". Moreover, these correspondences, in contrast to synaesthetic mappings, are shared by the general population. Although there are various correspondences between features of different modalities ranging from vision to audition, smell to shapes, shapes to taste (for a review, see Spence, 2011) the most studied correspondences are the audiovisual ones.

One of these correspondences is pitch and vertical position, or elevation. In this correspondence, there is a mapping between a high-pitched tone matched to a visual stimulus in a high visual position, and a low-pitched tone to a visual stimulus in a low position. This effect has been reported by several previous studies (Ben-Artzi and Marks, 1995; Bonetti and Costa, 2018; Melara and O'Brien, 1987; Mudd, 1963). Evans and Treisman (2009) showed the robustness of this correspondence through a series of speeded classification tasks, during which auditory and visual stimuli were co-presented. Subjects had to categorize properties related to pitch-spatial position correspondence (e.g., direct condition: classification of a tone as high or low-pitched) or properties irrelevant to this correspondence (e.g., indirect condition: is the tone produced by a violin or a piano). Interestingly, the findings revealed crossmodal congruency effects both for the direct and the indirect conditions, and even stronger effects for the direct condition. For instance, when the pitch of the tone was congruent with the presentation of the visual stimuli, the subjects categorized faster the tone compared to when the visual stimulus appeared in an incongruent spatial position. In addition, there is evidence for an early sensitivity in humans to a similar mapping, namely pitch-elevation correspondence (i.e., pitch rising and falling with visual stimulus rising and falling). Preverbal infants, tested with a preferential looking paradigm, looked longer at trials where the pitch

of a tone was congruent with the movement of a visual stimulus (e.g., a tone with increasing frequency matched to a ball moving upwards and vice versa for a low-pitched tone) (Dolscheid et al., 2014; Walker et al., 2010). Finally, sensitivity to pitch-spatial elevation has been reported even in newborns (Walker et al., 2018), suggesting an innate mechanism for this correspondence.

Another mapping of audiovisual features, identified in speakers of different languages is sound symbolism (for a review, see Lockwood and Dingemans, 2015). In sound symbolism, meaningless speech sounds which are ‘round’ or ‘sharp’ sounding match to a round or a sharp shape, respectively. Sound symbolism has been reported by Köhler (1929) with his classic “maluma-takete” example, in which he proposed that ‘round’ sounding ‘maluma’ fits better to a cloudish/round figure and a ‘sharp’ sounding ‘takete’ to a sharp one. Several other studies followed and replicated this correspondence (Asano et al., 2015; Kovic et al., 2010; Nielsen and Rendall, 2011, 2013; Ramachandran and Hubbard, 2001) and even in different age groups (for a review, see Fort et al., 2018), however most of them used a limited set of pseudowords.

Finally, there is another crossmodal correspondence, namely pitch-shape mapping, in which certain features of sound symbolism and pitch-spatial position are present. In pitch-shape correspondences, a high-pitched tone is better matched to a sharp shape, and a low-pitched tone to a round shape. O’Boyle and Tarte (1980) reported these pitch-shape mappings by presenting to subjects either a sharp or round shape, while subjects were asked to turn the dial to the frequency that best matched the presented shape with the usage of a radio oscillator. These results were replicated later in adults (Marks, 1987; Parise and Spence, 2012) as well in preverbal children (Walker et al., 2010).

The interaction between sound symbolic and pitch-shape mappings has not been investigated in the literature, and most of the studies have focused on testing these mappings separately. As far as we are concerned, the closest attempt to study the interaction of these two phenomena is the study of Shang and Styles (2017). However, the focus of

that study was to explore the linguistic use of pitch in sound symbolic mappings. To do that, they combined in the same task classic sound symbolic mappings with different linguistic tones present in Mandarin Chinese. Specifically, they tested native Chinese, English, and bilingual English-Chinese speakers in two different two-alternative forced choice tasks (2AFC) for sound symbolic mappings with the addition of variations of the Mandarin linguistic tones. In the first study, they presented one vowel (/u/ vs. /i/) produced from four different tone categories present in Mandarin Chinese. Each sound was followed by the presentation of two shapes, one sharp and one round, and subjects had to match the presented vowel to one of the two shapes. Their results showed that only the vowel type determined the responses of the subjects regardless of the linguistic tone; hence the vowel /u/ was matched to a round shape and the vowel /i/ to a sharp shape. In their second study, different subjects performed the same task, but this time they were presented with one shape followed by two sounds (i.e., every time the same vowel with two different tones). The findings of the second study revealed a linguistic tone-shape correspondence, while the vowel type was kept constant. Moreover, the effect of tone on shape preference was different across the three language groups. English speakers matched tones to shapes based on pitch height, hence a high-pitched tone was mapped to a sharp shape and a low-pitched tone to a round shape. Chinese speakers, on the other hand, matched tones to shapes based on pitch change (e.g., a tone with multiple changes matched to a sharp shape and a tone with fewer changes to a curve shape). The authors explain that the diversity in the tone-shape effects across speakers can be explained by the different structure of the linguistic sound systems of their languages. Hence, some linguistic sound systems such as the Chinese, focus on pitch change, whereas others on pitch-height.

In the present study, we wanted to validate first, the presence of sound symbolic effects in a group of speakers of different languages, then, the effects of crossmodal correspondences when tested with a forced choice task, and finally the interaction between pitch-shape and of sound symbolism mappings, beyond any language-specific use of pitch. For that reason, we tested speakers of different languages. As pitch-shape

and sound symbolic mappings have been reported in separate studies, we wanted to test which of these mappings will determine the shape selection of the subjects and if congruency on pseudoword and pitch type will improve subjects' performance. In other words, would pitch-shape mapping interfere with sound symbolic mapping? Moreover, since pitch is related to the property of spatial position, we introduced one more parameter, that of the spatial position of the visual stimulus. Secondly, with the addition of the pitch dimension, we wanted to test if pitch-spatial position correspondence would overshadow sound symbolic mappings and/or pitch-shape mappings. Would subjects match a high-pitched pseudoword to a shape on a high position and a low-pitched pseudoword to a shape in a low position regardless of the type of the shape (round vs. sharp)?

First, we expected that subjects will detect sound symbolic congruency effects. Moreover, we assumed that performance in at least one of the three mappings would be at chance level, given that some mappings may affect more the mapping strategy of the subjects. Secondly, we expected that in a given trial, congruency between pitch and pseudoword category on shape type would improve the sound symbolic congruency performance of the subjects. Specifically, subjects will perform better and faster on sound symbolic mappings, when both pseudoword type ('round' vs. 'sharp' sounding) and pitch (low vs. high-pitched) match to the same shape, compared to trials with incompatible features with the pseudoword (e.g., 'round' sounding pseudoword but high-pitched). To explore these hypotheses, we conducted a classic sound symbolic 2AFC task by adding the variable of pitch to the pseudowords (low or high-pitched) and the variable of spatial position (the two shapes were presented vertically).

2.2. Materials and Methods

Subjects

Twenty-four healthy right-handed adults (14 females, age $M=25.87$, $SD=5.08$) participated in the study. The subjects were native speakers of different languages (11 German,

3 Greek, 2 Italian, 2 Spanish, 1 French, 1 Bulgarian, 1 Russian, 1 Urdu, 1 Kurdish, 1 Afrikaans). Two of the subjects were bilinguals, one speaking Greek and Albanian, one Afrikaans and English. All subjects had normal hearing and normal or corrected-to-normal vision. Subjects were recruited from written announcements at the Freie Universität Berlin. All methods were approved by the Ethics Committee of the Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin, and were performed in accordance with their guidelines and regulations. All subjects provided written informed consent prior to the participation to the study and received 10 euros for their participation. ¹

Stimuli

The auditory stimuli were recorded and edited on Audacity (2.0.3) (Free Software Foundation, Boston, USA) by a female native Greek speaker. The visual stimuli were the same shapes as in the study of Margiotoudi et al. (2019) (see Table B.1), but this time presented in the middle upper and middle lower part of the screen. For the auditory stimuli, we used a set of 24 trisyllabic pseudowords in the form of (CVCVCV) (see Table A.1). There were twelve pseudowords in each category of ‘round’ and ‘sharp’ sounding stimuli. The combination of vowels and consonants was determined from the ratings described in the study of Margiotoudi et al. (2019). We recorded the auditory stimuli in a soundproof booth (sampling rate: 44.1.kHz). The average duration of all the 24 pseudowords was $M=745 \pm 29.84$ ms. We modified also the pitch of the auditory stimuli and created two subcategories of low and high-pitched stimuli. The low-pitched pseudowords had an average fundamental frequency (F0) of 214 Hz, the high-pitched ones 247 Hz and the initial baseline frequency averaged at of 232 Hz (see Fig. 2.1). After checking the normal distribution of the three pitch categories with Shapiro-Wilk’s test (for all three categories : $p < 0.05$), we ran a Kruskal-Wallis test ($\chi^2(2)=52.07$, $p < 0.001$) to check if the F0s were significantly different among the three categories. In

¹The subjects were the same as the subjects of Experiment 1 in Margiotoudi et al. (2019).

addition, the multiple comparisons tests showed significant differences between all categories (all $ps < 0.001$). In total, the list of auditory stimuli consisted of 48 pseudowords, 12 high-pitched ‘round’ sounding pseudowords, 12 low-pitched ‘round’ sounding pseudowords and the same number for the ‘sharp’ sounding category. After modifying the F0s, we normalized all auditory stimuli for sound energy by matching their root mean square (RMS) power to 24.0 dB.

Design and Procedures

The experiment was designed in E-Prime 2.0.8.90 (Psychology Software Tools, Inc., Pittsburg, PA, USA). Subjects performed a 2AFC task. In each trial, all four variables were crossed, namely pseudoword, pitch, shape, and spatial position (see Table 2.1). Each trial started with the presentation of a fixation cross for 500 ms, followed by the presentation of an auditory stimulus for 800 ms (i.e., a ‘round’ or ‘sharp’ pseudoword with high or low pitch). Next, two target shapes (always one round and one sharp) appeared vertically on the screen, one in the middle upper part of the screen and one in the middle lower part. The shapes remained on the screen for 1500 ms; during this time, responses were collected. Every trial ended with the presentation of a blank slide lasting 500 ms. All slides were presented on a grey background (RGB 192,192,192) (see Table 2.2).

The experiment consisted of 288 trials and was divided into three blocks (96 trials each). The blocks were separated by two pauses in between them. In each block, all the auditory stimuli were presented twice and all the visual shapes eight times. There was a total of 48 combinations unique in each block, and all trials were randomized within each block.

The procedure was identical to the Experiment 1 of Margiotoudi et al. (2019). Before the initiation of the experiment, subjects received the following written instructions: "During the experiment, two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures that matches to the

sound you hear". No specific instructions were given to the subjects regarding speed or accuracy.

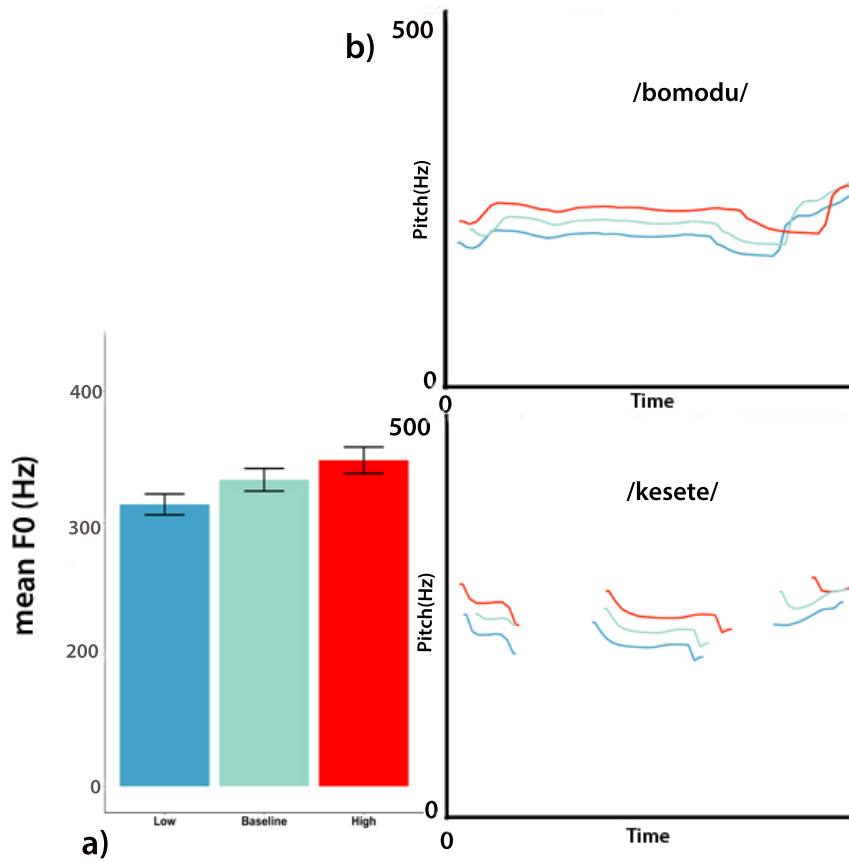


Figure 2.1.: a) Bar plots show average and standard deviations of fundamental frequencies (F0s) for high (red), low (blue) and baseline (green) categories for all pseudowords. b) Pitch contours for the three different pitch categories for the ‘round’ sounding pseudoword "bomodu" and for the ‘sharp’ sounding pseudoword "kesete".

2.3. Data analysis

For all analyses, trials with reaction times greater than 1500 ms or no response were excluded. One subject was excluded from the analysis, due to poor understanding





Pseudoword	Shape	Position
round		—
sharp		—
high		up
low		low

Table 2.1.: Congruency pairs between the pseudoword features (‘round’ vs. ‘sharp’ & low vs. high-pitched), the shapes (round vs. sharp), and the position of the shapes (high vs. low) during the task.

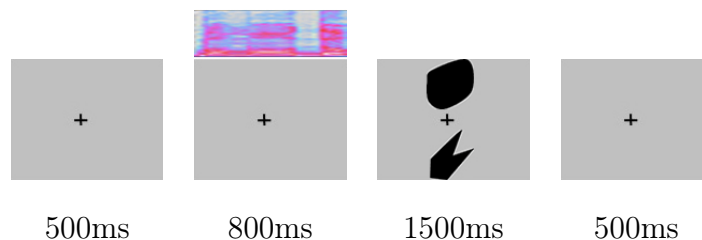


Table 2.2.: Schematic representation of the experimental design of the 2AFC crossmodal task.

of the English language. First, we checked the normal distribution of the data with Shapiro-Wilk’s test. After that, using three non-parametric one-sample Wilcoxon signed-rank tests against chance, we checked (1) if subjects’ performance under sound symbolic matching—selecting a round shape when a ‘round’ sounding pseudoword preceded and a sharp shape for a ‘sharp’ sounding pseudowords—exceeded chance levels, (2) if subjects’ performance under pitch-shape matching—selecting a round shape when a low-pitched pseudoword preceded and a sharp shape for a high-pitched pseudoword—exceeded chance levels and (3) if subjects’ performance under the pitch-spatial position matching—selecting the shape at the upper part of the screen when a high-

pitched pseudoword preceded and the shape at the low part of the screen for a low-pitched pseudoword—exceeded also chance levels.

In order to explore if shape selection was influenced by the pitch or pseudoword type, we fitted a generalized linear mixed model (GLMM) with a binomial error structure. As analysis tool, R version 3.4.3 was used including the package lme4 (Bates et al., 2014). The dependent variable was the selected shape (round vs. sharp); as fixed effects, we included word type ('round' vs. 'sharp' sounding pseudoword) and pitch of the pseudowords (low vs. high-pitched). As random effect, we included intercepts for subject and random slopes for each trial nested within these random effects. The likelihood ratio test (LRT) was used to check if the predictor variables improved the fit of the model; these were calculated by comparing the full model to a reduced model that included all terms except the fixed effect terms in question.

In addition, we conducted one analysis of variance (ANOVA) to check if the different combinations of pitch (high vs. low) and pseudoword ('sharp' vs. 'round' sounding) affected the reaction times of the subjects. For instance, would the subjects respond faster in trials where there was congruency between pitch and pseudoword type (e.g., 'round' sounding and low-pitched pseudoword). We performed ANOVA because normality was not violated for none of the categories for reaction time (High/Sharp : $W=0.93$, $p = 0.11$, Low/Sharp: $W=0.93$, $p = 0.14$, Low/Round: $W=0.96$, $p = 0.62$, High/Round: $W=0.95$, $p = 0.29$).

In order to explore sound symbolic congruency performance, we fitted one more GLMM model with a binomial error structure to check which variables affected the accuracy of the subjects on matching a 'round' sounding pseudoword to a round shape and a 'sharp' sounding pseudoword to a sharp shape. Fixed effects were the pseudoword type ('round' vs. 'sharp'), the type of pitch (low vs. high) and the number of trial. As random effects, we included intercepts for subject and random slopes for each trial. We used the likelihood ratio test (LRT) to check if the predictor variables improved the fit of the model. Chi-squares and p-values were computed using the function drop1 from

the R package lme4 for all the GLMM models.

Finally, in order to check any shape preference effects regardless of the type of the preceding sound, we calculated the proportion of times each subject chose a round or a sharp shape (independent of the previous acoustic stimulus) and performed a Wilcoxon signed-rank test. The same analysis was performed in order to check spatial position preference irrespective of the features of the preceding sound. In other words, we checked whether subjects selected more often the stimulus that appeared at the upper or at the lower part of the screen.

2.4. Results

A total of 3.34% trials were excluded from the analysis because no response was given or responses exceeded 1500 ms. Based on the results of the Shapiro-Wilk tests, normality was violated for the pitch-shape condition: $W=0.62$, $p < 0.001$, but not for the other two conditions, sound symbolism: $W=0.92$, $p = 0.10$, pitch-spatial position: $W=0.97$, $p = 0.81$. The one-sample Wilcoxon signed-rank tests, revealed significantly above chance performance only for the sound symbolic mapping ($V=225$, $p < 0.001$) 71.06%, but not for the pitch-shape correspondence ($V=170$, $p = 0.16$) 52.65% or for the pitch-spatial position with 50,70% ($V=163$, $p = 0.10$) (see Fig. 2.2, for further analysis on the pitch-spatial position and pitch-shape correspondences, see Appendix A).

The model which explored the shape selection, revealed an effect of word type ($\chi^2(1)=1389$, $p < 0.001$), and pitch type ($\chi^2(1)=24.45$, $p < 0.001$). Subjects selected more often round shapes after the presentation of a ‘round’-sounding pseudoword or/and a low-pitched pseudoword (see Fig. 2.3). We compared further these differences within each category with paired sample Wilcoxon tests. For the word type, there was a significant difference for shape preference ($V=10$, $p < 0.001$), with more selection of round shapes when the pseudoword was just ‘round’ sounding (round: 83.16% while for sharp: 42.82%). On the other hand, there was no significant difference between the two pitch categories on

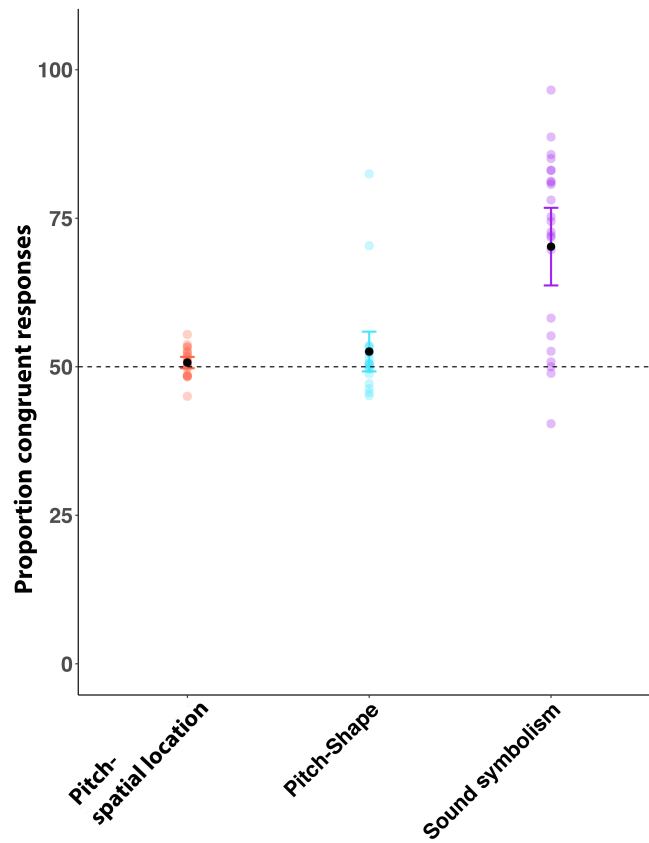


Figure 2.2.: Proportion of congruent responses for the three conditions. Light colored circles indicate congruent responses for each individual for the three conditions: pitch-spatial location (orange), pitch-shape (blue) and sound symbolism (purple). Black circles indicate average performance for each category. Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance.

the round shape preference ($V=91$, $p = 0.15$) (high-pitched : 60.77 % & for low-pitched: 65.87%).

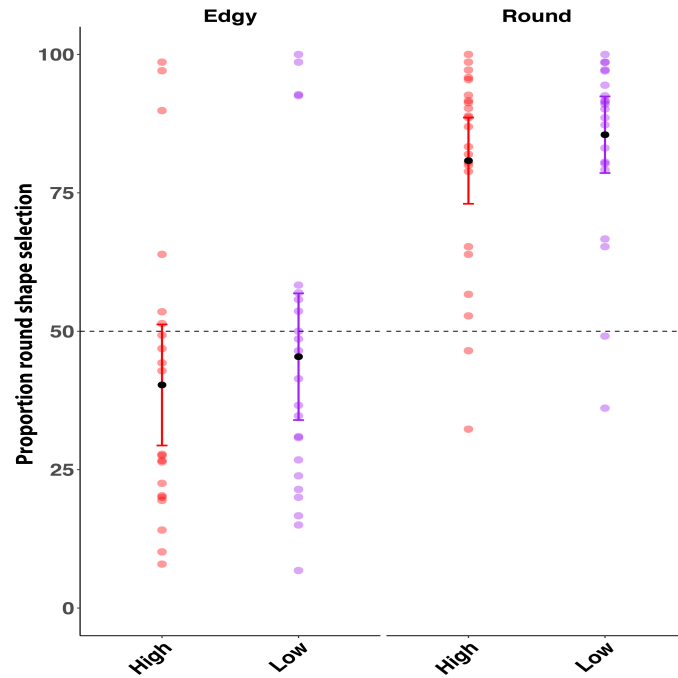


Figure 2.3.: Proportion of round shape selection for the two pitch (low vs. high) and ('round' vs. 'sharp' sounding) pseudoword categories. Light colored circles indicate percentage of round shape selection for each individual across all the four categories: high-pitched pseudowords (red), low-pitched pseudoword (purple). Black circles indicate average performance for each category. Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance.

The comparison of reaction times for the different combinations of pseudoword features revealed no significant differences across the four possible different combinations of pseudoword features ($F(3,92)=0.29$, $p = 0.82$). Hence, pitch and pseudoword type congruency on shape selection did not shorten the response times of the subjects.

As for the analysis on sound symbolic congruency, the full model for exploring the performance in the sound symbolic condition significantly differed from the reduced

model ($\chi^2(2)=641.26$, $p < 0.001$). There was an effect only of word type and not of pitch type. Specifically, a higher congruency detection was found for ‘round’ sounding pseudowords than for ‘sharp’ ones. On average, there was a percentage of 83.16% congruent responses for ‘round’ vs. 57.17% for ‘sharp’ pseudowords. Further analyses using Wilcoxon signed-rank tests ² showed that performance was above chance only for the ‘round’ sounding pseudowords ($V=298$, $p < 0.001$) and not for the ‘sharp’ ones ($V=205$, $p = 0.06$)(see Fig. 2.4a & b). Moreover, there was a significant difference between incongruent ‘round’ sounding pseudowords: 16.83% vs. incongruent ‘sharp’ sounding: 42.85% ($V=0$, $p < 0.001$).

Furthermore, we conducted an additional Kruskal-Wallis test to compare congruency performance in sound symbolism for the different combinations of pitch (high vs. low) and pseudoword types (‘round’ vs. ‘sharp’ sounding).³ The analysis revealed a difference in congruency levels for the four different combinations of pitch and roundness of pseudowords ($\chi^2(3)=31.68$, $p < 0.001$), hence the pseudoword characteristics had an effect on the subjects’s performance (see Table A.2). However, the pairwise comparisons revealed systematic significant differences between all categories that differed in the ‘roundness’ or ‘sharpness’ of the pseudowords but not in their pitch. Consequently, we exclude the possibility that congruency in pitch and pseudoword type had any effect or facilitated sound symbolic congruency detection.

In respect to any preference in the visual stimulus’ position or shape type, regardless of the features of the preceding sound, there were no significant preference for selecting the stimuli at the upper or at the lower part of the screen ($V=204$, $p = 0.12$) (low position: 50.81%; up position: 49.18%). In contrast, for the shape preference, subjects chose significantly more often round shapes versus sharp shapes (63.31% round vs 36.68%

²We used Wilcoxon signed-rank tests because normality was violated for both pseudoword categories (‘round’: $W=0.87$, $p < 0.05$ & for the ‘sharp’: $W=0.89$, $p < 0.05$) after running Shapiro-Wilk tests.

³Kruskal-Wallis test was performed, because normality was violated for these categories (Low/Sharp: $W=0.89$, $p = 0.2$,Low/Round: $W=0.79$, $p < 0.001$, High/Round: $W=0.89$, $p < 0.01$, High/Sharp: $W=0.87$, $p < 0.01$).

sharp) ($V=300$, $p < 0.001$). However, there were three subjects that showed an extreme bias on round shape preference, with about 80% of the times selecting round shapes (see Fig. 2.5). These same subjects had a similar performance for the sound symbolic task performed in Experiment 1 of Margiotoudi et al. (2019). Moreover, their performance across all three conditions (pitch-shape, pitch-spatial position, sound symbolism) reached chance levels.

In order to exclude any possible effect of the extreme performance of these three subjects on the total congruency obtained for the sound symbolic task for the two pseudoword categories ('round' vs. 'sharp' sounding), we ran a Wilcoxon signed-rank test for testing congruency for each pseudoword category against chance, and excluded these three subjects. Performance significantly improved for the 'sharp' sounding pseudowords reaching 64.79% ($V=204$, $p < 0.001$), as for the 'round' sounding ones, congruency levels remained significantly above chance, reaching 81.42 % ($V=229$, $p < 0.001$). In contrast, the comparison of incongruent responses between the two pseudoword categories remained significant (incongruent 'round' sounding pseudowords: 18.57% vs. incongruent 'sharp' sounding: 35.20% ($V=0$, $p < 0.001$)). Hence, although subjects performed above chance for the two pseudoword categories, they had significantly more incongruent responses for 'sharp' sounding pseudowords (see Fig. 2.4).

In contrast, the analysis on the shape preference excluding these three subjects showed that, on average, subjects selected 58.67% of the time round shapes irrespective of the preceding sound, and 41.32% of the time sharp ones, and these scores were significantly different from each other ($V=231$, $p < 0.001$).

2.5. Discussion

In the present study, we tested in the same 2AFC task, sound symbolism, pitch-shape, and pitch-spatial mappings by crossing in the same trial the different auditory and visual properties present in all three mappings. In every trial, a 'round' or 'sharp' sounding

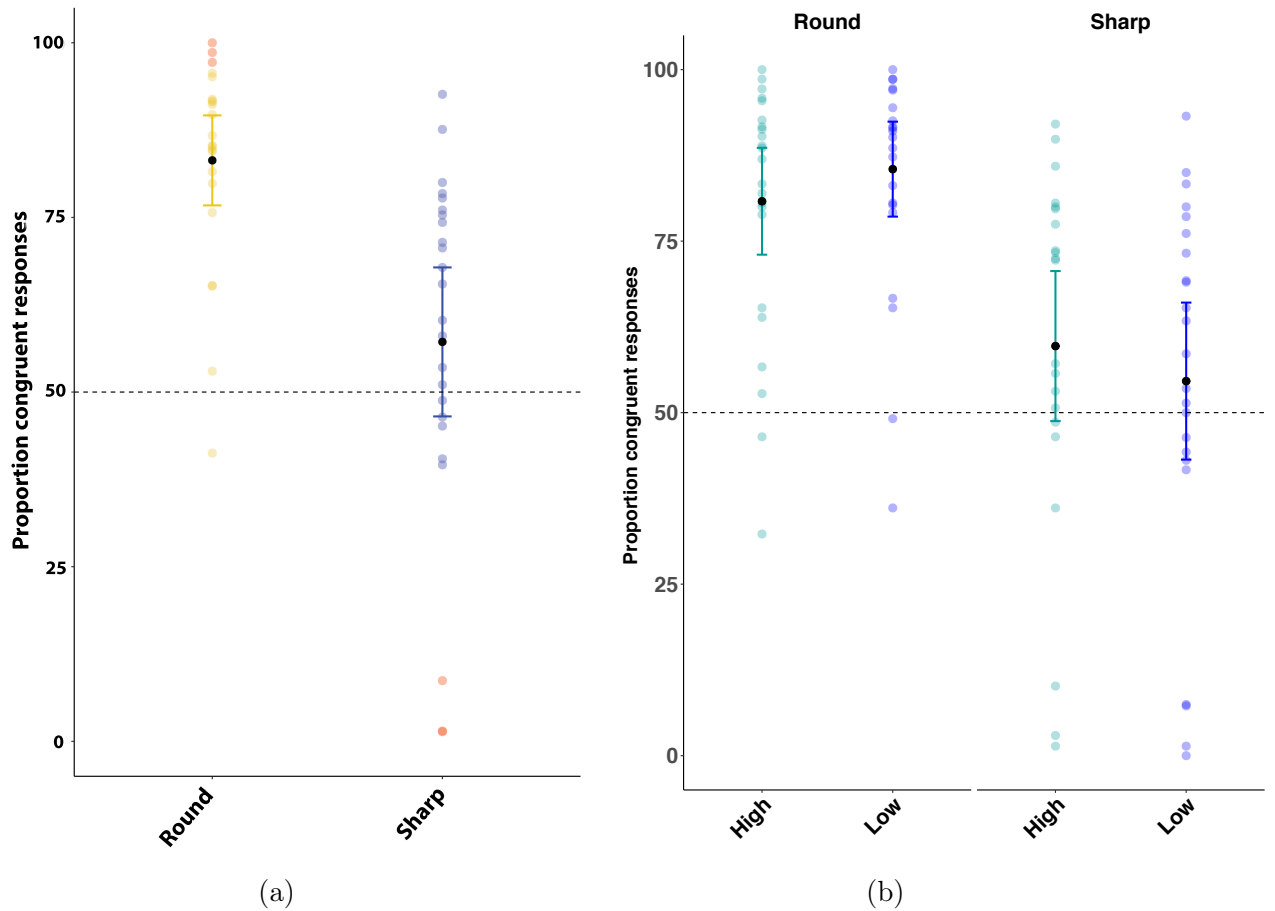


Figure 2.4.: a) Proportion of congruent responses for the two pseudoword categories in the sound symbolic condition. Light colored circles indicate congruent responses for each individual for the two categories: ‘round’ sounding (yellow) and ‘sharp’ sounding pseudowords (blue). Red circles indicate the subjects that selected more than 80% of the times the round shapes. Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance. b) Proportion of congruent responses for the sound symbolic condition for the different combinations of pseudoword features (low vs. high-pitched and ‘round’ vs. ‘sharp’-sounding). Light colored circles indicate congruent responses for each individual for the two categories: high-pitched (green) and low-pitched pseudowords (blue). Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance.

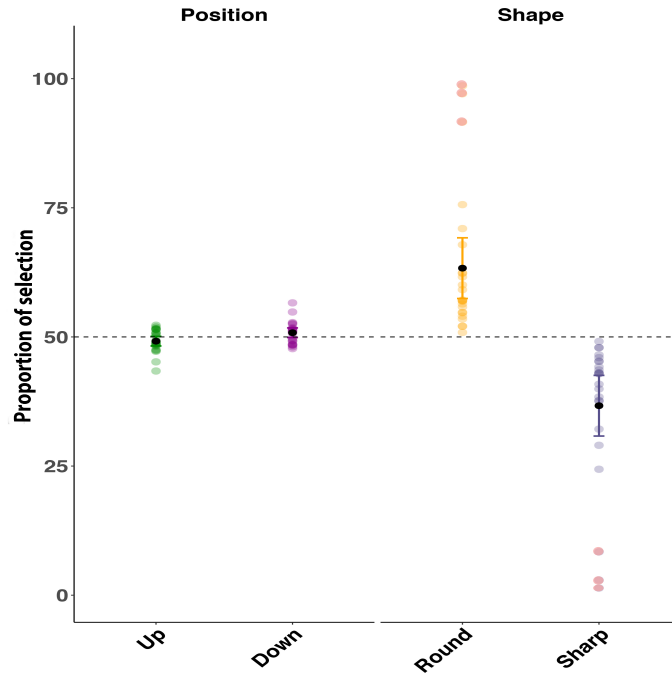


Figure 2.5.: Proportion of selecting the upper or the lower position of the screen (upper position: green; lower position: pink) and proportion of selecting one of the two shapes (round shapes: yellow; sharp shapes: purple) regardless of the preceding sound. Light colored circles indicate percentage of selection for each individual. Red circles indicate the subjects that selected more than 80% of the times round shapes. Black circles indicate average percentage of selection for each category. Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance.

pseudoword with high or low pitch was followed by the presentation of two shapes, one round and one sharp, presented vertically on the screen. The main finding of the study revealed above chance congruency detection only for the sound symbolic condition, but not for the pitch-spatial position, nor for the pitch-shape mappings. Therefore, in every trial, subjects mapped pseudowords to shapes based on the perceived ‘roundness’ or ‘sharpness’ of the pseudowords, irrespective of their pitch. Consequently, the spatial positions of the shapes (high vs. low) did not determine the shape selection of the subjects. In addition, congruency of pitch and pseudoword type for shape mapping (e.g.,

‘round’ sounding and low-pitched pseudoword are both congruent to a round shape or ‘sharp’ sounding and high-pitched pseudoword are both congruent to a sharp shape) did not facilitate the sound symbolic congruency detection of the subjects, when compared to trials in which pseudoword and pitch type were incongruent in respect to the shape. The present findings validate the presence of sound symbolic effects when tested with a 2AFC task, but not the emergence of pitch-spatial position and pitch-shape mappings. Subjects ignored the information of pitch as a criterion for shape or shape’s position selection.

Our results are in accordance with the first study of Shang and Styles (2017). The authors replicated the classic sound symbolic results of vowel-shape mappings in a 2AFC survey study, during which a vowel with a distinct tone was followed by the presentation of two shapes. In this study there was an effect of vowel type in the shape selection of the subjects but no effect of the Mandarin tones. However, in their second study—which was again a 2AFC survey but this time one shape presentation was followed by the presentation of the same vowel paired with two different tones—they showed an effect of tone on shape preference. Specifically, they found an effect of language specificity on tone-shape mappings. In more details, the English-speaking subjects mapped high tones to spiky shapes and low tones to round shapes, whereas the Chinese-speaking subjects mapped steady tones to curvy shapes and tones with dynamic changes to pointy shapes. The authors proposed that these differences in tone-shape mappings are due to their different structured linguistic sound systems, and that these differences can affect sound symbolism. For instance, the language of Chinese speakers requires attention to the dynamic changes of pitch in speech rather than to its low or high frequency. It is important to note that this language specificity emerged only in their second study, in which the vowels were kept constant and only the tones were changed. The findings of our study are in accordance with the above results and provide evidence that in a 2AFC task on matching speech sounds to shapes, sound symbolic information (either from vowels or pseudowords) dominates over pitch information.

We propose here that the absence of congruency or facilitatory effect of pitch on

sound symbolic performance was overshadowed by the complex phonemic properties of ‘round’ or ‘sharp’ sounding pseudowords. As Shang and Styles (2017) suggested, the vowel identity could have overshadowed some subtle tone differences. Similarly, in our present study, the sound symbolic features of the pseudowords were enough to determine shape selection and outweigh the pitch characteristics of pseudowords. In addition, we used even more complex sound symbolic stimuli, which consisted of both vowels and consonants (D’Onofrio, 2014; McCormick et al., 2015; Nielsen and Rendall, 2011). Hence, the combinations of ‘round’ and ‘sharp’ sounding vowels and consonants were sufficient to determine the preference of the subjects for sound symbolic mappings of pseudoword-shape type, and masked more low-level features, such as pitch of pseudowords.

Regarding the absence of any pitch-spatial position effects, the instructions administered to the subjects could possibly explain the present finding. The instructions guided the subjects to match sounds to shapes, hence they were explicitly instructed to pay attention to the shapes and not to the position of the shapes. It is very possible that these instructions introduced a bias regarding the attention and responses of the subjects to spatial position. Future research should test if instructions could bias the responses of the subjects, by testing the same mapping with a 2AFC task but under explicit instructions regarding the importance of the shapes’ spatial position.

The present results on sound symbolic congruency are in accordance with the findings of Experiment 1 in Margiotoudi et al. (2019), and more importantly with a different set of pseudowords. The same subjects participated in both studies in the same session. In contrast with Experiment 1, in which bisyllabic pseudowords were presented, in the present sound symbolic task, we used trisyllabic pseudowords with additional variables: pitch and vertical positions of the presented shapes. Subjects in both studies reached an average of about 70% congruency detection performance.

In addition, as in Experiment 1 of Margiotoudi et al. (2019), subjects showed a round shape preference regardless of the preceding sounds and selected 63.31% of the time round shapes. An effect which could be related to a natural aesthetic preference of

humans on curved over sharp contours (Bar and Neta, 2006; Bertamini et al., 2016; Palumbo et al., 2015). This effect resulted in a significantly higher congruency detection for ‘round’ sounding pseudowords compared to sharp ones. However, once we removed from the analysis three subjects with extremely high round shape preferences, sound symbolic congruency detection was significantly above chance for both ‘round’ and ‘sharp’ sounding pseudowords. Note that the comparison of incongruent responses remained significant with more incongruent responses for ‘sharp’ sounding pseudowords. Given that in these two studies we used different pseudowords (bisyllabic vs. trisyllabic), and that the word selection was based on a combination of ‘round’ vs. ‘sharp’ sounding phonemes which were rated before the studies (see Methods, Margiotoudi et al., 2019), we exclude the possibility that our ‘round’ and ‘sharp’ sounding pseudowords in both experiments were not well selected. In contrast, we suggest that this pattern on higher congruency for round sounding pseudowords, which has been previously reported (Fort et al., 2018; Jones et al., 2014), is present due to the overall preference of people for round over sharp shapes.

To conclude, in the present study, we validated the presence of sound symbolic mappings, when tested for a new set of pseudowords and in speakers of different languages, and showed that pitch-shape and pitch-spatial position mappings did not emerge, when tested and co-presented in the same two-alternative forced choice task. Although previous studies have reported crossmodal correspondences of pitch-shape and pitch position with different types of tasks (Evans and Treisman, 2009; O’Boyle and Tarte, 1980), these correspondences were not detected in the present 2AFC task. Moreover, sound symbolic properties of pseudowords and not pitch guided the decision of the subjects for detecting the sound-shape mappings. This finding suggests that ‘sharp’ or ‘round’ sounding pseudowords, despite possible differences in pitch, also have other features that differentiate them from each other and make them being perceived as ‘sharp’ or ‘round’. For example, the abrupt changes found in the pseudoword "ta-ke-te" are not the same with the smooth transitions present in a ‘round’ sounding pseudoword like "bou-ba". These differences could be one of these acoustic properties that make a pseudoword sound more

sharp or round. Thus, it is possible that pitch-shape and sound symbolic mappings share a similar mechanism as they are mapped to the same shapes. However, it is still necessary to explore what other additional acoustic properties make a pseudoword ‘sharp’ or ‘round’.

As for the pitch-spatial position mapping, a possible explanation for the effect is related to the type of instructions administered (not explicit instruction on pitch-position detection) but also in the type of task. Previous research found a significant effect for this correspondence on reaction times when tested with a speeded classification task (Evans and Treisman, 2009), while attention was required to the sound, and the spatial position of the visual stimulus was used as priming. In that study, subjects classified faster a sound, when the primed spatial position was congruent to this sound. The present results suggest that in a 2AFC in which subjects are not explicitly instructed to pay attention to pitch or spatial position, pitch-spatial position mappings are not detected. In addition, these results fit to the general framework of work in crossmodal attention, which suggests that attention to one modality can produce shifts of attention to other modalities and facilitate crossmodal links (for a review, see Driver and Spence, 1998). Here as attention was not required neither to pitch nor to spatial position, a crossmodal attention shift and hence crossmodal link was not possible.

There is certainly a need for further investigation on the combinations of different mappings. In our daily environment, we are presented with multimodal features that meet together and we have to make a perceptual choice beyond spatiotemporal parameters on which features ‘go together’ in order to group them in the same event, an ability known as “unity assumption” (Spence, 2007). Although there is extensive research on different crossmodal correspondences (for a review, see Spence, 2011), and on sound symbolism (for a review, see Lockwood and Dingemans, 2015), there is yet little known on the natural co-existence of these mappings, and how they affect the perceiver’s decisions. Future studies, using different types of tasks should focus on the interactions of these different mappings and on how they affect decision making processes in humans. These interactions could improve our knowledge of the perceptual hierarchy between

these mappings, of their mechanisms and shared properties. Most importantly, these interactions could help us better understand which mappings are more meaningful to humans when they need to make sense out of them.

3. Sound symbolic congruency detection in humans but not in great apes

This chapter is based on :

Margiotoudi, K., Allritz, M., Bohn, M., & Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific reports*, 9(1), 1-12.

doi: <https://doi.org/10.1038/s41598-019-49101-4>

The original article has been published under a (CC-BY) license. This version of the article may not exactly replicate the final version published. It is not the version of record. **Authors contributions:** study concept and design (KM and FP), material generation and data collection (KM), data analysis (KM), manuscript drafting (KM), revisions (KM, FP, MB and MA).

Abstract

Theories on the evolution of language highlight iconicity as one of the unique features of human language. One important manifestation of iconicity is sound symbolism, the intrinsic relationship between meaningless speech sounds and visual shapes, as exemplified by the famous correspondences between the pseudowords ‘maluma’ vs. ‘takete’ and abstract round and sharp shapes. Although sound symbolism has been studied extensively in humans including young children and infants, it has never been investigated in non-human primates lacking language. In the present study, we administered the classic “takete-maluma” paradigm in both humans (N=24 and N=31) and great apes (N=8). In a forced choice matching task, humans but not great apes, showed crossmodal sound symbolic congruency effects, whereby effects were more pronounced for shape selections following round-sounding primes than following edgy-sounding primes. These results suggest that the ability to detect sound symbolic correspondences is the outcome of a phylogenetic process, whose underlying emerging mechanism may be relevant to symbolic ability more generally.

3.1. Introduction

There has been a long debate in semantics as to whether the relationship between form and meaning of a sign is entirely arbitrary or not (Saussure, 1959; Hockett, 1960). A classic example of non-arbitrariness in human language is sound symbolism. Sound symbolism describes the phenomenon that humans match pronounceable but meaningless pseudowords to specific visual shapes. Köhler (1929), who discovered sound symbolism, had reporting that the pseudoword ‘maluma’ was judged to be a good match to a round shape whereas the pseudoword ‘takete’ was judged as better match to a sharp shape. Instead of "sound symbolism", other terms have been used, for example "phonetic symbolism" (Sapir, 1929) or "crossmodal iconicity" (Ahlner and Zlatev, 2010).

A number of studies have documented sound-meaning mappings in speakers of a broad

range of languages (Blasi et al., 2016), including South East Asian languages (Watson, 2001), African languages (Childs, 1994), Balto-Finnic (Mikone, 2001) and Indo-European (McCormick et al., 2015), thus ruling out language specificity as a possible factor. Although cross-modal sound symbolic relationships replicate across a wide range of experiments in human adults or children with variable language backgrounds using different stimuli, it still appears as a mystery why a majority of human subjects agree that certain speech items sound ‘rounder’ or ‘sharper’, and why certain visual and acoustic stimuli intuitively match with each other.

Sound symbolism has been claimed to facilitate language acquisition and development. For example, Maurer et al. (2006) have shown that 2.5 years old children matched oral sounds to shapes. Other studies tested whether sound symbolism facilitates verb learning and found positive evidence in 25-month-old Japanese (Imai et al., 2008) and 3-year-old English children (Kantartzis et al., 2011). In both studies, children learned novel verbs that sound-symbolically matched or did not match different actions. Based on their findings, children performed better on generalizing the novel verbs to the same actions but in different contexts (e.g., different actor performing the action), when the novel learned verbs sound-symbolically matched the described action during the learning phase. In contrast to these results from children already knowing some language, evidence for sound symbolic matching in preverbal infants is less conclusive. A sequential looking time study found that 4-month-old infants looked longer at incongruent correspondences between shape and sound than to congruent ones (Ozturk et al., 2013). However, Fort et al. (2013) found no such evidence in 5 and 6-months old infants tested in a preferential looking paradigm. A recent meta analysis (Fort et al., 2018) concluded that it is still unclear whether preverbal infants are capable of sound symbolic matching. Hence the sensitivity to sound symbolism in early life is an open issue.

Research in nonhuman animals has investigated the understanding of sound-image correspondence for familiar categories (e.g., vocalizations vs. faces of con-specifics, or human speech and human faces; Adachi et al. (2006); Adachi and Fujita (2007); Hashiya and Kojima (2001); Kojima et al. (2003); Martinez and Matsuzawa (2009); Proops et al.

(2009); for a review, see Izumi, 2006). Animal research and specifically research in non-human primates has not directly addressed the question of abstract sound-shape correspondences. However, Ludwig et al. (2011) tested crossmodal correspondences between luminance and pitch in great apes. In this study, 6 chimpanzees were trained to perform a speeded classification paradigm of squares with ‘high’ or ‘low’ luminance. During the testing phase, chimpanzees (as well as a human control group) performed the same task again, however now with ‘high’ - or ‘low’ -pitched sound co-presented. Chimpanzees, like humans, performed better when a congruent sound was presented (high-pitched sound with high-luminance square and low-pitched sound for low-luminance square) than an incongruent one. This finding suggests a general ability for cross-modality matching in great apes.

A number of recent studies tested production (Grosse et al., 2015) and comprehension (Bohn et al., 2018, 2016) of iconic gestures in chimpanzees and children. A study by Grosse et al. (2015) examined whether chimpanzees and 2-3-year-old children use iconic gestures to instruct a human experimenter on how to use an apparatus. Chimpanzees, unlike human children, did not produce iconic gestures to instruct the human experimenter. In a similar vein, Bohn et al. (2016) tested comprehension of iconic gestures in chimpanzees and 4-year-old children. In this experiment, the experimenter used either iconic or arbitrary gestures in order to inform the subject about the location of a reward. In contrast to children, chimpanzees showed no spontaneous comprehension of iconic or arbitrary gestures. A follow-up study also found no spontaneous comprehension when gestures were enriched with iconic sounds and preceded by a communicative training (Bohn et al., 2018). However, in the initial study, apes learned to associate iconic gestures with a specific location faster compared to arbitrary gestures. According to the authors, apes failed to spontaneously comprehend the gesture because they did not perceive it as communicative. Associative learning of gesture - location correspondence was enhanced in the iconic condition because seeing the gesture shifted apes’ attention to the corresponding apparatus by triggering a memory representation of the bodily movement, from which the gesture was derived, that was used to operate the apparatus. This

evidence might be taken as a hint that apes have at least some tendency toward correctly interpreting at least some iconic manual gestures, thus raising the possibility that also other forms of iconicity may be available to them, which may or may not include sound symbolic congruency detection and matching.

Few hypotheses address the mechanistic cause of sound symbolic mappings. For example, Ramachandran and Hubbard (2001) proposed the “synaesthetic account of sound symbolism”, which is based on putatively innate knowledge about correspondences between visual shapes and phonemic inflections. According to the authors the mechanism behind this effect has an articulatory account. For example, the sharp edges of a spiky shape mimic the sharp phonemic inflections and the sharp movement trajectory of the tongue on the palate when uttering the pseudoword ‘kiki’. The authors see such “synaesthetic correspondence” as important in the emergence of language.

The hypothesis that language and sound symbolic processing are intrinsically related to each other raises the question whether both of these effects are only present in humans, but not in non-human primates. In fact, brain organization in great apes, and in particular chimpanzees, shows reasonable similarity to humans, although there are, no doubt, anatomical differences, which have their correlate at the highest functional level in the presence and absence (or great limitation) of language.

Neuroanatomical studies have shown that a major difference setting apart humans from their closest relatives, chimpanzees, lies in the much stronger and richer development of a neuroanatomical fiber bundle called the arcuate fasciculus (AF) (Rilling et al., 2008; Rilling, 2014). This fiber bundle connects the anterior and posterior language areas in frontal and superior temporal cortex with each other, but also interlinks the ventral visual stream of object related form and color processing with the latter (Catani, 2009). The AF is known to be important for interlinking information about articulatory movements with that about acoustic signals produced by the articulations, thereby laying the ground for abstract phonological representations that span across modalities (López-Barroso et al., 2013; Pulvermüller and Fadiga, 2010; Yeatman et al., 2011). Sim-

ilarly, the AF may play a main role in linking letters to sounds, and it is likely that it stores other types of cross-modal symbolic relationships too. Experimental evidence has shown functional relationships of the AF in humans with their ability to store verbal materials (verbal working memory, VWM) and its general relevance for language processing (Schomers et al., 2017). We hypothesize that a strongly developed human-like AF is also involved in, and necessary for, the kind of abstract cross-modal information linkage required for sound symbolism. This position predicts a fundamental difference in sound symbolic ability between humans and apes which parallels their difference in language capacity.

It is evident that apes can differentiate forms and shapes (Matsuzawa, 1990; Tomonaga and Matsuzawa, 1992) and some research also indicates that they can perceive differences in human speech (Heimbauer et al., 2011; Kojima et al., 1989; Kojima and Kiritani, 1989; Steinschneider et al., 2013). Considering these two abilities, the present study aims to explore whether our closest living relatives process sound symbolic mappings between shapes and sounds. We attempted to replicate existing findings in sound symbolic matchings in human adults using a two-alternative forced choice (2AFC) task under explicit instructions, and performed a similar ape-compatible 2AFC task with a group of touchscreen trained great apes to investigate if any sound symbolic congruency effect would be present.

3.2. Experiment 1

3.2.1. Materials and Methods

Subjects

Twenty-four healthy human right-handed adults (14 females, age $M=25.87$, $SD=5.08$) participated in the study. The subjects were native speakers of different languages (11 German, 3 Greek, 2 Italian, 2 Spanish, 1 French, 1 Bulgarian, 1 Russian, 1 Urdu, 1

Kurdish, 1 Afrikaans). Two of the subjects were bilinguals, one speaking Greek and Albanian, one Afrikaans and English. All subjects had normal hearing and normal or corrected-to-normal vision. Subjects were recruited from announcements at the Freie Universität Berlin. All methods of the study were approved by the Ethics Committee of the Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin and were performed in accordance with their guidelines and regulations. All subjects provided written informed consent prior to the participation to the study and received 10 euros for their participation.

Design and Procedure

Sharp or round shapes were created in Power Point with the freeform tool and edited on GNU Image Manipulation Program (The Gnu Image Manipulation Program Development Team, 2010; www.gimp.org). Each shape was black (RGB 0,0,0) and 350×350 pixels in size. For the selection of the final shapes, a separate group of subjects ($N=110$, recruited online via mailing lists) judged how sharp or round each shape was on a 7-point likert scale, ranging from 1-sharp to 7-round. We selected the 12 most sharp ($M=2.00$, $SD=0.34$) and round ($M=5.32$, $SD=0.58$) shapes, respectively (see Table B.1). For all selected shapes, the sum of responses in the range 1-3 (sharp) or in the range 5-7 (round) was three times higher, than the number of responses for 4-point (neutral) or for the other half of the scale. Auditory stimuli were created based on a previous studies regarding the role of consonants (McCormick et al., 2015; Nielsen and Rendall, 2011) and vowels (Maurer et al., 2006) in sound symbolism. We used combination of vowels and consonants that have been previously reported sounding more ‘round’ or ‘sharp’ respectively. We created trisyllabic or bisyllabic pseudowords with combinations from the following letters : the front vowels /i/ and /e/, the back vowels /o/ and /u/, the fricatives /z/, /s/ and /f/, the voiceless plosives /p/, /t/ and /k/, the nasals /m/ and /n/ and the voiced plosives /g/, /d/ and /b/. A separate group of subjects ($N=92$, again recruited via online mailing lists) rated these pseudowords on a 7-point likert scale, ranging from 1-‘sharp’ to 7-‘round’ sound. The ‘sharpest’ pseudowords had the

combination of the front vowels /i/ and /e/, the fricatives /z/ and /s/ and /f/, and the voiceless plosives /p/ and /k/ (M=2.8, SD=0.22), whereas the ‘roundest’ words were combinations of the back vowels /o/ and /u/, the nasals /m/ and /n/ and the voiced plosives /g/ and /d/ (M=5.4, SD=0.34). For the final experiment, we decided to use bisyllabic pseudowords with a consonant-vowel-consonant-vowel (CiViCiVi) structure, for example “lolo” or “kiki”, based on the combinations of consonants and vowels determined by the online questionnaire. We included 10 ‘sharp’ and 10 ‘round’ pseudowords for each category. The auditory stimuli were recorded in a soundproof booth by a female native Greek speaker in Audacity (2.0.3) (<http://audacityteam.org/>) and afterwards normalized for amplitude. For the list with the final stimuli (see Table B.2).

Both humans and apes performed a 2AFC task. Evidence suggests that apes are able to perceive differences in abstract forms and shapes presented to them on computer screens (Matsuzawa, 1990; Tomonaga and Matsuzawa, 1992). Furthermore, it has been shown that under specific circumstances, apes also perceive differences between human speech utterances (Heimbauer et al., 2011; Kojima and Kiritani, 1989; Kojima et al., 1989). Each trial started with the presentation of a fixation cross for 500 ms followed by the presentation of an auditory stimulus for 800 ms. Next, the two target shapes always one sharp and one round appeared diagonally, on the screen from upper left to bottom right or reverse. These stayed on screen for 1500 ms; during this time, responses were collected. Every trial ended with the presentation of a ‘buzz’ sound lasting 500 ms (see Fig. 3.1). All slides were presented on a grey background (RGB 192,192,192). The experiment was divided into 3 blocks (80 trials each) separated by two pauses in between. In each block, 10 specific combinations assembled from the selected 12 shapes and 10 sounds were used. These repeated within blocks, but were different between blocks. All trials were randomized within each block.

Human subjects sat in a dimly lit room in front of a 23 in. LCD monitor (screen refresh rate 75Hz; screen resolution 1280×1024). The auditory stimuli were presented via two Logitech speakers (Model NO: Z130) located at each side of the screen. Responses were recorded via two-button press on a Serial Response Box™ (SRBox, Psychology

Software Tools, Inc). The experiment was designed in E-Prime 2.0.8.90 (Psychology Software Tools, Inc., Pittsburg, PA, USA). Before the initiation of the experiment, subjects received the following written instructions: “During the experiment two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures that matches the sound you hear.” No specific instructions were given to the participants regarding speed or accuracy. By the end of the experiment subjects completed a computer-based questionnaire about their strategies on shape selection and on their previous knowledge on sound symbolism.

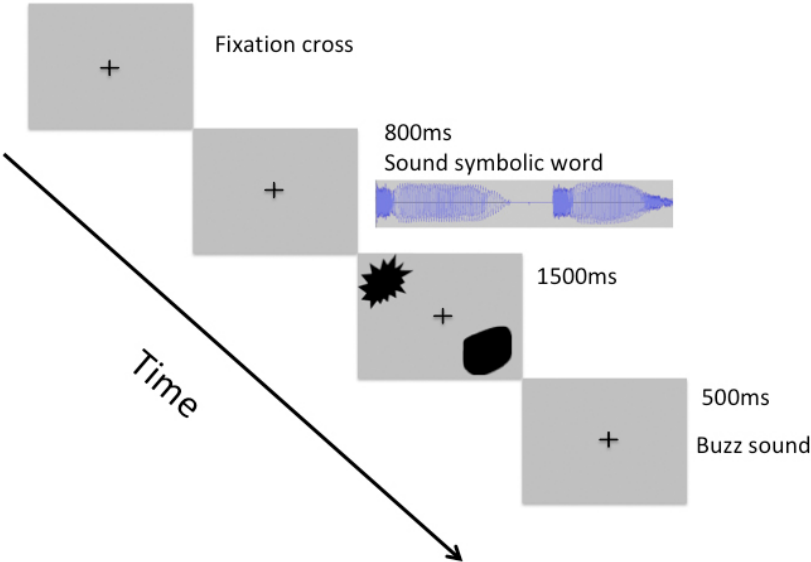


Figure 3.1.: Schematic representation of experimental design of the two-alternative forced choice (2AFC) task applied in humans.

3.2.2. Data analysis

For all analyses, trials with reaction times greater than 1500 ms or non-response were excluded. To check if subjects’ selection of shapes was influenced by the sound, we performed Wilcoxon signed-rank tests to compare the number of congruent (sound sym-

bolic) responses against chance level. In order to check if performance was further influenced by other variables, in an exploratory analysis, we fitted a generalized linear mixed model (GLMM) with a binomial error structure. As analysis tool, R version 3.4.3 was used including the package lme4 (Bates et al., 2014). The dependent variable was congruency that is whether the shape of the selected stimulus matched the shape of the primed sound. We included word type ('sharp' vs. 'round') and trial number as fixed effects. We used a maximal random effect structure with random intercepts for subject, word and for the combinations of the presented shapes and random slopes for each trial nested within these random effects. We used the likelihood ratio test (LRT) to check if the predictor variables improved the fit of the model; these were calculated by comparing the full model to a reduced model that included all terms except for the fixed effect term in question. Chi square and p-values were computed using the function drop1 from the R package lme4. In addition, we compared individual proportions of incongruent responses for 'round' and 'sharp' words using Wilcoxon signed-rank test as well as the individual proportions of congruent responses for each pseudoword category against chance level. Finally, we calculated the proportion of times each subject chose a round or sharp shape (independent of the previous acoustic stimulus) and performed a Wilcoxon signed-rank test.

3.2.3. Results

We excluded 3.1% of the trials obtained from humans, because reaction times were greater than 1500 ms or non-response was given. Humans showed a significant preference for image choices with sound symbolic correspondence to the preceding sounds ($V = 296$; $p = 0.001$; see Fig. 3.3). An average of 71.33% of congruent shape choices contrasted with only 28.67% incongruent responses. In addition, the predictor variable of word type significantly improved the model ($\chi^2(1)=27.30$, $p = 0.001$). Specifically, there were more congruent responses for 'round' than for 'sharp' pseudowords (see Fig. 3.4a). Incongruent responses were primarily seen for 'sharp' words being classified as

‘round’ (6.16% incongruent ‘round’ responses vs. 22.16% incongruent ‘sharp’ responses ($V = 300$, $p = 0.001$) (see Fig. B.1). Corresponding result was revealed by the analysis on the proportion of congruent responses for each pseudoword category against chance, with ‘sharp’ pseudowords perhaps showing a tendency but not significantly exceeding chance level ($V=188$, $p = 0.14$) and ‘round’ congruent responses clearly ending up above chance ($V=300$, $p = 0.001$). Humans selected round shapes in 66.8% of cases, significantly more often than sharp shapes ($V=0$, $p = 0.001$) (see Fig. B.3). Figure 3.3 shows that the range of human performance varied widely from chance to 71.33% congruent responses. Closer examination of the individual subjects’ behavior and performance was conducted to assess whether all subjects performed the task as instructed. It turned out after the experiment when filling out the post-experiment questionnaire, that one individual’s understanding of the English language – the language in which instruction were given – was very limited. Three other participants showed an extreme preference for round shapes, which they chose over 80% of the trials. This is quite unusual behavior (also not paralleled by any of our apes) and we therefore excluded these four ill-behaving subjects. Their results are highlighted in pink in Figure 3.3. Please note that any sound congruency effects in these subjects’ responses were absent, with performance approximating chance. A new analysis conducted on the data from the remaining 20 individuals confirmed the presence for sound symbolic congruent over incongruent responses ($V = 210$; $p = 0.001$). The comparison of individual proportion of incongruent responses for the two pseudoword categories remained significant (5.76% incongruent ‘round’ responses vs. 18.42% incongruent ‘sharp’ responses ($V = 210$, $p = 0.001$). On the other hand, the analysis on the proportion of congruent responses for each pseudoword category against chance revealed that both ‘sharp’ pseudowords ($V=173$, $p = 0.004$) and ‘round’ ($V=210$, $p = 0.001$) exceeded chance levels.

3.3. Experiment 2

3.3.1. Materials and Methods

Subjects

Six chimpanzees (3 females) and two gorillas (2 females) (age $M=20.75$, $SD=13.18$) housed at the Wolfgang Köhler Primate Research Center (WKPC) at Leipzig Zoo, Germany, participated in the study. Apes were never food or water deprived. Food rewards from the study were given in addition to their regular diet. Participation was voluntary and apes could abort the experiment at any time. The study was approved by an internal ethics committee at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Research was non-invasive and strictly adhered to the legal requirements of Germany. Animal husbandry and research complied with the European Association of Zoos and Aquaria (EAZA) Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria and the World Association of Zoos and Aquariums (WAZA) Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums.

Design and Procedure

The study was conducted in the apes' familiar observation or sleeping rooms. We installed an infrared touchscreen (Nexio NIB-190B infrared touch screen) outside to the testing cage. The screen was connected to a 19 in. computer monitor with a resolution of 1280×1024 (aspect ratio 5:4) fixed behind the touchscreen. Sound was played through two loudspeakers placed on the floor next to each side of the monitor, 2 Logitech speakers (Model No: X-120) for the chimpanzees and 2 Logitech speakers (Model No: x-140) for the gorillas. The stimuli were the same as the ones used for humans. We made the following adjustments to the setup: The background of the slides was black (RGB 0,0,0) and the shapes were white (RGB 255,255,255) in order to have high contrast and to maintain the attention of the apes (see Fig. 3.2).

The ape experiment was designed to be as similar as possible to the human one. However, some modifications were necessary to accommodate between-species differences and especially to replace the verbal instruction given to humans by a training procedure with direct reinforcement. Every trial started with an initiation symbol that the ape had to touch in order to start the trial. This self-initiation procedure has been used before to ensure that the apes are attentive at the beginning of the trial (Allritz et al., 2016; Munar et al., 2015). If the ape did not engage with the touchscreen for a certain amount of time, the session was terminated prematurely. After touching the initiation square, a bisyllabic pseudoword was presented for 800 ms followed by the presentation of two shapes diagonally. The response time window was the same as for humans (1500 ms). The last slide was the reward or no reward slide, namely a black blank slide that remained on the screen for 2000 ms, which was either combined with a reward-announcing ‘chime’ sound (Windows XP Default) or not. Within these 2000 ms after the ‘chime’ sound, a reward (a piece of apple) was delivered. We used the ‘chime’ sound, as it has been previously used to announce the delivery of the food reward in the same apes (Munar et al., 2015). There was a 50% chance for a given trial to be followed by a reward-announcing sound and actual reward. This random rewarding procedure was implemented to maintain the subject’s motivation to continue partaking. Note however, that the type of response, whether a sharp or round shape was selected, did not influence the likelihood of the reward, thus excluding any bias toward ‘congruent’ or ‘incongruent’ responses.

To familiarize apes with the 2AFC task, they had to perform up to three habituation sessions. A habituation session consisted of 80 trials in which different combinations of bisyllabic pseudowords, irrelevant to the experiment were presented along with random combinations of shapes used in the experiment. The apes were rewarded every time they selected one of the two shapes followed by the positive ‘chime’ sound. The purpose of the habituation sessions was to assure that the apes would not be surprised or mildly agitated by the sound stimuli and they would touch one of the two shapes within the specific response time window. Five chimpanzees completed one habituation session, one

completed two and both gorillas completed three. In order to move from the habituation to the testing phase the ape had to make a selection 80% of the times within the specific response time window and look at the touchscreen during every trial.

The experiment consisted of 6 blocks of 80 trials each. As in Experiment 1 the combinations of sounds and shapes differed across the 3 blocks. These 3 blocks were the same as those used with humans; with apes, they were repeated to yield the overall 6 blocks. The same sound-shapes trial was not presented in more than one block across the first 3 blocks. The order of trials was randomized within each block. Apes were tested in one block per day to avoid any habituation effects.

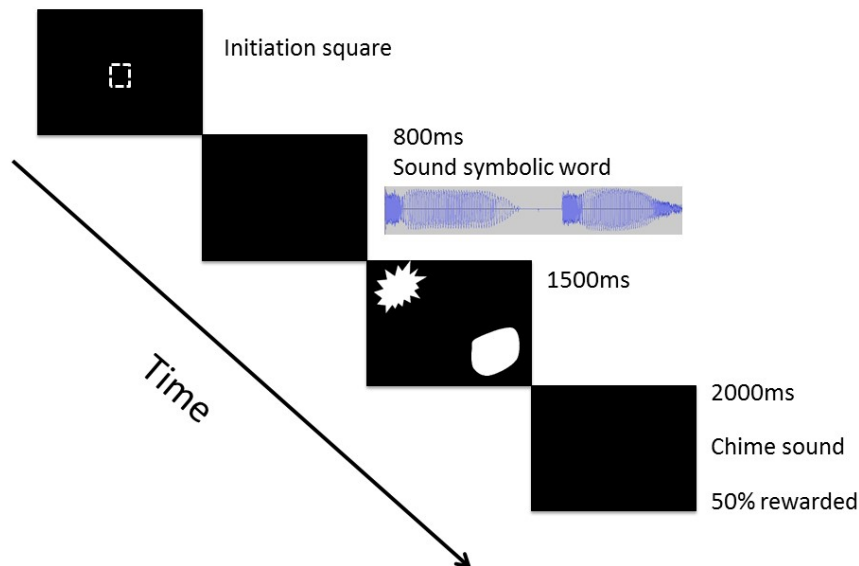


Figure 3.2.: Schematic representation of experimental design of the two-alternative forced choice (2AFC) task applied in apes.

3.3.2. Data analysis

The analyses were similar to the ones conducted for Experiment 1. For apes, we excluded 13.88% trials with responses above 1500 ms or non-responses. The GLMM model

for apes included as dependent variable congruency, that is whether the shape of the selected stimulus matched the shape of the prime sound, and as fixed effects word type, trial and block. We used a maximal random effect structure with random intercepts for participant, word and the trial-specific combination of shapes, as well as random slopes for trial and block. In order to explore any effect of the reward schedule on the performance of apes, we fitted generalized linear mixed models. In the first model, we included as dependent variable the shape category the apes selected ('sharp' or 'round') for each trial and as fixed effects the shape category selected in the previous trial if and only if this trial had been rewarded, as well as the fixed effects trial and block. We used a maximal random effect structure with random intercepts for participant and word, as well as random slopes for trial and block. The likelihood ratio test (LRT) was applied to check if the predictor variable improved the fit of the model; these were calculated by comparing the full model to a reduced model that included all terms except for the fixed effect term in question. Chi square and p-values were computed using the function `drop1` from the R package `lme4`. Finally, we used Mann-Whitney U test to compare the congruency responses between Experiment 1 and 2.

3.3.3. Results

Due to the small sample size of the two non-human primate species we could not make any statistical inferences on their performance separately. However a visual inspection of the results showed no difference in the performance of chimpanzees and gorillas. Numerically, both species performed similarly, with gorillas reaching 51.17% congruent responses and chimpanzees 50.75% (see Fig. B.2). Apes, showed no preference for sound symbolic correspondences ($V=21$; $p = 0.27$) (see Fig. 3.3). There was also no significant difference between the full and the reduced model ($\chi^2(1)=2.28$, $p = 0.13$), indicating that word type (round or sharp), block and trial, considered in conjunction, did not improve the predictive accuracy of the model. Moreover, they tended to have similar congruency effects for 'sharp' and 'round' words (27.16% incongruent 'round'

responses vs. 22.07% incongruent ‘sharp’ responses, $W = 20$, $p = 0.23$) (see Fig. B.1). Furthermore, apes did not indicate also any bias towards selection of one of the two shape types (45.11% round vs. 54.88 % sharp responses; $W=44$, $p = 0.23$) (see Fig. B.3). The result of the reward analysis revealed that the subjects’ choices did not differ significantly depending on whether a reward on a preceding trial was received after touching a round vs. sharp image. Specifically, there was no significant difference between the full and the reduced model after ($\chi^2(1)=0.25$, $p = 0.61$). Thus, in a trial by trial analysis, the shape selected and rewarded in a given trial did not affect the shape selected in the following trial. Comparing the result patterns between Experiment 1 and 2, it can be seen that apes and humans show almost non-overlapping distributions of sound-congruency effects ($W=175$, $p = 0.001$). The four human subjects that performed at a level similar to apes were those with evidence for non-cooperative task performance; after their removal, the distributions were fully distinct. Calculating chi-square tests for each participant, including apes and humans, there was significant above-chance performance for 20 out of 24 human subjects but for none of the apes.

3.4. Interim Discussion

The results from Experiments 1 and 2 suggest that sound symbolic congruency effects are present in humans but not in great apes. However, before we will discuss this putative conclusion in detail, an obvious caveat of the preceding experiments needs to be taken into account. Human subjects were explicitly instructed to perform sound symbolic matchings, whereas apes were trained to respond to pairs of visual displays by selecting one, without any task instruction or other hint about the ‘desired’ outcome being given. This obvious difference and potential confound of the previous results was addressed in Experiment 3 where a new set of human subjects was tested without explicit task instruction hinting at the sound symbolic correspondences our research targets.

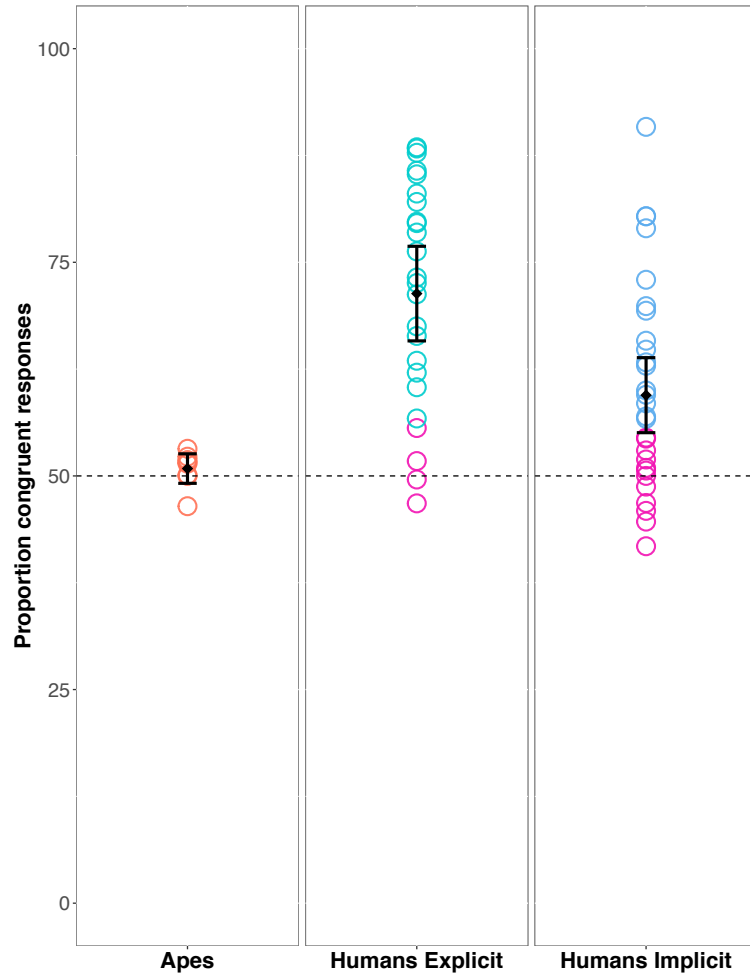


Figure 3.3.: Percentage of sound symbolic congruent responses for apes and for humans performing on the explicit and the implicit 2AFC task, quantified as the proportion of times each individual matched a ‘sharp’ sound to a sharp shape or a ‘round’ sound to a round shape. Orange, cyan and blue circles show the percentage of congruent responses for individual apes and humans for the explicit and implicit instructions separately. Pink circles represent the human subjects that reached the ape performance. Black diamonds represent the average responses for each species and the whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance.

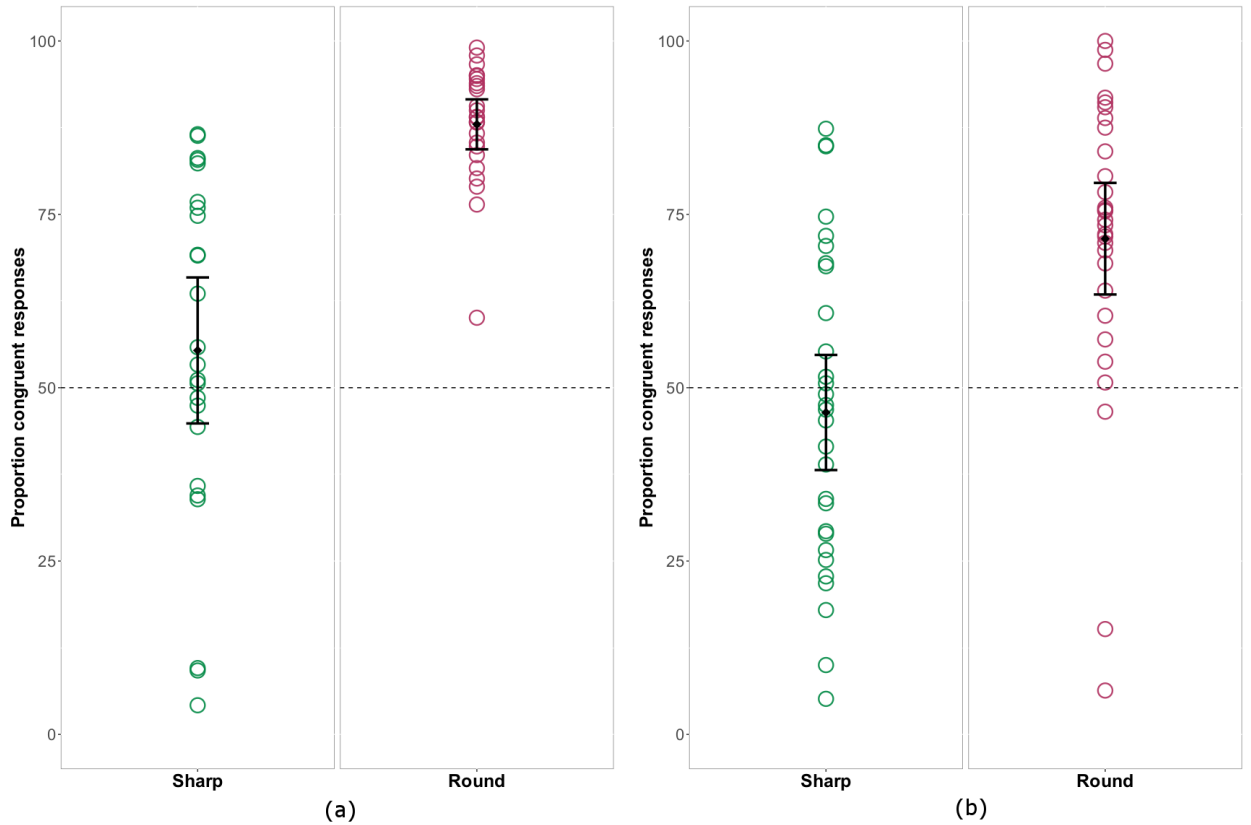


Figure 3.4.: a) Proportion of sound symbolic congruent responses in humans for the two pseudoword categories in the explicit 2AFC task. Green and maroon circles show the percentage of congruent responses for each individual for ‘sharp’ and ‘round’ pseudowords separately. Black diamonds represent the average responses for each pseudoword category and whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance. b) Proportion of sound symbolic congruent responses in humans for the two pseudoword categories in the implicit 2AFC task. Green and maroon circles show the percentage of congruent responses for each individual for ‘sharp’ and ‘round’ pseudowords separately. Black diamonds represent the average responses for each pseudoword category and whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance.

3.5. Experiment 3

3.5.1. Materials and Methods

Subjects

Thirty-one healthy right-handed adults (17 females, age $M=25.35$, $SD=3.56$) participated in the study. The subjects were native speakers of different languages (11 German, 3 English, 3 Spanish, 2 Mandarin, 2 Greek, 2 French, 1 Bulgarian, 1 Italian, 1 Romanian, 1 Czech, 1 Polish, 1 Malaysian). Two of the subjects were bilinguals, one speaking English and Spanish, one Spanish and German. All subjects had normal hearing and normal or corrected-to-normal vision. Subjects were recruited from announcements at the Freie Universität Berlin. All methods of the study were approved by the Ethics Committee of the Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin and were performed in accordance with their guidelines and regulations. All subjects provided written informed consent prior to the participation to the study and received 10 euros for their participation.

Design and Procedure

In order to explore further any possible effect of the explicit instruction given in Experiment 1 on the performance of humans, we conducted an additional experiment in humans similar to Experiment 1. The materials were the same as in Experiment 1. The experimental design and procedure were also alike with the following modifications. We reduced the total number of trials into 2 blocks (80 trials each) separated by one pause in between. In each block, 10 specific combinations assembled from the selected 12 shapes and 10 sounds were used. These were repeated within blocks, but were different between blocks. No ‘buzz’ sound was presented at the end of each trial. All trials were randomized within each block. In addition we modified the written instructions given to the subjects before the initiation of the experiment.

To provide a social motivation for performing on the task, subjects were informed about their reimbursement before the experiment and they received the following written instructions: "During the experiment two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures". Note that this instruction lacks any information about any type of matching to be performed. If such matching is observed in this experiment's context, it cannot therefore be driven by instruction. Furthermore, the instruction did not specify response speed or accuracy. After the experiment, subjects completed a computer-based questionnaire about their strategies on shape selection and on their previous knowledge on sound symbolism.

3.5.2. Data analysis

For all analyses, trials with reaction times greater than 1500 ms or non-response were excluded. Data analyses were the same as in Experiment 1.

3.5.3. Results

We excluded 2.5% of the trials, because reaction times were greater than 1500 ms or non-response was given. As in Experiment 1 humans showed a significant preference for image choices with sound symbolic correspondence to the preceding sounds ($V = 367$; $p = 0.001$). An average of 59.44% of congruent responses contrasted with 40.56% incongruent responses. However, the performance of more subjects dropped to chance level compared to Experiment 1 (see Fig. 3.3). The predictor variable of word type significantly improved the model ($\chi^2(1)=30.05$, $p = 0.001$). In accordance with Experiment 1, there were more incongruent responses for 'sharp' than for 'round' pseudowords (see Fig. 3.4b). Once again, the incongruent responses were primarily seen for 'sharp' words being classified as 'round' (13.65% incongruent 'round' responses vs. 26.89% incongruent 'sharp' responses, $V = 433$, $p = 0.001$) (see Fig. B.1). A corresponding result emerged from the analysis of the proportion of congruent responses for each pseudoword category

against chance, with ‘sharp’ pseudowords not exceeding chance level ($V=200$, $p = 0.82$) and ‘round’ congruent responses being significantly above chance ($V=446$, $p = 0.001$). As in Experiment 1 humans selected round shapes in 63.21% of cases, significantly more often than sharp shapes ($V=60$, $p = 0.001$) (see Fig. B.3). Calculating chi-square tests for each participant there was significant above-chance performance for 18 out of 31 human subjects. The comparison between Experiment 2 and 3 revealed again that apes and humans show non-overlapping distributions of sound-congruency effects ($W=177$, $p = 0.03$). In contrast, there was no significant difference in the performance of human subjects between Experiment 1 and 3 ($W=188$, $p = 0.99$).

3.6. Discussion

The present study used a 2AFC task to test whether humans and great apes spontaneously detect sound symbolic correspondences between abstract visual shapes and meaningless word-like combinations of speech sounds. Results indicate that humans’ forced choices of shapes were significantly biased by prime sounds towards selection of shapes that showed sound symbolic congruency with the primes, whereas great apes did not give evidence of any such sound symbolic congruency detection. In our human populations, this sound symbolic effect was mainly carried by ‘round sounding’ pseudowords. Whereas they may have tended to select sharp shapes more frequently than round ones after perceiving correspondingly ‘sharp sounding’ syllable combinations, only the opposite preference in favor of round shapes was clearly manifest after ‘round sounding’ syllables. The same general result was obtained after explicit task instruction to “match shapes to sounds” (Experiment 1) and similarly when humans were given just a picture selection task instruction with the sound symbolic task aspects remaining fully implicit and opaque (Experiment 3). When humans were explicitly instructed to match pseudowords to shapes in Experiment 1, only four of them performed, similarly to all of the apes, at chance level, which suggests lack of task instruction understanding in these human individuals. However, performance dropped to chance level for 18 out of 31

human subjects in the ‘implicit’ Experiment 3, while still remaining significantly above chance at the level of group statistics.

There are obvious limitations in testing different species on tasks aiming at higher cognitive abilities such as cross-modal congruency processing. Although we chose a general task applicable to humans, chimpanzees and gorillas, namely the 2AFC, we had to introduce some modifications to adjust it to testing great apes. We will discuss these differences between the 2AFC tasks one by one. First, humans performed all testing blocks in one session, whereas apes performed one block per testing session, completing six sessions in total. Included in our main analysis, the predictor "block" did not modulate the apes’ performance, thus arguing against this difference being relevant for explaining between-species differences in performance. Second, humans registered their answers through a keyboard, whereas apes used a touchscreen; however, it is not obvious why humans should have responded differently in the Experiments 1 or 3, had they been using a touch screen. Concerning potential touch location biases that the apes may have shown, note that the position of the round and sharp shapes were balanced across trials so that such a bias could also not have influenced the results. It was critical to provide randomly delivered food rewards to the apes to train them for the task, to compensate for the impossibility to use verbal instruction, and in order to continuously motivate them and keep them engaged across testing. Indeed the 50% administration of food reward which was orthogonal to the task was efficient in that subjects were motivated to complete the data collection. Moreover, our reward analysis showed that the presence of the reward had no effect in shaping the choice of shape selection on a trial by trial basis in apes. Although humans received no reward in a trial by trial schedule, they were socially rewarded by monetary compensation, and they were made aware of such social reward before starting the experiment. Especially the social reward for performing choice responses to pictures (without further instruction) in Experiment 3 seems to us a reasonable match of the unspecific food reward our apes received.

The verbal instruction for humans in Experiment 1 to select a shape “that matches the sound” were reasonably efficient as well, although three out of 24 subjects did not follow

them well. In order to exclude any possible effect of the explicit verbal instructions on humans' performance in Experiment 1 and on explicitly paying attention to the pseudoword, we conducted Experiment 3. In this Experiment we performed a similar 2AFC task but with 'implicit' instructions. This means that participants were not instructed to pay attention to the sounds and they were not asked to select a shape that "matches the sound" as in Experiment 1. Instead, they were just instructed to select a shape with any sound symbolic aspects of the task remaining fully implicit and opaque to the participant. The results indicate that participants' performance dropped compared to Experiment 1 and relatively more humans performed at chance level. However, and crucially, the different task instructions of Experiment 1 and 3 did not significantly alter the sound-symbolic performance pattern in humans.

In the future, it may be worthwhile to adopt a direct reinforcement paradigm to humans to potentially efficiently motivate consistently cooperative task performance in this species too. This could be done by using a food reward as with the apes, or, more conventionally, by providing the monetary reward piecemeal, on a trial by trial basis. However, it seems unlikely that such 'reinforcement instruction' may change the strong preference of human subjects for sound congruent responses as showed in Experiment 1 and 3. After all, social reward by reimbursement at the end of the experiment and possibly the self-reward resulting from the knowledge of acting as a cooperative experimental subject were already sufficient for allowing sound symbolic effects to emerge. Therefore, we do not believe that the remaining differences between the tasks applied in this study had a significant influence on the patterns of results obtained, and especially on the presence of the sound symbolic effect in humans.

Crossmodal similarity processing in apes and humans

Even though the present study found no sound-shape correspondences in great apes, there is evidence that apes are sensitive to crossmodal mappings. As mentioned, Ludwig et al. (2011) showed that apes are able to process crossmodal correspondences between

pitch and luminance, as they matched a high luminance stimulus to a high-pitched sound and a dark stimulus to a low-pitched sound. In a similar vein, another study showed that great apes can detect visual-auditory structural isomorphic patterns. In that study, two apes were trained to choose a symmetric visual sequence (e.g., two identical geometrical shapes separated by a different third shape in between, for example $\circ\square\circ$) (Ravignani and Sonnweber, 2017). During the testing phase, the apes were presented with the trained symmetric visual pattern and with a non-symmetric pattern (e.g., two identical shapes followed or preceded by a third shape, for example $\circ\circ\square$). The visual presentation was preceded by an auditory pattern, either a symmetric (e.g., two high tones separated by a low tone) or non-symmetric one (e.g., two high tones preceded or followed by a low tone), which was either congruent or incongruent with the structure of the target trained symmetric visual sequence. When the presentation of the pattern was preceded by congruent auditory patterns, response latency to the symmetric visual patterns were shorter compared to when they were preceded by incongruent auditory patterns. The authors interpreted this result as evidence for crossmodal structure processing (priming) in chimpanzees.

In spite of these indications that apes can process cross-modality structural similarities, we did not find evidence for a matching between the visual and auditory domain for spoken pseudowords and contour stimuli. To what degree this lack of crossmodal interaction depends on the specific sound and visual stimuli used, their familiarity and specificity to the species, requires further study. Correspondences of the pitch-luminance type, could be explained by a common neuronal system of magnitude or energy (high vs low acoustic/light energy) across modalities, or simply by a ‘more or less’ in sensory neuronal activation (Walsh, 2003). The analogy between symmetric and asymmetric patterns across modalities can be formulized in terms of abstract structural patterns such as ‘ABA’ vs. ‘AAB’, and could be taken as evidence for abstract processes generalizing away from the individual stimuli and across modality-independent patterns. In contrast, the sound symbolic congruency between abstract shapes and pseudowords is not easily captured by comparable abstract rules or differences in magnitude or en-

ergy. If, for example, the articulatory account of sound symbolism is true, which posits that the maluma-takete effect stems from similarities between shapes and the tongue’s movement trajectory in the mouth, this may explain why apes in our study did not give evidence of processing this congruency that humans apparently perceived. Still, one may object that apes are well-capable of lip smacking and tongue clicking (Bard, 1998; Fedurek et al., 2015; Parr et al., 2005), thus offering a potential basis for sensorimotor knowledge about sound symbolic correspondences, too. Based on this inconsistent picture, a question remains whether a similar congruency effect could be found in great apes, if picture and sound stimuli were more attuned to their species. This provides a possible reason why apes did not give evidence of processing such congruency. However, it still leaves open the important question which features of visual and acoustic materials make these items subject to sound symbolic congruency.

Bias toward congruency for ‘round’

In both Experiment 1 and 3, humans gave more ‘congruent’ than ‘incongruent’ responses to ‘round’ than for ‘sharp’ pseudowords, and the predominance for congruent over incongruent responses was consistently significant only for the ‘round’ items. One may argue that the ‘round’ pseudowords we selected were more sound symbolic on average than the ‘sharp’ pseudowords, or that sound symbolic effects are generally carried by ‘round’ items only. However, the average scores on the ratings of the selected pseudowords could not support these hypotheses. The average “round- vs sharpness” ratings for the ‘round’ words were ($M=5.4$, $SD=0.34$), and the ‘sharp’ words ($M=2.8$, $SD=0.22$) were both equally far from the midpoint of the Likert scale (4.0) for ‘sharp’ words ($V=0$, $p = 0.001$) and for ‘round’ words ($V=190$, $p = 0.001$). Even in the absence of a general bias in stimulus selection, a natural propensity in favor of congruent round responses was reported previously in the literature on sound symbolism (Fort et al., 2018; Jones et al., 2014). Human children show an earlier and stronger sound symbolic effect for ‘round’ pseudowords (Fort et al., 2018), but a much weaker effect for ‘sharp’ ones. A

possible explanation for this general stronger effect of sound symbolism for ‘round’ pseudowords could be the natural tendency of people to prefer round versus sharp shapes, which has been reported earlier (Bar and Neta, 2006; Bertamini et al., 2016; Palumbo et al., 2015). A strong preference for preferring round over sharp shapes was also clearly evident from human performance in the present experiments (Experiment 1: 66.8 vs. 33.2%; Experiment 3 : 63.21 vs 36.79%). The observed difference in favor of ‘round’ pseudoword congruency responses and to the disadvantage of ‘sharp’ sounding words therefore appears to be the result of a response bias.

A preference for curved contours was found previously also in apes on a 2AFC task. Apes, in contrast to humans, preferred curved contours only when the presented items remained on screen until a response was registered, whereas humans preferred curved contours only after short presentation (80 ms) of the two item types (Munar et al., 2015). Our present experiment with apes did not show any significant bias in favor or round shapes. Contrasting with the human pattern, our apes tended to have similarly absent congruency effects for ‘sharp’ and ‘round’ words as well as similar probabilities of selecting round and sharp shapes (see Results for Experiment 2).

The lack of a preference for curvature in our study of eight apes stands in contrast to the findings by Munar et al. (2015), whose study of apes was conducted at the same facility and used a similar method. Their sample of apes was also of similar size ($N = 9$), four of whom participated in the present study. Differences in the types of stimuli that were used may explain why the original finding in apes was not replicated in this study. It is also possible that curvature preference may be too subtle to be detected reliably in small samples of apes, or it may be subject to procedural moderators.

Sound symbolism is specific to humans

Our present data show that apes and human subjects produce clearly distinct response patterns of sound symbolic congruency effects. Whereas humans in both Experiments 1 and 3 consistently showed clear significant sound symbolic preferences at a population

level, not a single ape did so. Even with non-cooperative subjects included in the human sample, there was a clearly significant between-species difference in the group analysis both between Experiment 1 and 2 ($W=175$, $p = 0.001$) and between Experiment 2 and 3 ($W=177$, $p = 0.03$). Although sound symbolic congruency detections in humans seemed to be more clearly apparent in Experiment 1 than in Experiment 3, there was no significant performance difference between their results ($W=188$, $p = 0.99$). This robust difference may be related to the fundamental difference between the species in language ability. Humans share complex languages with large vocabularies and great combinatorial power as tool kit for communication whereas in apes, such a system is absent. We therefore suggest that sound symbolism may emerge from the same neuroanatomical connectivity that is also necessary and essential for the brain's neuronal language circuits. If correct, this implies that human specificity of sound symbolism can be tracked down to anatomical differences between apes and humans revealed by comparative neuroanatomical data (Rilling, 2014). Comparative data suggest an expansion of the connectivity between perisylvian cortical areas involved in language in humans, which those in apes largely lack (Rilling, 2014). In particular, the AF, a left-lateralized long-distance corticocortical connection between inferior-frontal and posterior-temporal cortex, is relatively more strongly developed in humans (Rilling et al., 2008; Rilling, 2014). Recent evidence from a computational model in human and non-human primates' perisylvian language networks, showed better verbal working memory in humans (Schomers et al., 2017) explaining in part the weaker auditory memory documented in non-human primates (Scott and Mishkin, 2016; Scott et al., 2012). The limited verbal working memory in apes prevents their word learning and phonological retrieval capacities, and these may also be fundamental for creating a repertoire of sound symbolic associations for social-interactive communication. It is also possible that, all other things being equal, humans exploit their AF connections when learning associating speech sounds/words and visual stimuli/abstract shapes. This is because the AF connects anterior language areas with both visual and auditory sites. The better developed AF in humans may therefore contribute to the possibility to store and process sound symbolic congruency, as it is crucial

for building the brain's language and phonological network. However, it is important to note that this is still a hypothesis. On its background, testing a language-trained ape for sound symbolic congruency processing appears as relevant. If anatomical connectivity structure determines sound symbolic processing ability, a language trained ape should still be unable to show it. In case sound symbolism is closely linked to language learning, we may predict sound symbolic congruency processing in apes with some linguistic competence.

To conclude, these results show no behavioral indication that great apes spontaneously perceive, recognize or infer cross-modal congruencies between speech sounds and abstract visual displays, whereas humans clearly show this type of crossmodal effect in both explicit and implicit 2AFC tasks. We suggest that the human specificity of sound symbolism may be linked to neuroanatomical differences between humans and apes in the connectivity structure of the perisylvian cortex which provides the basis for human language and possibly sound symbolic congruency too. Sound-shape mappings of this type might indeed have played a significant role in shaping human language.

4. Action sound-shape congruencies explain sound symbolism

This chapter is based on :

Margiotoudi, K., & Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports*, 10(1), 1-13.

doi: <https://doi.org/10.1038/s41598-020-69528-4>

The original article has been published under a (CC-BY) license. This version of the article may not exactly replicate the final version published. It is not the version of record. **Author contributions:** study concept and design (KM and FP), material generation and data collection (KM), data analysis (KM), manuscript drafting (KM), revisions (KM and FP).

Abstract

Sound symbolism, the surprising semantic relationship between meaningless pseudowords (e.g., ‘maluma’, ‘takete’) and abstract (round vs. sharp) shapes, is a hitherto unexplained human-specific knowledge domain. Here we explore whether abstract sound symbolic links can be explained by those between the sounds and shapes of bodily actions. To this end, we asked human subjects to match pseudowords with abstract shapes and, in a different experimental block, the sounds of actions with the shapes of the trajectories of the actions causing these same sounds. Crucially, both conditions were also crossed. Our findings reveal concordant matching in the sound symbolic and action domains, and, importantly, significant correlations between them. We conclude that the sound symbolic knowledge interlinking speech sounds and abstract shapes is explained by audiovisual information immanent to action experience along with acoustic similarities between speech and action sounds. These results demonstrate a fundamental role of action knowledge for abstract sound symbolism, which may have been key to human symbol-manipulation ability.

4.1. Introduction

Sound symbolism is an umbrella term that covers the non-arbitrary associations between meaningless speech sounds and sensory or other meanings Hinton et al. (2006)(for a review, see Lockwood and Dingemans, 2015). The iconic links between pseudowords and abstract visual shapes is the most popular demonstration of this phenomenon. In the present study, the term "sound symbolism" will refer to these latter associations. In his seminal book entitled "Gestalt Psychology", Köhler (1929) described the classic "maluma-takete" paradigm in which humans match a round figure to a ‘round’ sounding pseudoword, such as ‘maluma’, and a sharp figure to a ‘sharp’ sounding pseudoword such as ‘takete’, thus presupposing an abstract ‘resemblance’ between the otherwise meaningless symbol (pseudoword) and the corresponding shape, possibly based on shared

modality general abstract properties. Many experimental studies confirmed Köhler's example and demonstrated the postulated iconic speech-sound/meaning mappings across languages (Blasi et al., 2016; Dingemanse et al., 2016; Perniss et al., 2010), even at early age (for a meta-analysis, see Fort et al., 2018) and across stimulus modalities (Koppensteiner et al., 2016; Shinohara et al., 2016). Furthermore, the ability to perform well on sound symbolic tasks has been related to word learning capacity in young children (Imai et al., 2008; Kantartzis et al., 2011; Maurer et al., 2006).

These results led to some skepticism towards the linguistic Saussurean position that the relationship between form and meaning of signs is arbitrary (Saussure, 1959) and even suggest an important role of sound symbolic mechanisms in language development (Perniss and Vigliocco, 2014) and evolution (Imai and Kita, 2014). Specifically, vocal iconic mappings between infants' first spoken words and the referents these words are used to speak about appear to be substantial, so that iconic signs may have a special status for our ability to talk about things not present in the environment, a feature sometimes called 'displacement in communication' (Perniss et al., 2010). Today, iconicity and sound symbolism along with their bootstrapping role in language development and evolution are widely upon agreement (Imai and Kita, 2014), with recent evidence coming from a study in great apes showing the human specificity of sound symbolic mappings. Margiotoudi et al. (2019) tested humans and great apes in the same two-alternative forced choice (2AFC) task. Both species were presented with different 'round' vs. 'sharp' sounding pseudowords and were required to select a (round vs. sharp) shape that best matched the pseudoword. Humans but not great apes showed significant congruency effects. These results suggest that, similar to language, sound symbolism is a human-specific trait. It has also been argued that sound symbolism may depend on human-specific neuroanatomical connectivity also relevant for language (Rilling et al., 2008; Rilling, 2014), in particular on the presence of strong long-distance connection between frontal and temporal perisylvian areas (Margiotoudi et al., 2019).

Despite the numerous studies documenting sound symbolism, few theories attempt to explain the underlying mechanism. Sound symbolism may be considered as a spe-

cific type of crossmodal correspondence implicating the matching of shared sensory or semantic features across modalities (Spence, 2011). In this spirit, the frequency code theory proposed by Ohala (1994) states that the association of large (small) objects with segments of low (high) frequency, such as vowels having low (high) second formant (i.e., /o/ vs. /i/) is due to the statistical co-occurrence of these features in nature. For instance, large (small) animals vocalize in low (high) frequencies due to differences in the size of their vocal apparatuses; large animals have large vocal apparatuses resulting in the production of lower frequencies compared to smaller animals. However, whereas this explanatory scheme applies nicely to phonetic-acoustic correspondences, to small vs. large shapes, it is not immediately clear why sharp and round shapes should tend to co-occur with certain phonemes and articulations. Therefore, this approach seems to be too limited to provide a full account of sound symbolic effects. A related perspective puts that crossmodal links between acoustic and visual information may be based on the amount of energy across modalities, and therefore on a ‘more or less’ in sensory neuronal activation (Walsh, 2003). Whereas this position seems well-suited to provide a candidate account for the correspondences of ‘vivid’ and ‘flat’ speech sounds and colors (Johansson et al., 2019a; Moos et al., 2014), it would need to be shown how an explanation of the mapping of round abstract figures on the pseudoword ‘maluma’ and one of spiky stars and edges on "takete" could be marshalled along these lines. Therefore, also this approach seems to be too limited to provide a full account of sound symbolic effects.

An eminent and highly cited theory addressing the mechanism of sound symbolism specifically, also highlighting its putative importance for the emergence of protolanguages in language evolution, is that of Ramachandran and Hubbard (2001). The authors propose a "synaesthetic articulatory account" of "maluma-takete" type of associations between meaningless pseudowords and abstract visual forms. In their "bouba-kiki" example, the authors explain that the sharp edges of a spiky shape mimic the sharp phonemic inflections and the sharp movement trajectory of the tongue on the palate when uttering the pseudoword "kiki". Hence, the principal idea is that there are non-arbitrary

mappings between features of tongue movement trajectories which characterize the articulatory act and lead to the production of characteristic speech sounds. Ramachandran and Hubbard (2001) propose that these spatial characteristics and acoustic effects of the articulatory act provide the glue essential for sound symbolic iconic knowledge and that this knowledge became the basis for the emergence of protolanguages and for linking spoken signals to referent objects. However, as to the best of our knowledge, there is no strong experimental evidence supporting this synaesthetic articulatory model.

Ramachandran and Hubbard's proposal can be criticized on theoretical and empirical grounds. The knowledge most crucial for bridging between visual and speech modalities, that about the movement trajectory of the tongue, is part of procedural knowledge and therefore not necessarily and easily accessible to the cognizing individual Ouni (2011). Decades of phonetic research were necessary to document articulatory trajectories, first with x-ray and later- on with electromagnetic articulography (Bresch et al., 2008; Schönle et al., 1987), to find out about the complex and sometimes surprising moves and turns of different articulators in speech production (Browman and Goldstein, 1992; Fowler and Saltzman, 1993; Fuchs and Perrier, 2005). A simple abstract shape, such as a spiky star, appears as a quite distant approximation of such complexity. Unfortunately, the most important articulator, the tongue, is hidden in the mouth and therefore not visible to speakers or listeners. Making the visual features of these movements the key component of the explanation of sound symbolism may therefore appear as questionable from a theoretical perspective. Until now, a systematic comparison of articulatory trajectory features characterizing the production of pseudoword forms such as "takete-maluma" and the abstract shapes these spoken items respectively relate to according to sound symbolic experiments is still missing, so that it remains unclear whether this model can account for the range of phonetic contrasts leading subjects toward selection preferences for sharp and round shapes.

Furthermore, experimental evidence can be marshalled against the most established explanation attempt for sound symbolism: It is well known that dark and light vowels, such as /u/ vs. /i/ lean toward 'sharp' vs. 'round' interpretations (Maurer et al.,

2006; Nielsen and Rendall, 2013) although these are not associated with clear movement trajectory contrasts that could motivate such sound symbolic links. As shown in Fig. 4.1a & b the shapes of the classic "maluma-takete" example show little resemblance to the shapes of the tongue position of a typical 'sharp' sound, /i/, or that of a typical 'round' one, /u/. Both tongue shapes look very similar to each other and differ only with respect to the (backness) position (high at the front vs. back) of the anterior part of the tongue, without showing different sharpness vs. roundness features for the two vowels. Similarly looking at the kinematic trajectories of the tongue root while vocalizing different vowels (see Fig. 4.1c), there is nothing such as edgy shapes in the trajectories for the 'sharp' sounding vowels /i/ and /e/, or cloudy shapes for the 'round' sounding vowels /o/, /u/ and /a/. Likewise, when looking at lip trajectories recorded with articulatory during the production of syllables such as /pi/ vs. /ba/, which again lie on opposite sides of the round-sharpness continuum, the movements appear equally smooth (see Fig. 4.1d). These examples seem incompatible with the idea of similarities between the 'round-' and 'sharpness' of speech sounds on the one hand and articulator shapes or trajectories on the other; thus, they argue against the proposed articulatory account of sound symbolism.

Whereas the tongue and a range of other important articulators are hidden in the mouth, other body parts are clearly visible to the acting individual. Particularly hand movements, are clearly visible to the person performing them and to any interacting partners. When learning to move and, later on, to perform complex goal directed actions, the information about how to perform an act and the perceptual aspect, how the movement is carried out and how the gestures look and sound like, go together and can be associated in a Hebbian learning process (for discussion, see Pulvermüller, 1999, 2018a). As a result, sensorimotor representations develop in the brain. Computer simulations of learning in cortex indicate that these multimodal representations are carried by distributed and connected groups of neurons interlinking action and perception knowledge, so-called action perception circuits. These multimodal neuronal devices can provide a basis of crossmodal information exchange and for the computation of the shape

of a movement trajectory based on the motor schema or vice versa. We here explore the possibility, that these action perception circuits for hand actions provide the mechanistic basis of sound symbolic associations. If this is the case, we would not only expect that human subjects show corresponding abilities a) to detect sound symbolic congruencies and b) to match hand action sounds to the visual forms resulting from action trajectories, but we would also expect these abilities to be correlated across individuals, so that experts in sound symbolism would also be excellent sensorimotor action mappers and vice versa, whereas individuals less skilled in one of the tasks should also perform not-so-well on the other. This leads to the *primary hypothesis*, that there is a significant correlation between subjects' ability to perform sound symbolic mappings and their performance on solving sound-shape mapping tasks for hand actions. In particular, any such correlation should be significantly stronger than any correlation between the performances on the sound symbolic task and a control condition closely matched to the latter, which, in our present case, was the 2AFC. The new model would also postulate that sound symbolic mappings are a by-product of action mappings, due to analogies and physical correspondences between the acoustic features of action sounds and speech and similarities between typical sound symbolic shapes and the shapes resulting from action trajectories. This latter postulate implies further important secondary predictions: that there are further significant correlations between subjects' abilities to map information about actions and sound symbolic entities across modalities and domains, that is, between action sounds and abstract visual shapes and, furthermore, between maluma-takete-like pseudowords and the shapes of hand action trajectory shapes.

To test these novel predictions, we performed using the same 2AFC paradigm, (1) the classic sound symbolic (or *SoSy*) "maluma-takete" experiment along with three others. (2) A hand *Action* condition examined the matching of visual and acoustic aspects of pen drawing, whereby the sounds of the pen moving on the paper when drawing elementary visual shapes led to the acoustic stimuli and the corresponding visual items were the visual shapes, produced by moving the pen. In both, conditions (1) and (2), half of the stimuli were round and the other half sharp. The remaining two conditions

resulted from crossing of the former two, so that (3) hand action-produced visual stimuli had to be selected for sound symbolic pseudowords (*Crossed1 condition*) and (4) sound symbolic abstract shapes (or the "maluma-takete" type shapes) for hand action sounds (*Crossed2 condition*). As a further condition, a control 2AFC task was administered with animal pictures and the sounds the depicted animals typically produce, so as to probe general sensorimotor knowledge unrelated to shape-sound correspondences intrinsic to human-specific actions. The Animal task was administered to obtain an estimate of performance with variations in 2AFC task performance, evaluating general attentional, motor or perceptual skills across the test population. At the end of the experiment, an additional paper-and-pencil attention test (6) was administered to control for variability in the subjects' performance level on a sustained attention task. We predicted that, if action knowledge links underlie the sound symbolic mapping of auditory to visual features and vice versa, specific significant correlations across all action and sound symbolic tasks would emerge, that is, across conditions (1)-(4), but not between tasks (1)-(4) and any of the control tasks (5) or (6).

4.2. Materials and Methods

Subjects

Twenty-four right-handed adults (20 females, age $M=25.04$, $SD=3.47$) participated in the study. The subjects were native speakers of different languages (8 German, 3 Turkish, 2 Mandarin, 2 English, 2 Greek, 2 Arabic, 1 Spanish, 1 Italian, 1 Albanian, 1 Cantonese, 1 Hungarian). To assure that all subjects understood the oral instructions given in English, all participants successfully completed the online Cambridge Assessment English test for the English language prior to the experiment. In order to be eligible for the study, subjects had to have on the aforementioned test a score equal to or above the B1 level in English. All subjects had normal hearing and normal or corrected-to-normal vision. One subject could not complete the experiment due to health issues and her data

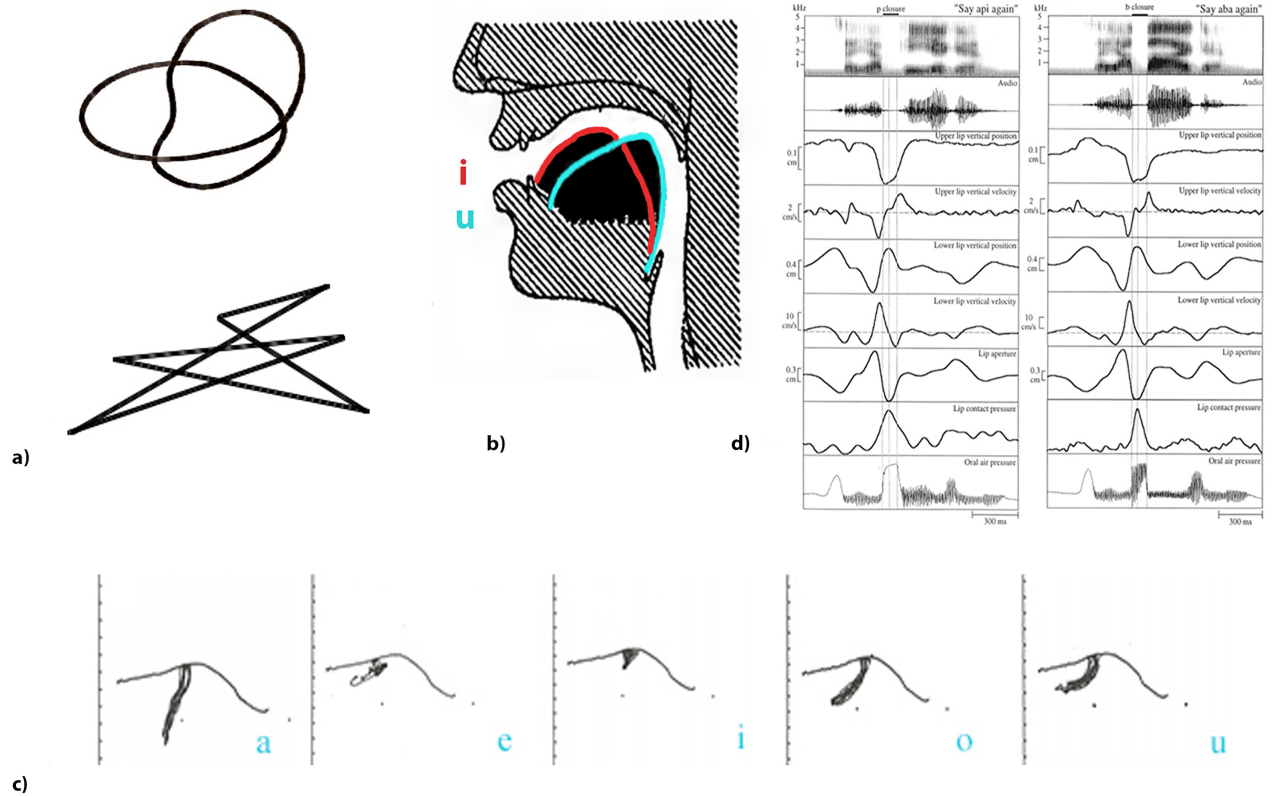


Figure 4.1.: a) Köhler's original stimuli "maluma-takete". The upper shape corresponds to the pseudoword 'maluma' and the lower to 'takete'. Reproduced from Köhler (1929). b) Tongue positions of the vowels /i/ (in red) and /u/ (in turquoise). The shape of the tongue for the vowel /i/ does not resemble the edgy "takete" figure depicted at Köhler's work. Adapted from Jones (1922). c) Kinematic trajectories of the tongue root while uttering the vowels /a/, /e/, /i/, /o/ & /u/ Schönle et al. (1987). d) Movements/velocities of lips during the production of the pseudowords "api" (left panels) and "aba" (right panels). Note the absence of any similarity between movement trajectories and 'sharp' shapes (such as the lower item in panel a). Reproduced from (Löfqvist and Gracco, 1997).

was therefore excluded from the analysis. Subjects were recruited by way of written announcements at the Freie Universität Berlin. All methods of the study were approved by the Ethics Committee of the Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin and were performed in accordance with relevant guidelines and regulations to the Declaration of Helsinki. All subjects provided written informed consent prior to the participation to the study and received 20 euros for their participation.

Stimuli

We included the following stimulus types:

- Sound symbolic abstract shapes (SS_{sh}): Twenty shapes, all of them similar to shapes commonly used in experiments on sound symbolism, were selected from Margiotoudi et al. (2019), 10 sharp and 10 round ones. However, whereas filled versions had previously been used, we here followed Köhler’s original strategy using black-on-grey (RGB 0,0,0 vs. RGB 192,192,192) line drawings (size: 350×350 pixels). This was done to achieve similarity to the action shapes (see Fig. C.1).
- Sound symbolic pseudowords (SS_{pwd}): Twenty bisyllabic SS_{pwd} previously used and described in the Experiment of Margiotoudi et al. (2019). These included items typically used in sound symbolic experiments, such as "kiki" and "momo". We adopted 10 ‘sharp’ and 10 ‘round’ sounding SS_{pwd} . All recordings were saved at 44.1 kHz sampling rate with an average duration of all SS_{pwd} $M = 578 \pm 41.28$ ms.
- Action shapes ($Action_{sh}$): Action shapes were generated by drawing a selection of abstract shapes. We focused on elementary geometric shapes, such as circle, oval, sine wave and triangle, saw tooth, plus slightly more complex figures including two of the elementary shapes, e.g., small circle/triangle embedded in a larger one, figure-of-eight/hourglass figure. We selected the 10 shapes whose corresponding sounds had previously been rated the 5 best ‘round’ and ‘sharp’ sounding ones. A further rating ($N=13$, by subjects recruited online via mailing lists) ascertained

that the 10 action shapes selected were also among either the five most ‘sharp’ ($M=1.30$, $SD=0.47$) or the five most ‘round’ rated ones ($M=6.53$, $SD=0.32$); the ratings of these stimulus grounds significantly different from each other ($W=25$, $p = 0.01$) as revealed by a Wilcoxon signed-rank test (see Fig. C.2).

- Action sounds ($Action_{snd}$): A pen producing a clearly audible (but not uncomfortable) sound was used to generate sounds while drawing the abstract shapes of the action shape condition described above. Recordings were taken in a sound-proof booth, using a stereo built-in X/Y microphones Zoom H4n Handy Recorder (Zoom Corporation, Tokyo, Japan) saved at 44.1.kHz. For rating the action sounds, a separate group of subjects ($N=41$, recruited online via mailing lists) judged the ‘sharpness’ or ‘roundness’ of each hand drawing recording on a 7-point Likert scale, ranging from 1-completely ‘sharp’ to 7-completely ‘round’. We selected the five action sound recordings receiving the highest ‘sharp’ ratings ($M=2.18$, $SD=0.23$) and the five ‘round’ ones ($M=5.23$, $SD=0.40$). These corresponded to the shapes selected for the action shape category described above. The rating scores obtained for these two subgroups of action sounds were significantly different from each other ($W=25$, $p = 0.007$), as revealed again by a Wilcoxon signed-rank test. The $Action_{snd}$ were edited to make them acoustically comparable to the bisyllabic SS_{pvd} , which all consisted of two syllables. To this end, we restricted the length of each action sound so that it included only the first two acoustic maxima and therefore resembled a bisyllabic speech item (see Fig. 4.2 a & b). Moreover we applied fade in and out functions for the first and the last 100 ms, so as to remove any on-and offset artifacts. The average duration of action sounds was $M = 934 \pm 473.19$ ms.
- Animal pictures: Twenty pictures of common animals, two for each animal species, were selected. As preliminary testing showed ceiling performance on the animal-picture-sound matching task, animal pictures were slightly blurred to introduce a level of difficulty in the task and require subjects to be attentive.

- Animal sounds : Finally, we chose ten different common sounds produced by the well known animals whose pictures were selected for the task control condition. Each animal sound had a duration of 300 ms.

All auditory stimuli were normalised for sound energy by matching their root mean square (RMS) power to 24.0 dB and they were edited using the programs Audacity (2.0.3) (Free Software Foundation, Boston, USA) and Praat (Institute of Phonetic Sciences, University of Amsterdam, the Netherlands). The visual stimuli were edited on Adobe Photoshop CS5.1 (Adobe Systems Incorporated, San Jose, CA, USA).

Design and Procedure

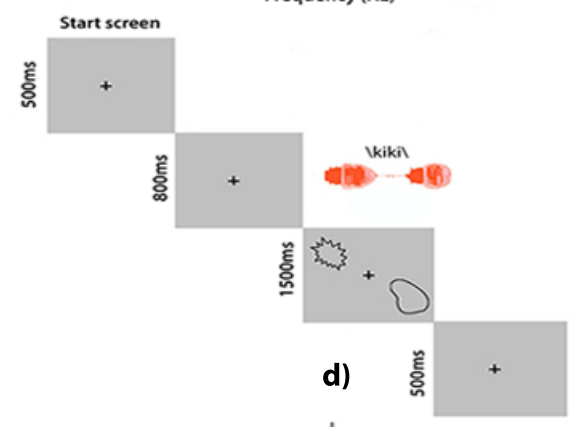
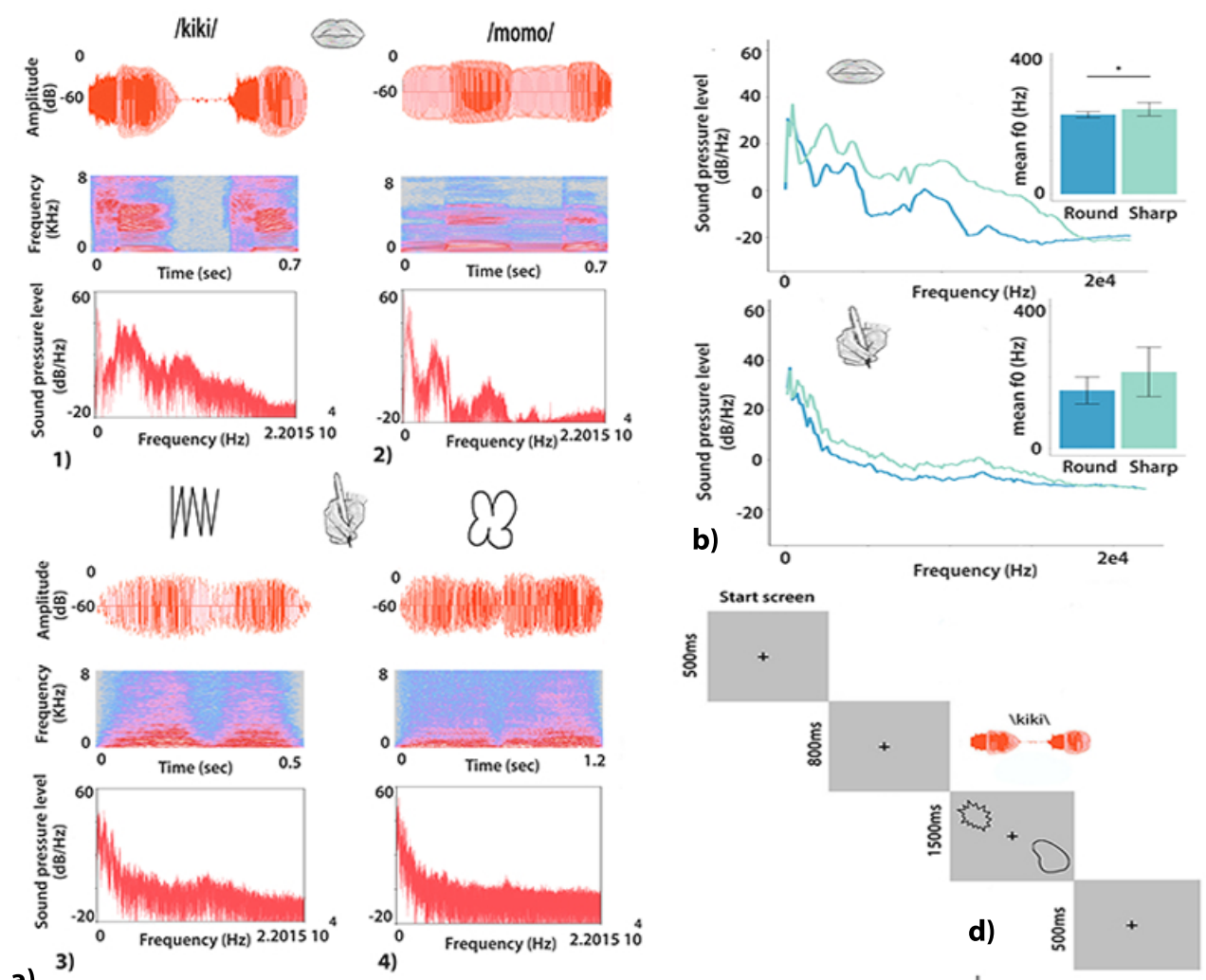
The experiment was programmed in E-Prime 2.0.8.90 (Psychology Software Tools, Inc., Pittsburg, PA, USA). During the 2AFC task, subjects are presented with a sound/pseudoword, followed by two alternatives (pictures/shapes) and they have to make a forced-choice on which picture/shape is the target stimulus that best matches to the preceding sound. Subjects performed a 2AFC task with five different conditions. In the first four conditions, we explored any congruency effects between the different sound symbolic and hand action related visual and auditory stimuli. Specifically, in the first condition (sound symbolic, SoSy) subjects had to match SS_{pwd} to SS_{sh} . In the second condition (Action) they had to match $Action_{\text{snd}}$ to $Action_{\text{sh}}$ stimuli. For the third and fourth conditions, we crossed the auditory and visual stimuli of the previous two conditions. Hence for the Crossed1 condition we used the SS_{pwd} with $Action_{\text{sh}}$ and for the Crossed2 condition the $Action_{\text{snd}}$ with the SS_{sh} . Condition five, the Animal task, was introduced for any effects (e.g., attention, perception, motor responses) induced by the 2AFC task itself that could affect the performance of the subjects generally. Finally, the last paper-and-pencil d2-attention test was introduced in order to control for variable levels of sustained attention for each subject (see Fig. 4.2 c).

In all five alternative forced choice conditions, each trial started with the presentation of a fixation cross for 500 ms followed by the presentation of an auditory stimulus ‘the

prime'. Due to the different nature of the sounds (SS_{pwd} , $Action_{\text{snd}}$, Animal sounds) presentation time of these prime stimuli were either 800 ms (SS_{pwd}) or 1700 ms ($Action_{\text{snd}}$), or for 300 ms (Animal sounds). Next, the two shapes, always one sharp and one round, appeared diagonally on the screen, one on the upper left, the other on the bottom right or in the other two corners. One of these visual stimuli was the target matching with the previous prime sound. During the fifth condition, two animal pictures were presented with only one of them matching to the preceding animal sound. The two visual stimuli stayed on screen for 1500 ms ($SS_{\text{pwd}} / Action_{\text{sh}}$). Presentation time was shortened to 1000 ms (Animal picture) so as to slightly challenge the subjects in the otherwise too easy Animal task. Responses were collected while visual stimuli were on screen. Every trial ended with the presentation of a blank slide lasting for 500 ms (see Fig. 4.2 d). All visual stimuli were presented on a grey background (RGB 192,192,192). Each condition consisted of 160 trials. Half blocks of 80 trials were separated by a pause screen. The subjects decided when to resume the next half block. Within each condition, trials were randomized; the combinations of auditory and visual stimuli were unique in each half block.

In a sound proof and dimly lit room, subjects sat in front of a 23 in. LCD monitor (screen refresh rate 75Hz; screen resolution 1280×1024). The auditory stimuli were presented via two Logitech speakers (Model NO: Z130) (Logitech Europe S.A., Lausanne, Switzerland) located at each side of the screen. Responses were recorded via two buttons on a Serial Response Box™ (SRBox, Psychology Software Tools, Inc, Pittsburg, PA, USA). Before the initiation of the experiment and at the beginning of every new condition, subjects received on the screen the following written instructions: "During the experiment, two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures that best matches the sound you just heard". No specific instructions were given to the participants regarding speed or accuracy. Button presses had to be given with the index and middle fingers of the right hand. The up/down button was used for selecting the visual stimuli appearing at the corresponding side of the screen. After completing the computer experiment, all

subjects completed in English the d2 cancellation test (Brickenkamp and Zillmer, 1998). The d2 paper-pencil test is a psychometric measure of sustained attention. During the test, takers are asked to discriminate between different visual stimuli, and cross out the target stimuli (the letter "d" with two dashes). The d2-test procedure lasted about 5 minutes. Finally, subjects performed a questionnaire in which they rated on a Likert scale the roundness and sharpness of the two maxima action sounds, the action shapes and the sound symbolic shapes (see Fig. C.3).



c) Stimulus categories and d2-Test.

	SoSy	Action	Crossed1	Crossed2	Animals	d2-Test
Auditory						
Visual						

Figure 4.2.: a) Waveforms, spectrograms and power spectral densities (PSD) of SS_{pwd} (top panels) and $Action_{\text{snd}}$ (bottom panels), 1) "kiki", a 'sharp' rated bisyllabic SS_{pwd} , 2) "momo", a 'round' rated bisyllabic SS_{pwd} , 3) a 'sharp' and 4) a 'round' sounding $Action_{\text{snd}}$. b) Average PSD for both 'sharp' and 'round' sounding SS_{pws} (top panel) and $Action_{\text{snds}}$ (bottom panel), segmented in 145 bins. Mann-Whitney-U-tests were used to calculate the difference of PSD average values between round and sharp categories. For both SS_{pwd} ($W=14919$, $p < 0.001$) and $Action_{\text{snd}}$ ($W=7526$, $p < 0.001$) there was a significant difference of PSD values between 'sharp' and 'round' sounding stimuli. Bar plots show average and standard deviations of fundamental frequencies (F0) for 'sharp' and 'round' sounding categories. Mann-Whitney-U-tests revealed a significant difference only for the SS_{pws} ($W=17$, $p < 0.01$) between 'sharp' and 'round' sounding categories and not for the $Action_{\text{snds}}$ ($W=5$, $p = 0.15$) for the F0 measure. c) The table summarizes the combination of auditory and visual stimuli for the five forced choice tasks. The sixth column depicts an example from the d2 attention task as presented in the paper-pencil version. d) Schematic representation of the experimental procedure for the SoSy condition. The procedure was the same for all the forced choice tasks with modifications on presentation times depending on the type of the stimulus.

4.3. Data analysis

All analyses were performed on the analysis tool R (version 3.4.3, R Development Core Team) (Team et al., 2013). Trials with reaction times greater than the time response window or without button-press were excluded. All variables were checked for normality using the Shapiro-Wilk normality test. To check if subjects' selection of shapes was influenced by the preceded sound, we performed Wilcoxon signed-rank tests to compare the number of congruent responses against chance level for every condition

separately after controlling for multiple testing with Bonferroni correction (adjusted threshold $p = 0.05/5 = 0.01$). Moreover, we compared the congruency performance between the four conditions with a Kruskal-Wallis test with pairwise multiple comparison adjusted using Bonferroni correction. In order to explore whether the congruency detection performance of the subjects in a given AFC condition was correlated with their congruency detection performance with the other AFC conditions and with their performance on the d2 test, we performed a number of correlations. Specifically, we calculated Spearman’s correlation coefficients to assess pairwise linear relationships for the number of congruent responses of each subject between AFC conditions, and between each AFC condition and the scores acquired from the d2 test. From the d2 test, we calculated the concentration performance (CP) score, which is the number of correctly crossed-out items minus the errors of commission. CP scores can provide an index of sustained attention and takes into account both speed and accuracy of the performance. The higher the CP score the higher the attention of the subject. A false discovery rate correction (FDR, threshold set at 0.05)(Benjamini and Hochberg, 1995) controlled for multiple comparisons using the `p.adjust` function in R. Furthermore, for comparing the size of the correlation coefficients among the sound symbolic, action and crossed conditions and between with the control AFC task, we performed 12 multiple comparisons with Steiger’s Z one-tailed tests on these coefficients. All p -values were adjusted with Bonferroni correction for multiple testing ($p = 0.05/12 = 0.004$).

In order to check, whether performance in the first four conditions was further influenced by other variables, we fitted a generalized linear mixed model (GLMM) with a binomial error structure using the package `lme4` (Bates et al., 2014). The dependent variable was congruency, that is, whether the selected shape matched the shape corresponding to the primed sound. We included $SS_{\text{pwd}}/\text{Action}_{\text{snd}}$ (‘sharp’ vs.‘round’) and trial number as fixed effects. We used a maximal random effect structure with random intercepts for subject, SS_{pwd} or $\text{Action}_{\text{snd}}$ and for the combinations of the presented shapes and random slopes for each trial nested within these random effects. We used the likelihood ratio test (LRT) to check if the predictor variables improved the fit of

the model; these were calculated by comparing the full model to a reduced model that included all terms except for the fixed effect term in question. Chi-square and p-values were computed using the function `drop1` from the R package `lme4`.

4.4. Results

Across all five conditions, a total of 5.8% of trials were excluded from the analyses because of null or long-delay responses. Shapiro’s-Wilk tests, performed on the percentage of congruent responses obtained from each subject for each of the five conditions, revealed that normality was violated for two conditions, (Action: $W=0.75$, $p < 0.001$) and for (Crossed2: $W=0.83$, $p < 0.001$) and hence non-parametric statistics were performed. In each condition, subjects showed above chance performance on congruency detection between the presented sound and the selected pictures. In particular, for SoSy, subjects performed above chance ($V=273$; $p = 0.001$) with an average 70.64% congruent responses. Similarly, above chance performance was observed for the Action condition with an equally strong congruency bias of 81.50% congruent responses ($V=244$, $p = 0.001$). Comparable results were obtained for the two crossed conditions, Crossed1 and Crossed2 with 76.59% of congruency ($V=270$, $p = 0.001$) and 80.56% ($V=266$, $p = 0.001$) congruent responses. The Animal task yielded 90.20% congruent responses ($V=276$, $p = 0.001$)(see Fig. 4.3a). In addition, the Kruskal-Wallis test showed a statistically significant difference between congruency performance levels across the first four conditions ($\chi^2(3)=8.45$, $p = 0.04$). However, none of the pairwise differences survived Bonferroni correction ($p < 0.012$, $= 0.05/4 = 0.012$).

Next, we addressed the *primary hypothesis* whether the roundness and sharpness classifications of sounds were related to each other across the SoSy and the Action conditions. Spearman rank correlations revealed a significant positive correlation between subject specific congruency percentages obtained from the SoSy and the Action conditions ($\rho = 0.50$, $p = 0.01$ before and $p = 0.03$ after FDR correction). Notably, correlations of SoSy task performance with that on the closely matched 2AFC control task failed to reach

significance (see Fig. 4.3b). One may argue that the significant correlation between Action and SoSy conditions and its absence in the comparison between SoSy and 2AFC control task may just reflect a threshold effect. To address this possibility, the Steiger’s Z test was used to assess any significant differences between correlation coefficients. Using this test, the crucial correlation of SoSy and Action condition performances (SoSy vs. Action) was significantly greater (Steiger’s $Z = 1.67$, $p = 0.04$) than that between SoSy and 2AFC control task results.

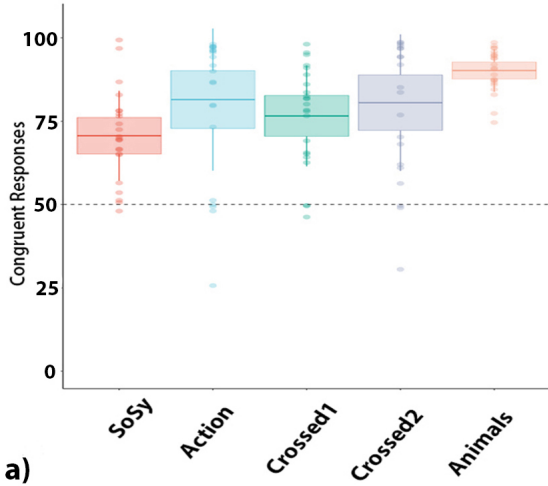
To address the secondary hypotheses, that the mapping between SoSy and Action conditions was in part due to similarities in acoustic and visual stimuli used across these tasks, we calculated all pairwise correlations between the SoSy, Action and Crossed conditions (FDR corrected). The highest positive correlations were observed between Action and Crossed2 conditions ($\rho=0.88$, $p = 0.001$), followed by SoSy and Crossed1 ($\rho=0.76$, $p = 0.001$). These condition pairs both share the same sounds: the Action and Crossed2 conditions the action sounds and the sound symbolic and Crossed1 conditions the pseudowords. Therefore, the correlations indicate that subjects generalized very well across shape types: they performed similarly on matching SS_{sh} and $Action_{sh}$ to the same sounds. This implies a degree of similarity between SS_{sh} and $Action_{sh}$, which is obvious, as the same visual elements resulting from elementary round and edgy movements were the components of these shapes. Clearly significant, although slightly less impressively than the former, were the correlations between conditions that shared the same shapes, i.e., $Action_{sh}$ or SS_{sh} . $Action_{sh}$ were similarly well matched to $Action_{snd}$ as to SS_{pwd} ($\rho=0.58$, $p = 0.01$), and the same applied for the SS_{sh} ($\rho=0.56$, $p = 0.02$). These results indicate a similarity in processing the different sound types, of actions and speech sounds, a topic to which we will return in discussion below. Moreover, a positive correlation was also observed between the two crossed conditions, Crossed1 and Crossed2 ($\rho=0.52$, $p = 0.03$)(see Fig. 4.3c).

Remarkably, there was not a single reliable correlation between the closely matched action-unrelated 2AFC task using animal pictures and sounds and any of the four experimental conditions addressing SoSy and Action related knowledge (see Fig. 4.3c).

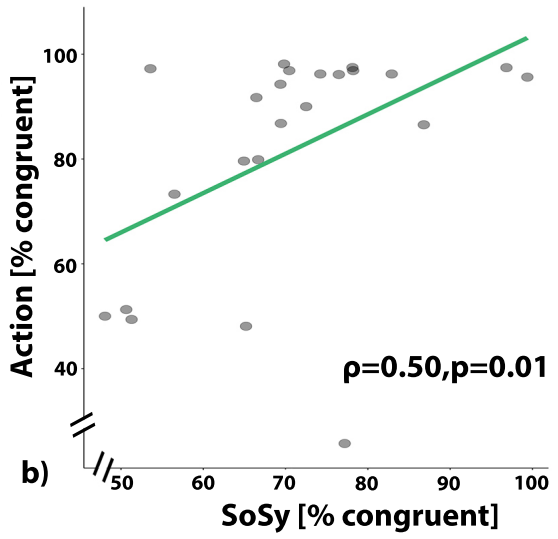
Likewise, the secondary control task, d2 test performance using the CP score index, failed to yield any significant correlations with any of the sound symbolic or action related conditions. Also, performance on the two control tasks was uncorrelated. The absence of correlations with any of the two control tasks shows that the performance variability of our subjects in sound symbolic and action related conditions was not related to attention or to the cognitive and motor demands of the forced choice task.

To address possible threshold effects related to the secondary hypothesis, the Steiger Z-test was used once again, now to more systematically compare all possible pairings of correlation coefficients across SoSy-Action domains on the one hand – the ‘within-domain correlations’ – and correlations between these and the task-control condition on the other – ‘between-domain correlations’. The 12 tests performed between ‘within’ and ‘between-domain’ correlations revealed 8 significantly different correlation coefficients ($p < 0.05$), and even after most conservative Bonferroni correction (corrected critical $p = 0.05/12 = 0.0042$), five of these remained significant (for details, see Table C.1). This is evidence for the specificity of correlations across action- and sound-symbolic domains.

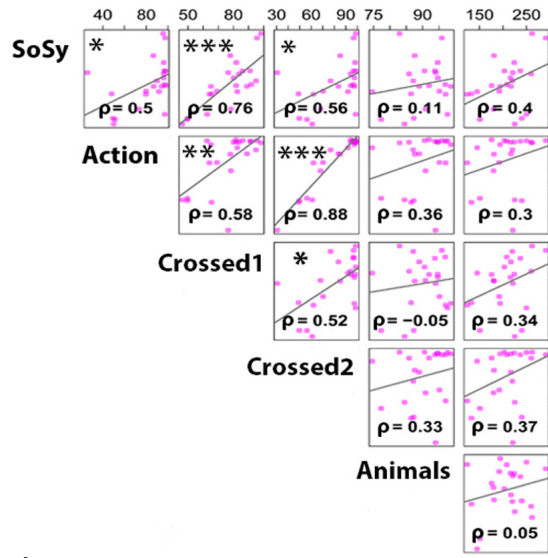
The predictor variable of SS_{pwd} type significantly improved the model for SoSy condition ($\chi^2(1)=12.72$, $p = 0.001$) with subjects having more congruent responses for ‘round’ sounding SS_{pwd} than ‘sharp’ ones, a finding previously reported by Margiotoudi et al. (2019), which may indicate a ‘roundness bias’ in the matching choices of pseudowords in sound-symbolic experimental context. This effect was, however, not seen in other conditions. The factor $SS_{\text{pwd}}/Action_{\text{snd}}$ type did not improve any of the other models with Action not reaching significance ($\chi^2(1)=3.26$, $p = 0.07$), not either in the conditions Crossed1 ($\chi^2(1)=1.7$, $p = 0.18$), or Crossed2 ($\chi^2(1)=0.96$, $p = 0.32$). Therefore, any roundness bias was not present in the crossed conditions sharing either SS_{pwd} or SS_{sh} stimuli with the SoSy condition. As a result, the roundness bias specifically observed in the SoSy condition cannot be driven by the pseudoword or shape stimuli shared between SoSy and Crossed conditions.



a)



b)



c)

d2-Test

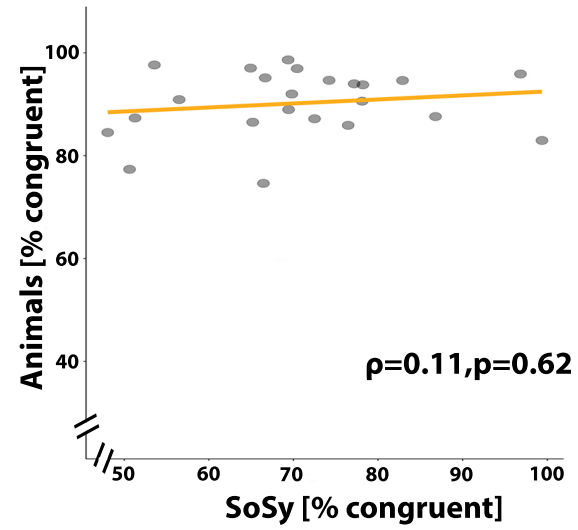


Figure 4.3.: a) Percentages of congruent responses for the two-alternative forced choice conditions, the SoSy (red), the Action (blue), the Crossed1 (green) and Crossed2 (purple) and the Animals tasks (orange). For the first four conditions, congruency is quantified as the proportion of times each individual matched a ‘sharp’ sounding $SS_{\text{pwd}}/Action_{\text{snd}}$ to a sharp shape or a ‘round’ sounding $SS_{\text{pwd}}/Action_{\text{snd}}$ to a round shape. For the Animal task, congruency means correct matching of sound and selected animal picture. Light colored circles show the percentage of congruent responses for each individual. Boxplots show standard deviations, lines show means and the whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance. b) Bivariate scatterplots with regression lines and correlation coefficients (ρ values) of Spearman correlations between SoSy and Actions (green), and between SoSy and Animal task (yellow). c) Bivariate scatterplots with regression lines and correlation coefficients (ρ values) of Spearman correlations calculated across congruency scores of subjects obtained for all possible condition pairs, including the five alternative forced choice conditions and the concentration performance (CP) scores of the d2-test. Significant correlations after FDR correction (threshold set at: 0.05) are marked with asterisks ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

4.5. Discussion

In the present study, we used several two-alternative forced choice and control tasks to investigate the role of action knowledge in sound symbolism, i.e., the human-specific ability to detect abstract iconic correspondences. We replicated the well-known classic "maluma-takete" effect in the sound symbolic or SoSy condition and found similar and statistically even more impressive result for an Action condition, where subjects had to match abstract shapes drawn with a pen and the sounds produced by drawing them. Notably, by crossing both conditions and thus pairing action shapes with pseudowords

(Crossed 1) and abstract shapes with action sounds (Crossed2), we also found that our experimental subjects consistently judged sound symbolic correspondences, across SoSy and Action stimuli, thus classifying some shapes and sounds coherently as either round and others as sharp.

However, subjects performed differently well on such classification and we therefore asked, whether their differing levels of ability to interlink meaningless speech with abstract symbolic shapes might be systematically related to their performance on associating the shape of hand movements with the sounds produced when performing such movements. Surprisingly, when correlating the subjects' performance on the sound symbolic and the action task, there was a significant correlation, which even exceeded that found between the 2AFC control task unrelated to sound symbolic or action knowledge. Furthermore, when investigating all four sound symbolic, action and crossed tasks, we consistently found significant correlations across these. No significant correlations were found between sound symbolic or action related conditions and the main control tasks examining general performance on the 2AFC control task and sustained attention abilities.

These results, demonstrate that human subjects' sound symbolic ability to associate meaningless speech with abstract shapes is intrinsically related to their knowledge about the sounds of bodily actions performed with the hand and the shapes of the trajectories of such movements. We submit that this knowledge about sound symbolic relationship in our experimental subjects is best explained by associative learning between manual actions and the observed shapes and sounds they produce, along with visual similarities between action and sound symbolic shapes as well as by acoustic similarities between action sounds and speech.

One may argue that articulatory sounds and their related articulatory trajectories may provide an alternative explanation for the sound-symbolic capacity of humans, as previously stated by Ramachandran and Hubbard (2001) (henceforth R&H). These authors stated that “[...] the sharp changes in visual direction of the lines in the [takete] figure

mimics the sharp phonemic inflections of the sound kiki, as well as the sharp inflection of the tongue on the palate.” (ref. Ramachandran and Hubbard (2001), p.19). Therefore, the postulate is about (i) a correspondence between ‘sharp’ visual line arrangements and ‘sharp’ sounds and about (ii) a correspondence between ‘sharp’ visual arrangement and ‘sharp’ inflections of the tongue. Statement (i) appears to us rather metaphorical. The word ‘sharp’ means different things in the context of visual shapes and sounds. Any ‘similarity’ needs explanation, but cannot be taken for granted and used to provide an explanation. The crucial question is why we perceive ‘sharp’ shapes and sounds as somewhat similar, and this question remains unanswered by R&H’s statement. Whereas their first statement does not provide an explanation, R&H’s postulate (ii) comes with an empirical implication: that visual shapes of the abstract figures must in some way or another, resemble the “inflections of the tongue on the palate”. Unfortunately, the authors do not provide empirical or experimental evidence. Meanwhile, a body of data is available addressing this issue. So is there in fact resemblance between sharp and edgy figures and sharp tongue or articulator inflections on one side and rounding figures and round and smooth articulator movements?

As mentioned in the Introduction above, knowledge about the trajectories of our articulations is implicit and procedural so that one may dispute conscious access to it. As most articulators and their trajectories are not visible to the speakers or interlocutors, it may therefore be asked how any knowledge about these trajectories could come in into play in the cognitive task of sound symbolic matching. Decades of phonological and phonetic research were necessary to uncover these articulatory trajectories, so that it appears as a little optimistic to assume that the relevant knowledge is freely available as a basis for explicit sound-symbolic decisions. If we focus on articulators that are visible, as for example the lips, only a limited fraction of relevant features can be covered. But even worse: as we will elaborate below, there seems to be a lack of evidence for resemblances between abstract shapes and the shapes of articulators or articulatory trajectories while uttering phonemes that contribute to the perception of a pseudoword as either ‘sharp’ or ‘round’-sounding.

The actual movement trajectories revealed by articulatory phonetic research do not seem to exhibit edges, but, instead, appear as similarly smooth and round for round-sounding and sharp-sounding phonemes. In the case of consonants, even the most ‘spiky’ examples, such as /p/, are produced by similarly smooth lip movements as the ‘round’ sounding /b/ (see Fig 4.1d). Rather than being based on sharp and round articulatory movements, their acoustic differences relate rather to the precise timing of articulatory movements or the level of oral air pressure released Ladefoged and Disner (2012); Löfqvist and Gracco (1997). Turning to vowels, one may want to point to examples such as /i/ and /u/ - where one phoneme is ‘round’ from a sound symbolic perspective and from a phonetic perspective too (the /u/) – as it requires lip rounding, whereas the other one is sound-symbolically ‘sharp’ and not-rounded phonetically (the /i/). However, in spite of the existence of such matches, mismatching counterexamples are easy to find. Items that are uniformly classified as ‘rounded’ from a phonetic perspective, such as /y/ and /u/ –since both require lip rounding–end up at different ends of the sound-symbolic roundness-sharpness continuum (see for /y/ Ahlner and Zlatev (2010) and for /u/D’Onofrio (2014)). Sharp-sounding but phonetically ‘rounded’ /y/ violates the correspondence as do round-sounding but phonetically not-rounded /α/ and /a/ (D’Onofrio, 2014; Chow and Ciaramitaro, 2019). Therefore, lip rounding as a phonetic feature is not a reliable indicator of sound symbolic categorization.

The unreliable status of articulatory movements as indicators of sound symbolic properties is further confirmed when observing the trajectories of articulators hidden in the mouth. The tongue shape and trajectory while articulating the vowel /i/, a high front vowel producing a strong bias towards ‘sharp’ sound-symbolic judgements, does not show features of a spiky figure, nor would the ‘round’ sounding /u/ and /o/ exhibit any smoother tongue trajectories (see Fig 4.1c). This phoneme, /i/ is produced with the tongue close to the roof of the palate, thus creating a large cavity at the back of the mouth, which does not mirror a sharp structure nor is edgier compared to the tongue shape characteristic of /u/, which has the back of the tongue close to the palate (see Fig. 4.1b). Moreover, the resulting shapes of the kinematic trajectories in Figure 4.1c,

do not resemble the sharp and round shapes of the classic "maluma-takete" shapes depicted in Figure 4.1a. Similarly for the sharp vs. round sounding syllables /pi/ and /ba/, we explained above that the movements of the articulators do not have corresponding 'round' vs. 'sharp' features (see Fig. 4.1d). Studies, when mapped tongue movements online, for example with articulatory tractography, found comparable trajectories, for example of the back of the tongue, for different vowels Schönle et al. (1987). Therefore, it appears that the envisaged 'similarity' of articulatory movements to sharp and round shapes cannot be used as an explanatory basis of sound symbolism.

Although the similarity between articulatory movements and round vs. sharp shapes cannot account for the general phenomenon of sound symbolism, we do not wish to exclude that an acoustic-articulatory speech component might contribute in some way to such an explanation. In contrast, however, the shared roundness and sharpness features of overt hand movements shared across the visual shapes of their trajectories and the sounds of these actions are well supported by our current data and generally applicable to various speech sounds. Therefore, they offer a perspective on explaining sound symbolism.

Given that correlations across experimental subjects' performance were observed, one may argue that any significant effects may be due to general between-subject differences, such as, for example, differences in arousal, sustained or visual attention, or swiftness and skill in solving computerized tasks requiring button presses. As one possibility, it could have been the relatively greater level of attention of individual subjects to sounds and figures along with their acoustic and visual details that co-determined comparatively better performance on both sound-symbolic and action alternative forced choice tasks. To explore these possibilities, two control tasks were administered. The first task was designed to closely match the 2AFC task frequently used in sound symbolic experiments, but for the control task, no sound symbolic or action related information was involved. Subjects had to match animal pictures to sounds produced by animals, a task not drawing upon information about human action. Note that this task did not only control for possible differences in attention levels but likewise for putative variabil-

ity in perceptual or motor skills (e.g., slow vs. fast responders). We consider this AFC task the main control condition, as it most closely controlled for various features of the critical tasks. Furthermore, a second control task was administered, the d2-test, which provides an estimate of levels of sustained attention. Interestingly, whereas all correlations of performance across the four sound symbolic/action related conditions achieved significance, (at least at a level of significance uncorrected for multiple comparisons), all correlations between one of the latter and a control task were insignificant. Note that the large number of tests made it necessary to control for multiple comparisons, and, as mentioned in results, even after most rigorous correction a relevant number of tests were still significant, thus providing strong evidence for the proposed action-based explanation. However, the primary hypothesis of our current study addressed one and only one correlation, that between SoSy and Action matching tasks (and thus did not call for multiple comparison correction). As this correlation was significant and, crucially, proved significantly stronger across subjects than that between the sound symbolic and the main (2AFC) control task, we can conclude that the primary hypothesis, that sound symbolic and sensorimotor action mappings are intrinsically related, receives strong support. Our results also show that the sosy-action correlation we observed across individuals is not explained by perceptual, task-performance-related or general cognitive differences between experimental subjects.

The significant correlations in the crossed conditions together with those between SoSy and Action condition indicate that some acoustic features are shared between the ‘round’ sounding SS_{pwd} and $Action_{\text{snd}}$ produced in creating roundish hand movements and lines tracing them and likewise for the ‘sharp’ category. As the correlation between crossed and SoSy/Action conditions that shared their acoustic stimuli –either SS_{pwd} or the $Action_{\text{snd}}$ drawings –led to the most impressive results, with ρ values ranging around 0.8, it appears that these visual stimuli differing between these condition pairs resembled each other. This was doubtlessly the case, because the two visual shape categories, that is sound symbolic and elementary action shapes, shared edges/spikes or smooth curves. Based on these visual similarities, performance correlations between conditions sharing

acoustic stimuli can easily be explained.

Likewise, for the conditions sharing the visual shapes, there were significant results. This indicates that, also across the acoustic stimuli, the SS_{pwd} and $Action_{\text{snd}}$, there was a degree of similarity. Looking at individual stimuli, this hypothesis can be supported. Figure 4.2a & 4.2b show acoustic wave forms, spectrograms and frequency composition of sound stimuli (from the SS_{pwd} and $Action_{\text{snd}}$ categories) commonly judged as ‘sharp’ or ‘round’. It can be seen that in both, the ‘sharp’ $Action_{\text{snd}}$ and SS_{pwd} have brief breaks or sudden pronounced sound energy drops between the two maxima of the sound, whereas, the ‘round’ sounding stimuli lack such an abrupt break or substantial dip. Also, the ‘sharp’ items typically exhibit relatively more power in the high frequency range, which is either absent or much reduced for the ‘round’ items; instead the latter include relatively more energy at the lower frequencies (see average power spectra in the bottom diagrams in Fig. 4.2b). These observations were supported by statistical analyses. We found significantly different overall spectral power for both ‘sharp’ $Action_{\text{snd}}$ ($W=7526$, $p < 0.001$) and SS_{pwd} ($W=14919$, $p < 0.001$) as compared with their respective ‘round’ categories. In addition, the first peak of the Fourier spectrum was found at significantly lower frequency for ‘round’ stimuli than for ‘sharp’ ones for the SS_{pwd} (round: 236.9 vs. sharp: 252.8 Hz $p < 0.01$). Similar patterns were revealed for the $Action_{\text{snd}}$ (round: 162.7 vs. sharp: 214.7 Hz $p > 0.05$), although the differences did not reach significance in this case, maybe due to the limited number of actions (five per category).

In summary, our results revealed a reliable correlation between our subjects’ performance on the classic task of sound symbolism and an action condition. This finding is best explained by the similarities between stimulus categories, in particular between sound-symbolic shapes and the drawn shapes on the one hand and between the pseudowords and the sounds resulting from shape production on the other. The correlation suggests that, due to these physical similarities, similar mechanisms are at work in the processing of actions and sound symbolism.

These results offer a novel explanation of sound symbolism. As the link between

abstract shapes and meaningless speech is difficult to explain, similarities between these shapes and the correlation between the trajectories and sounds of hand actions can easily be learned when observing oneself or another person drawing or otherwise producing such shapes. Hence, it is possible to explain sound symbolic knowledge as a consequence of action knowledge, i.e., the learnt correspondences between the shapes and consequent sounds of hand movement.

It is worth mentioning that previous studies have already shown that, beyond sound-shape associations, round and sharp dynamic body movements can also be associated to ‘maluma’ vs. ‘takete’ pseudowords Koppensteiner et al. (2016) as well as to certain speech sounds Shinohara et al. (2016). Shinohara et al. (2016) reported that front vowels and obstruents are more likely to be associated to sharp than round dynamic gestures and demonstrate a further fact of abstract cross-modal sound symbolism. In this study, the takete-maluma-type sound symbolism is considered just one type of sound symbolism and the movement-phonemic links represent a different one, so that all of these cross-modal links are instantiations of “a general feature of our cognition”. These findings, although providing great evidence for the link between actions and round or sharp sounding speech sounds, do not address whether action knowledge may be the basis of abstract sound symbolic knowledge. In addition, the actual sounds created by executing these body movements were not investigated. Here, in contrast to Shinohara et al. (2016), we propose that there are not different types of sound-symbolic knowledge – e.g., for static figures and for actions – but that one type (action knowledge) explains the other seemingly ‘abstract’ types by experience-based associative learning and physical similarity, rather than by pre-established abstract links.

One may object that the visual and acoustic stimuli used in this experiment were too limited to fully support such general conclusion. Other visual shapes, for example more complex ones than the elementary ones used in this study, may show other relevant features not explored here. However, we believe that these possible caveats do not generally invalidate our argument. If other, for example more complex shapes allow for additional sound-shape associations, this does not invalidate the links obvious from our

present stimuli using elementary figures. Other ways of producing sounds –for example produced by ‘drawing’ shapes with a sword in the air, or the tip of the foot in the sand –will certainly produce different sounds. Nevertheless, it seems plausible that the acoustic physical features varying between a sharp and round on-paper drawing are similar to the features emerging from the same shapes being drawn with sword or foot. In fact, we have experimented with different ways of producing action trajectories and sounds and finally selected the pen-on-paper strategy because it led to stimuli that were easy audible and easy to control for a range of acoustic properties (see Methods). Although we have not investigated this systematically, our data indicate that acoustic and visual features differences are shared across different ways of action production. Therefore, these differing features may provide the cues for visual-acoustic binding of information essential in sound symbolic knowledge.

The knowledge about an action together with its visual and auditory aspects must be stored in the cortex by a memory trace. Such traces may be local neuron circuits localized in a specific part of the brain devoted to semantics, a so-called ‘semantic hub’ (Patterson et al., 2007). However, this type of model does not explain the knowledge link between memory mechanism and the perceptual and action-related knowledge it needs to connect with (grounding). Therefore, grounded memory models propose distributed neuron circuits as the carriers of memory (Fuster, 2015). These distributed circuits interlink neurons in sensory and motor systems also relevant for perceptual and action-execution mechanisms by way of neurons in multimodal areas (Garagnani et al., 2008; Pulvermüller, 2018a; Tomasello et al., 2017). The distributed nature of these ‘action perception circuits’ makes it necessary to use cortical long-distance connections for linking together the motor, acoustic, visual and other perceptual knowledge of engrams and connect them with those parts of the distributed circuits most relevant for memory storage. One of the long-distance connections of the human brain especially important for interlinking action to visual and acoustic information is the arcuate fasciculus, AF, which connects frontal premotor and prefrontal with temporal visual, auditory and multimodal areas (de Schotten et al., 2012; Rilling et al., 2008; Rilling, 2014). If, as

our results suggest, sound symbolic knowledge is based on the co-storage of visual and acoustic information along with the motor aspects of overt bodily actions, the AF will have a main role in sound symbolic processing. From this theoretical consideration, a range of future predictions follow, including the following two: 1) the strength and development of the AF, which are known to vary across individuals (Lopez-Barroso et al., 2011; Yeatman et al., 2011), might determine or co-determine and therefore correlate with subjects' variable abilities to make sound-symbolic judgements, 2) subjects with dysfunction of the AF, due to developmental disorders (Moseley and Pulvermüller, 2018) or cortical lesions, should show no or much reduced ability to perform on sound-symbolic tasks. Hence, it will be an important task for future research to test these predictions and therefore further assess the theoretical proposal about action-perception circuits a basis of sound symbolism. A third prediction is that animals very similar to humans, but without strongly developed AF, should not show any sound-symbolic effects. The latter finding has recently been reported (Margiotoudi et al., 2019), thus providing at least some independent evidence for the proposed model.

Summary

We found that healthy human individuals perform similarly well on sound-symbolic matching of 'round' and 'sharp' pseudowords and abstract shapes as they are able to match diagrams of motor trajectories to the sounds of these same 'round' and 'sharp' actions. Likewise, the crossed matching of these two conditions worked equally well. Interestingly, there was a significant correlation between our subjects' performance on sound symbolic and action matching tasks, and this correlation exceeded the level of the relevant control tasks. In addition, similar correlations emerged across sound symbolic, action and crossed conditions, but were absent for when comparing performance on the latter and on control tasks. These results indicate common mechanisms of sound-symbolic and action matching and offer an explanation of the hitherto not well-understood iconic link between pseudowords and abstract forms. Although previous models attempted at an explanation based on speech sound production and the presumed shapes of articula-

tory gestures (Ramachandran and Hubbard, 2001), closer examination shows that this type of account is insufficient. The novel explanation of sound symbolism based on physical stimulus similarities to the sounds and shapes of bodily actions offers perspectives on modelling the relevant mechanism in a neurobiological framework. Most excitingly, this model offers a biological framework for understanding one type of semantic knowledge, which has long been proposed to lie at the heart of human's ability to acquire language and interlink abstract symbols with their abstract meanings.

In essence, the present study reports behavioral evidence for a role of action knowledge in explaining sound symbolic congruencies. Our findings are of vital importance from anthropological, linguistic and neurobiological perspectives, as they (1) offer a plausible mechanism behind sound symbolic congruencies relying on the human brain's action-perception networks and (2) show how body-environment interaction could have contributed to the generation of semantic vocal iconic signals carrying abstract meaning.

4.6. Preliminary studies

In this section are presented the methods and results of two preliminary studies that explored the mappings between ‘round’ and ‘sharp’ sounds of hand-drawings to round and sharp shapes produced by these hand actions. These studies were conducted before the final study reported above (Margiotoudi and Pulvermüller, 2020).

In Study 1, we tested with a classic 2AFC task whether there were any congruency detection effects between ‘round’/‘sharp’ action sounds and sharp/ round abstract visual shapes, similarly to those reported in sound symbolic studies. In Study 2, we further elaborated the 2AFC task on action sound-shape mappings and added one more condition, that of the classic 2AFC sound symbolic task, to be able to examine the performance of the same subjects in both tasks. Moreover, we introduced in both studies a control 2AFC condition, in order to check any attentional, motor, or perceptual biases induced by the 2AFC task that could affect the performance of the subjects in the action and sound symbolic mappings. Finally, we could examine the performance of the subjects on a 2AFC task under both explicit (Study 1) and implicit (Study 2) instructions on matching a sound to a picture/shape.

4.6.1. Study 1

Materials and Methods

Subjects

Twenty-four right-handed adults (17 females, age $M=24.20$, $SD=3.94$) participated in the study. The subjects were native speakers of different languages (7 German, 4 English, 4 Turkish, 3 Spanish, 4 Japanese, 1 Chinese, 1 Italian). All subjects had normal hearing and normal or corrected-to-normal vision. Subjects were recruited from written announcements at the Freie Universität Berlin. All methods were approved by the Ethics

Committee of the Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin, and were performed in accordance with relevant guidelines and regulations of the Declaration of Helsinki. All subjects provided written informed consent prior to their participation to the study, and received 10 euros for their participation.

Stimuli

Auditory stimuli were edited on Audacity (2.0.3) (Free Software Foundation, Boston, USA) and the visual stimuli on Adobe Photoshop CS5.1 (Adobe Systems Incorporated, San Jose, CA, USA). We included the following stimuli types:

- Sound symbolic abstract shapes: Twenty shapes were taken from the set used in (Margiotoudi et al., 2019) (see Table B.1). Each shape was filled with black color (RGB 0,0,0) presented on a grey background, 350×350 pixels in size.
- Action sounds: To produce ‘round’ or ‘sharp’ sounding action sounds, we recorded the sounds generated by gesturing round or sharp movements while holding these various objects (e.g., plastic, wooden and metal sticks, and leather or paper made bands) before the final selection. Due to pure quality of audio recordings from all the previous materials, the best recordings were achieved by drawing with a pen these sharp and round shapes (see Fig. 4.4). Action sounds were recorded in a sound proof room. We recorded the sounds with a stereo built-in X/Y microphones Zoom H4n Handy Recorder (Zoom Corporation, Tokyo, Japan). In order to select the ‘sharpest’ and ‘roundest’ action sounds we performed online ratings described in Margiotoudi and Pulvermüller (2020) (see Fig. C.2). For this experiment we used the total duration of every action sound (see Table C.2).
- Animal sounds: Ten different sounds produced by well-known animals were chosen (duration: $M=1310$, $SD=54.91$ ms). Sound pressure levels were equalised based on the mean root square amplitude.
- Animal pictures: Twenty corresponding pictures of the selected animals, two for

each animal, were selected. The animal pictures were selected under Creative Commons Attribution License. All pictures were colored and presented on a white background and 350×350 pixels in size (see Table C.3).

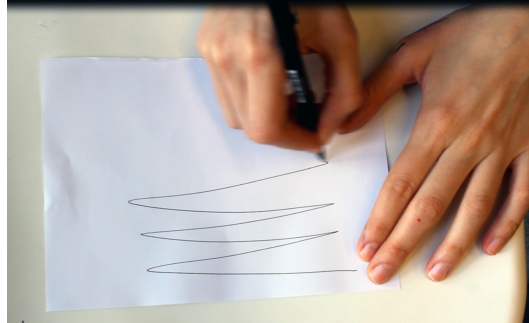


Figure 4.4.: Original hand drawing during action sound recordings.

Design and Procedure

The experiment was designed in E-Prime 2.0.8.90 (Psychology Software Tools, Inc., Pittsburg, PA, USA). Subjects performed a 2AFC task on two different conditions counterbalanced. In the action sound condition, we explored any congruency effects between the different sound symbolic abstract shapes used in the study of Margiotoudi et al. (2019), and the different ‘sharp’ or ‘round’ sounding actions sounds. We introduced the animal control condition (i.e., matching an animal sound with the correct animal picture), to monitor attention and perceptual effects induced by the 2AFC task. The design of the 2AFC task was almost identical between the two conditions, except the different time windows, due to the different auditory and visual stimuli used in the two conditions. Each trial started with the presentation of a fixation cross, lasting 500 ms and followed by the presentation of an action sound (2500ms) or an animal sound (1500ms). Next, the two target sound symbolic shapes always one sharp and one round appeared diagonally on the screen for 1500 ms. Same for the two animal pictures, which remained for the screen also for 1500ms. Every trial ended with the presentation of a blank slide lasting 500 ms. All slides were presented on a grey background (RGB 192,192,192) (see Table C.4a & b). Each condition consisted of 200 trials. Half blocks

of 100 trials were separated by a pause screen. The subjects decided when to resume the next half block. Within each condition, trials were randomized; the combinations of auditory and visual stimuli were unique in each half block. At the beginning of the animal task, we introduced five testing trials, to familiarize the subjects with the task.

The testing room, equipment and facilities for the present experiment were identical to the study of Margiotoudi and Pulvermüller (2020). Before the initiation of the experiment, and at the beginning of every new condition, subjects received on the screen the following written instructions: "During the experiment, two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures that best matches the sound you just hear". No specific instructions were given to the subjects regarding speed or accuracy. Subjects rated at the end of the study, on a Likert scale, the roundness and sharpness of the action sound recordings, and the sound symbolic shapes (see Fig. C.5).

Data analysis & Results

We excluded from the analysis a total of 2.6% of responses because no response was given or responses exceeded the 1500 ms time window. Before conducting any inferential statistics, normality of the data was checked with a Shapiro-Wilk test. For both the action ($W = 0.86$, $p = 0.005$) and the animal control condition ($W = 0.57$, $p < 0.001$), normality was violated. Hence we performed non-parametric statistics. In order to compare the congruency performance in both tasks, we performed a Wilcoxon signed-rank test. There was a significant difference in the percentage of congruent responses between the two conditions ($W = 15$, $p < 0.001$). Congruency was above chance also both for the action sound ($V = 300$, $p < 0.001$) with 89.51%, and the animal conditions ($V = 300$, $p < 0.001$), with the later reaching 99.32% congruency detection, most probably because the task was trivial for the subjects (see Fig. 4.5a).

We further compared the congruency performance of the subjects for the two action

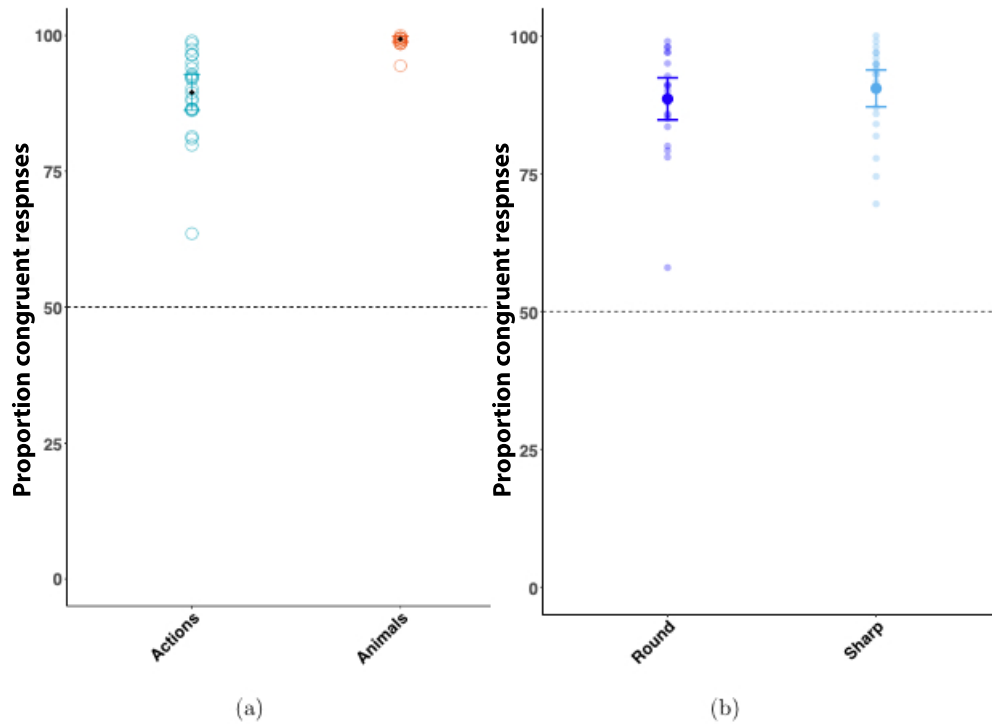


Figure 4.5.: a) Percentage of congruent responses for the two conditions. Congruency is quantified as the proportion of times each individual matched the congruent picture to the preceding sound. For the action sound condition, congruency is quantified as the proportion of times each individual matched a ‘sharp’ sounding action to a sharp shape or a ‘round’ sounding action to a round shape (blue). For the animal control condition, congruency is the matching between the sound and the animal picture selected (orange). The black diamonds show the means and the whiskers show 95% confidence intervals (CIs) b) Percentage of congruent responses for two action sound categories. Dark blue dots show the percentage correct for ‘round’ sounding action sounds and light blue circles for ‘sharp’ sounding ones. The light colored circles in both categories depict individual performance. The whiskers show 95% confidence intervals (CIs). In both graphs the dashed line at 50% shows chance-level performance.

sound categories, in order to check if the specific action sound category preceding the two visual shapes affected their performance. There was no significant difference in the congruency percentage between the action sounds for the two categories, as revealed by the Wilcoxon signed-rank test ($W=332, p > 0.05$)(see Fig. 4.5b).

Summary Study 1

The findings of Study 1 revealed that mappings between ‘sharp’/ ‘round’ action sounds and abstract visual round/harp shapes can be easily detected or inferred by healthy human subjects in a 2AFC task under explicit instructions. Moreover, there was no significant difference on congruency detection between ‘sharp’ and ‘round’ action sounds, and subjects performed equally well in both categories. In parallel, the performance on the animal control condition reached an average of 99.32% , as the task appeared to be not much demanding to the subjects.

Given that subjects reached a very high congruency performance in both conditions, in preliminary Study 2 we modified both the auditory and visual stimuli in both conditions in order to make the task more difficult for the subjects.

4.6.2. Study 2

In Study 2, we added some modifications in the two conditions (action sound & animal control condition) in order to avoid performances close to 100% of accuracy. Specifically, we decreased the total duration of the action and the animal sounds. We also modified the animal pictures in order to make them less recognizable by the subjects.

Furthermore, we introduced the classic sound symbolism task described in the study of (Margiotoudi et al., 2019), in order to compare the performance of the subjects in matching action sounds to abstract visual shapes and pseudowords to abstract shapes. If action knowledge and knowledge of the audiovisual by-products of these actions can be related to human sound symbolic ability, then we expected that “good” mappers and

subjects who can detect congruencies between ‘sharp’ and ‘round’ action sounds and abstract shapes will also detect congruencies between ‘sharp’ and ‘round’ pseudowords and the same shapes.

Finally, in order to check whether task instructions could affect the performance of the subjects in all three conditions (action, sound symbolism and animal control condition), no explicit instructions were given on matching a given sound to the shape/picture that best fits to the sound.¹ Here we expected an effect of implicit instructions on the performance of the subjects across the three conditions, with lower congruency detection performance in contrast to the previous results obtained from Study 1, based on explicit instructions.

Materials and Methods

Subjects

Thirty-three subjects participated in the study (18 females, age $M=25.24$, $SD=3.53$). The subjects were native speakers of different languages (11 German, 3 English, 3 Spanish, 2 French, 2 Greek, 2 Malayalam, 2 Mandarin, 1 Romanian, 1 Czech, 1 Polish, 1 Bulgarian, 1 Italian). Three of the subjects were bilingual, one in German-Spanish, the second in Romanian-Hungarian and the third one in English-Spanish. Two of the subjects were excluded from the analysis, due to technical problems during the experiment. Recruitment and ethic approvals were identical to preliminary Study 1.

¹The present results on the sound symbolic condition are reported in Experiment 3 (Margiotoudi et al., 2019), where we compared human performance on sound symbolism after explicit versus implicit instructions.

Stimuli

Materials and methods were identical to Study 1, with modifications in the auditory and visual stimuli. We included the following stimuli types:

- Sound symbolic shapes and sounds: Same shapes and sounds described in the study of (Margiotoudi et al., 2019).
- Action sounds : Here we used the same ten action sounds used in preliminary Study 1. However, we modified the duration of the sounds in order make the task a bit more difficult, and more similar to the duration of the sound symbolic pseudowords. The new duration of the sounds was limited to 700 ms.
- Animal sounds: The animal sounds were the same as the ones used in the preliminary Study 1. Here again, we shorten the duration to 300 ms, in order to make the task more demanding.
- Animal pictures: Finally, the same twenty animal pictures, as in Study 1, were used with few modifications. This time, the animal pictures were black and white, blurred, and were presented on a grey and 350×350 pixels in size (see Table C.4).

Design and Procedure

A 2AFC task was conducted under three different conditions. The order of the conditions was always the same. The first condition was the classic sound symbolic experiment of Margiotoudi et al. (2019). The second and third conditions were similar to Study 1 with few modifications in the auditory and visual stimuli. The presentation times of the slides for each condition differed between the animal control condition and the other two conditions (sound symbolism & action sounds), due to differences in the duration of the auditory stimuli (see Fig. C.6).

Each condition consisted of 160 trials. A pause screen separated blocks of 80 trials. The subjects decided when to resume the next half block. Within each condition, trials

were randomized; the combinations of auditory and visual stimuli were unique in each block. Before the initiation of the experiment and at the beginning of every new condition, subjects received on the screen the following written instructions: "During the experiment, two pictures will appear, one low and one high on your screen, presented after a sound. Please choose one of the two pictures". The instructions lacked information on explicitly matching sounds to shape/picture that best fitted to the sound. The present modification of the instructions was introduced for exploring performance on both the sound symbolic and action condition. Finally, subjects rated at the end of the study on a Likert scale the roundness and sharpness of the new action sounds of 700 ms, as well as the sound symbolic shapes (see Fig. C.7)

Data analysis & Results

Two subjects were excluded from the analysis because they responded only 50% of the time in one of the three conditions. Twenty-nine subjects were included in the final analysis. From all three conditions, 4.37% of responses were excluded from the analysis since no response was given or responses exceeded the response time windows. For sound symbolism, the average congruency detection performance of the subjects reached 59.84%, for the action sounds 75.20%, and for the animal control condition 84.62%. We checked for normal distribution of the data with Shapiro-Wilk tests, across all three conditions. Normality was violated only for the sound symbolic condition ($W=0.94$, $p > 0.05$) but not for both the action ($W= 0.90$, $p = 0.01$), and the animal control conditions ($W=0.88$, $p < 0.01$). For that reason, we performed non-parametric statistics. We tested the performance of the subjects against chance separately for each condition with a Wilcoxon signed-rank test. Subjects performed above chance for all three conditions (sound symbolism: $V=318$, $p < 0.001$, action sound condition : $V=398$, $p < 0.001$, animal control: $V=435$, $p < 0.001$) (see Fig. 4.7a). Moreover, with a Kruskal-Wallis rank sum test, we compared the performance across the three conditions. The analysis showed a significant effect of condition in the performance of the subjects

($\chi^2(2)=30.39$, $p = 0.001$).

In addition, we checked separately for each action sound and pseudoword category, the congruency detection with a Wilcoxon signed-rank test. Congruency detection for the ‘sharp’ action sounds was significantly above chance, with subjects selecting a sharp abstract shape after a ‘sharp’ action sound ($V=422$, $p < 0.001$). Above chance congruency performance was also observed for the ‘round’ action sounds ($V=347$, $p = 0.001$). For the sound symbolic pseudowords, the performance was above chance only for the ‘round’ sounding pseudowords ($V=390$, $p < 0.001$) and not for the sharp ones ($V=175$, $p = 0.36$) (see Fig. 4.7 b).²

Finally, pairwise Spearman’s correlations were conducted on the congruency performance between the three conditions. Significant positive correlations were observed between the action sound and the sound symbolic conditions ($\rho = 0.42$, $p < 0.05$) (see Fig. 4.8a) between the action sound and the animal conditions ($\rho = 0.41$, $p < 0.05$), but not between the animal control condition and the sound symbolic one ($\rho = 0.25$, $p = 0.19$) (see Fig. 4.8b).

Discussion

For all three tasks, we found an above chance congruency performance on matching a pseudoword, action, or animal sound to the corresponding shape or picture. First, we replicated the classic sound symbolic congruency detection in humans under explicit instructions. Furthermore, the higher congruency detection rates for ‘round’ sounding pseudowords is also in agreement with the findings of the study (Margiotoudi et al., 2019). In contrast, no such effect was observed for the action sound experiment, where both ‘round’ and ‘sharp’ sounding actions exceeded chance level.

Regarding the action sound and animal control conditions, congruency exceeded sig-

²The results are similar to the ones observed in Experiment 3 of Margiotoudi et al. (2019), note however that here we included 29 subjects in the analysis and not 31 like in Experiment 3, because two of our subjects responded in less than 50% of the total trials in the action and animal condition.

nificantly chance levels, but this time, it did not reach an extreme high performance close to 90-100% (see Fig. 4.9). A decrease in the congruency detection could be explained first by the modification of the visual and auditory stimuli used in both conditions, and by the shorter response windows introduced for the animal control condition. Also, the implicit instructions could have possibly affected subjects' performance in action sound and in the animal control task, as they did on the sound symbolic mappings.

The most striking finding of the analysis was the significant positive correlation between performances on the sound symbolic and action sound conditions. Subjects who performed well on the sound symbolic condition performed equally well on the action sound condition and vice versa for subjects who did not perform well. This correlation reveals that, possibly, the two mappings (action and sound symbolic) share the same mechanism and that 'round' and 'sharp' action sounds share similar physical properties to 'round' and 'sharp' pseudowords, as both were mapped to the same sharp and round abstract visual shapes. In the study of Margiotoudi and Pulvermüller (2020), this finding was explored further and discussed in more detail.

The second positive significant correlation reported was between the action sounds (75.20%) and animal control conditions (84.62%). This correlation could be attributed to the high congruency performance of the subjects in these two conditions; as the mapping of auditory to visual features was still an easy task for the subjects like in Study 1.

Altogether, the present study replicated, on the one hand, sound symbolic and action sound congruency effects in a group of healthy subjects under implicit instructions, and on the other hand revealed a positive correlation in the performance of the subjects in mapping action sound and pseudowords to the same sharp and round abstract visual shapes. Given that the aim of the present chapter was to explore the mechanism behind sound symbolic congruencies and how they could be related to action sound-shape congruencies and action sound/shape physical properties, we improved and elaborated further our experimental design and added a final control task to evaluate the sustained

attention levels of the subjects (Margiotoudi and Pulvermüller, 2020).

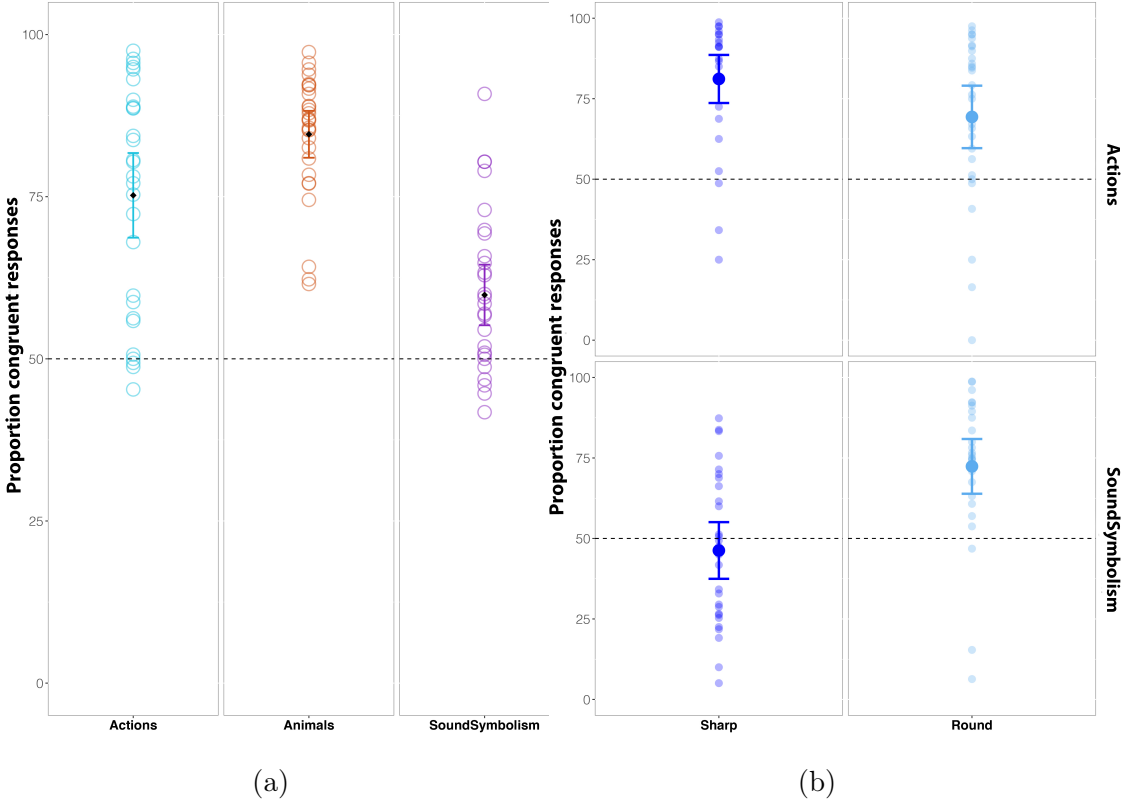


Figure 4.7.: a) Percentage of congruent responses for the three conditions. Congruency is quantified as the proportion of times each individual matched the congruent picture to the preceding sound. For the action sound condition, congruency is quantified as the proportion of times each individual matched a ‘sharp’ sounding action to a sharp shape or a ‘round’ sounding action to a round shape (blue). For the animal control condition, congruency is defined as the matching between the sound and the animal picture selected (orange). Finally, for the sound symbolic condition congruency is quantified as matching a ‘sharp’ sounding pseudoword to a sharp shape or a ‘round’ sounding pseudoword to a round shape (purple). Colored circles show the percentage of congruent responses for each individual. The black diamonds show means and the whiskers show 95% confidence intervals (CIs). b) Percentage of congruent responses for two action sound & pseudoword categories. Dark blue dots show percentage correct for ‘round’ sounding action sounds and pseudowords and light blue circles for ‘sharp’ sounding ones. The light colored circles in both categories depict individuals’ performance for action sound/pseudoword category. The whiskers show 95% confidence intervals (CIs). In both graphs the dashed line at 50% shows chance-level performance.

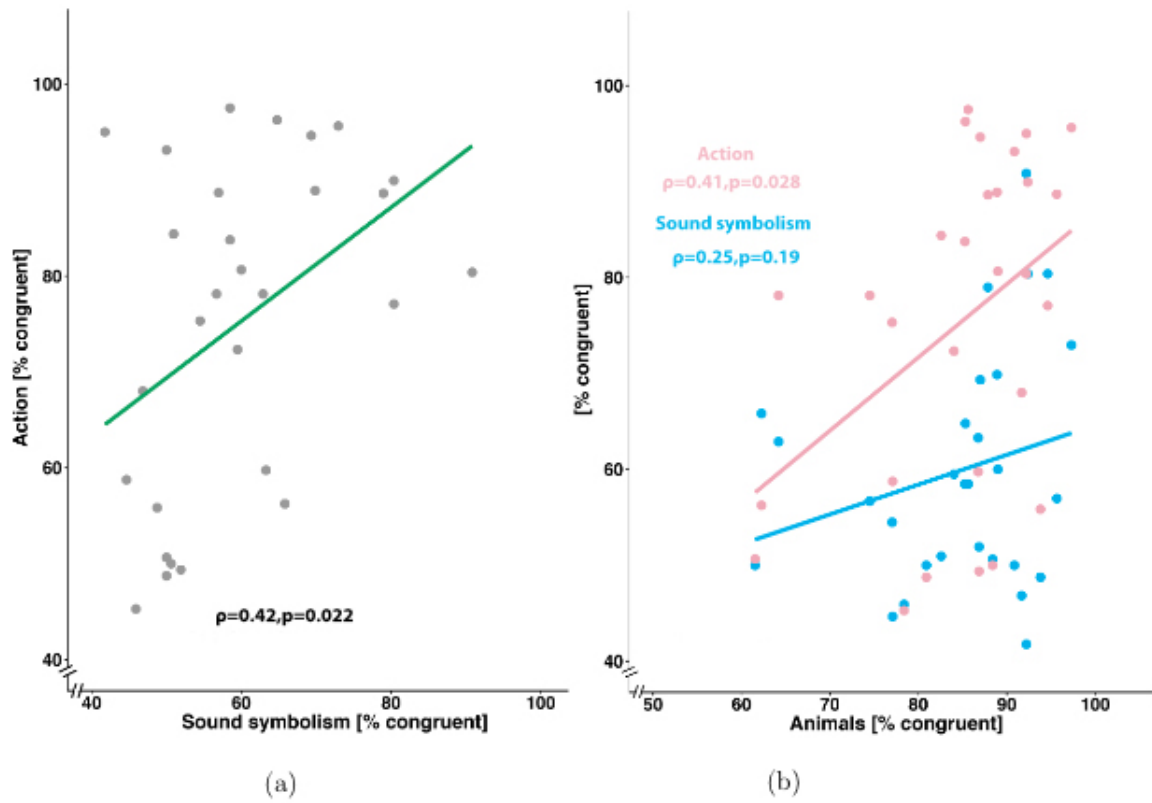


Figure 4.8.: Bivariate scatterplots with regression lines and correlation coefficients (ρ values) of Spearman correlations (a) between sound symbolic and action condition (b) between animal control condition and sound symbolism (blue) and animal control condition and action condition (pink).

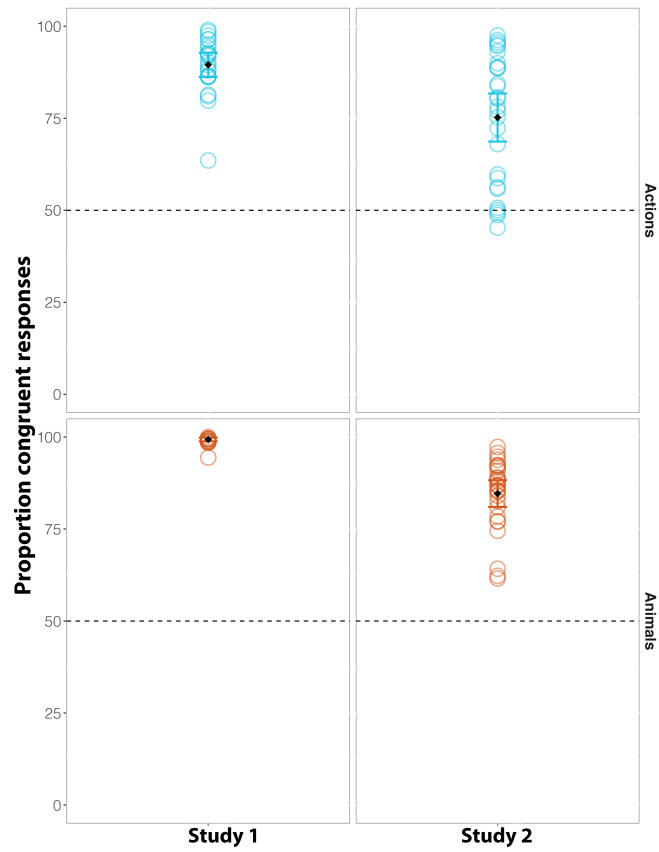


Figure 4.9.: Percentage of congruent responses for the action sound and the animal control conditions in Study 1 & Study 2. Colored circles show the percentage of congruent responses for each individual for the action sound (blue) and the animal control condition (orange). The black diamonds show the means and the whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance.

5. General Discussion

5.1. Summary of findings

5.1.1. Chapter 2

In order to understand the mechanism behind sound symbolic mappings, it is important to first validate this effect. Moreover, sound symbolism is a phenomenon similar to immanent mappings between features from different modalities, known as crossmodal correspondences. Often, there is an overlap of modality-specific features among these mappings. For instance, a round contour can be mapped both to a ‘round’ sounding pseudoword and to a low-pitched sound. However, previous studies have focused on the individual effects of these mappings and not on their interactions. Chapter 2 tests sound symbolic effects with a forced choice task and the interaction of this mapping with two audiovisual correspondences (i.e., pitch-shape, and pitch-spatial position).

The findings of Chapter 2 reveal significant above chance congruency detection only for sound symbolism and not for the other two mappings. Congruency across mappings did not improve the performance of the subjects. The rich information available from the phonemic properties of the pseudowords determined the mapping strategy of the subjects, overshadowing the low-level audiovisual properties of pitch and spatial location. The present findings validate the presence of sound symbolism when tested with a 2AFC task and show that low-level audiovisual crossmodal mappings are not detected when tested with a 2AFC task together with sound symbolism. These results stress out the

importance of studying further the interactions of these mappings to better understand their shared properties and mechanisms.

5.1.2. Chapter 3

Theoretical views on sound symbolism have highlighted its distinct role in language evolution and in the emergence of protolanguages. However, no previous study has explored the phylogenetic origin of this ability. In Chapter 3, with three 2AFC tasks, sound symbolic correspondences were tested in humans and great apes.

The findings of this chapter indicate (1) that sound symbolic ability is specific to humans, and (2) that this ability is present in humans when they are instructed both implicitly and explicitly to detect sound symbolic associations. The present findings are of great importance for the origins and mechanism of sound symbolic ability in humans, and suggest that this ability could be related to the distinct neuronal connectivity of the brain's language network in humans. Specifically, stronger left-lateralized long-distance cortico-cortical connections between inferior-frontal and posterior-temporal areas in the human brain, could support the learning of associations between abstract visual shapes and phonological units. Most importantly, these same neuroanatomical connections link motor and sensory cortices and carry visual, auditory, and perception information. The strong human-specific connectivity of this neuroanatomical infrastructure could support the model of action knowledge and knowledge of the audiovisual by-products of actions as a plausible explanation for sound symbolism, as this network would carry the perceptual and motor information of actions. Finally, as this network is stronger and more developed in human than in non-human primates, it can also explain the human specificity of this ability.

5.1.3. Chapter 4

Numerous studies have reported the effects of sound symbolism. Nevertheless, an empirically supported theory that explains the mechanism underlying sound symbolic mappings is still lacking. The prominent theoretical view in the literature proposes that the movements of our articulators imitate the contours of round and sharp shapes. Chapter 4 provides evidence against this theory and suggests an alternative regarding sound symbolic associations.

Chapter 4 investigates action knowledge as the mechanistic basis of sound symbolic mappings. In a series of 2AFC tasks, human subjects had to perform the classic sound symbolic associations, and in a second paradigm, they had to map the sounds of sharp or round action movements to the sharp or round visual by-products of these movements. Both conditions were also crossed. Finally, subjects' attention levels were evaluated with two control tasks. Overall, congruency detection was significantly above chance for all the forced choice tasks. The most striking result emerged from the significant correlation of subjects' performances between the sound symbolic and the action sound-shapes tasks but with none of the attentional tasks. "Good" mappers in sound symbolism were equally "good" mappers in action sound-shape mappings, and vice versa for the "bad" mappers. A detailed comparison of the audiovisual by-products of the hand movements and of the auditory and visual properties present in sound symbolism, revealed physical similarities. In addition, two preliminary studies showed the same effects and provide robust evidence for the correlations between the mappings of action sounds to shapes and sound symbolic performance. These results show that action knowledge and knowledge of the audiovisual products of these actions can explain the mappings between meaningless speech sounds to abstract shapes. Actions are the missing link behind mapping meaningless speech sounds and abstract shapes. The model of hand actions can be supported from a neurobiological perspective by distributed neuronal circuits in the human brain (action-perception theory) that carry, via long-distance cortical connections, information between motor and sensory cortices, and hence the perceptual and motor

knowledge of our hand actions.

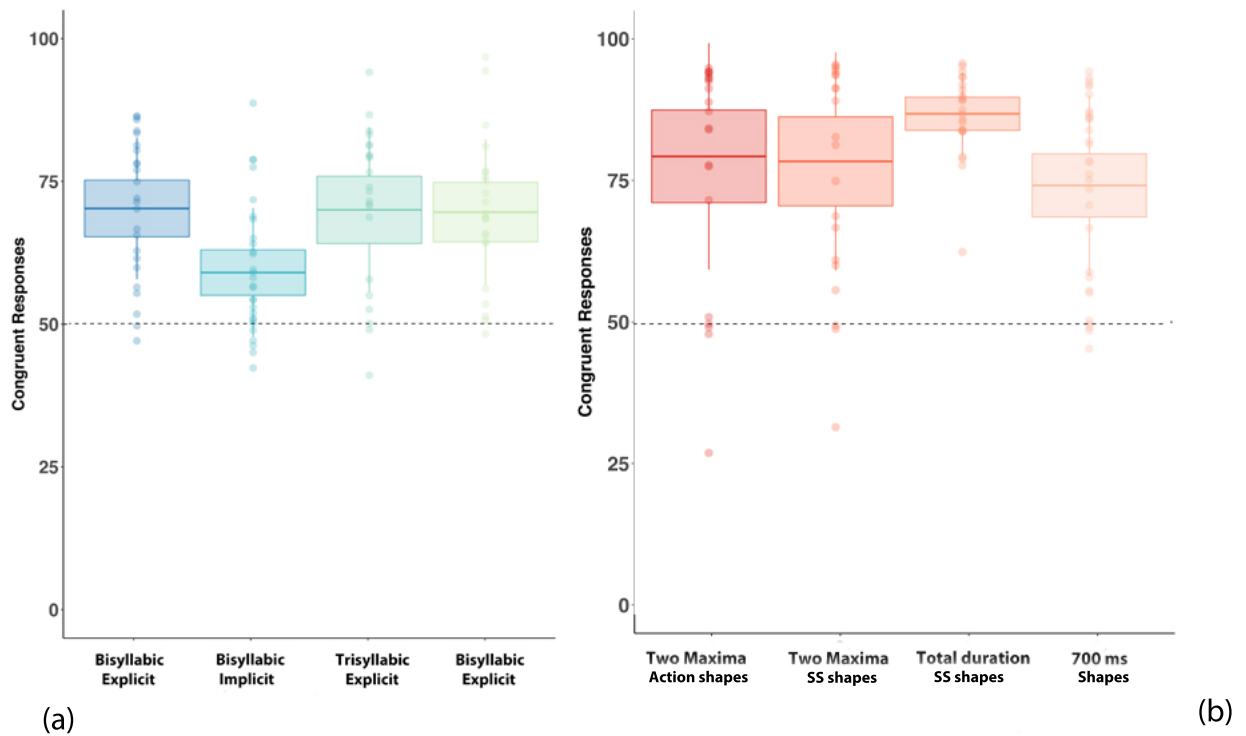


Figure 5.1.: a) Proportion of congruent responses for all studies testing sound symbolism.

The x-axis depicts the number of syllables of the pseudowords and the type of instructions (explicit vs. implicit). b) Proportion of congruent responses for all the experiments including action sounds. The x-axis depicts the durations of the action sounds and the type of the shapes presented (action shapes vs. sound symbolic shapes). Boxplots show standard deviations, lines show means and the whiskers show 95% confidence intervals (CIs).

5.2. Interpretation of findings

The aim of the present dissertation is to explore the origins and the mechanisms of the most-studied sound symbolic mapping in humans—namely, the intrinsic relationship between meaningless speech sounds (‘round’ vs. ‘sharp’) and abstract round and sharp visual shapes. Specifically, the three main objectives were the following: (i) validate

sound symbolic effects in a 2AFC task and explore the relationship of sound symbolism with the two crossmodal mappings (pitch-shape and pitch-spatial position) (ii) investigate the phylogenetic origins of sound symbolic ability, and (iii) test the hypothesis that action knowledge can provide a mechanistic explanation for sound symbolic mappings.

A series of behavioral experiments with different variations of a 2AFC task were used to test human and non-human primates and revealed the following three major findings: (i) validation of sound symbolic effects in a 2AFC task, but no detection of two other intuitive audiovisual mappings, (ii) sound symbolic ability is a human-specific ability, and (iii) a novel proposal on the role of action knowledge behind the mechanism of sound symbolic ability. It is worth mentioning that the findings in humans have strong experimental reliability; the results for both sound symbolism and action mappings were reproducible across studies, even under some modifications either in the auditory and/or visual stimuli (see Fig. 5.1).

As discussed in the Introduction, sound symbolic ability is a topic of special interest for theoretical and empirical approaches on language evolution and acquisition, as humans across the world share the ability to make intuitive mappings between meaningless speech sounds and abstract visual shapes, regardless of their native languages. Sound symbolism allows for the direct expression of semantic knowledge in respect to the sensory properties of a referent. A neurobiological model that explains why sound symbolism is specific to humans (Chapter 3), and correlates with the audiovisual mappings of hand actions (Chapter 4), should involve a brain network recruited for higher cognitive abilities in humans. Moreover such a model would propose that this brain network developed differently during the course of evolution and supports the integration of sensory and motor information. Action-perception circuits (APCs) for hand actions are the best neurobiological model at hand to explain the present findings and the evolution of symbolic ability in humans.

The following discusses in more detail the role of action-perception theory in humans' sound symbolic ability, the relevance of the findings in the evolution of human language,

and crossmodal matching ability. Finally, the last section presents limitations and future research perspectives.

5.2.1. Action-perception circuits (APCs) and the arcuate fasciculus

In order to store the knowledge acquired by our sensory and motor-based experience, we need its memory traces. These traces need to be linked to the motor and sensory cortices in our brain that carry this knowledge. Such a link cannot be explained by theories proposing that our conceptual knowledge is stored in specific semantic hubs in the human brain (Patterson et al., 2007). In opposition, the view that memory traces of this multimodal knowledge are carried by distributed neuronal networks in the human brain (Fuster, 1999, 2009) can support the connection between our perceptual and motor knowledge. These grounded memory models suggest the presence of distributed neuronal ensembles in the brain that link information from different cortical areas, known as APCs (Garagnani et al., 2008; Tomasello et al., 2017; Pulvermüller, 2018a).

These APCs (cell assemblies) are built on the principles of Hebbian learning (Hebb, 1949). According to Hebbian correlation learning, distributed cell assemblies emerge based on the principles of long-term potentiation (LTP) and long-term depression (LTD) of neurons; this phenomenon is most popularly paraphrased as “cells that fire together wire together” (Shatz, 1992). In other words, when neurons fire together, they create a cell assembly, that is a set of neurons that are strongly connected to each other, whereas when they are desynched (LTD), they delink. The correlated activity of auditory, visual, and action modules gives rise to the creation of cell assemblies, or APCs.

These APCs can provide a neurobiological ground for language production and perception (Pulvermüller, 1999; Pulvermüller and Fadiga, 2010). Theories on distributed neural circuits propose that these networks are reused to carry higher cognitive abilities in humans, such as language (Anderson, 2010, 2016; Pulvermüller, 1999), while initially they supported other basic motor and sensory functions.

Notably, these circuits are not shaped only by experience and associative learning, but also require both genetically determined neuroanatomical structural connectivity and associative learning in order to emerge (Pulvermüller and Fadiga, 2010; Pulvermüller, 2018a). A key brain structure relevant in supporting these circuits is the arcuate fasciculus (AF).

The arcuate fasciculus

The arcuate fasciculus (AF), a left lateralized white-matter fiber track with frontotemporal connections, offers a neuroanatomical ground for the connection between motor and sensory knowledge. The AF is already present in human infants (Dubois et al., 2009) and spreads from the inferior frontal to the posterior temporal cortices (de Schotten et al., 2012; Rilling et al., 2008; Rilling, 2014). Individual structural connectivity differences of this frontotemporal circuit are related to phonological memory in humans (Yeatman et al., 2011) and word learning abilities (López-Barroso et al., 2013).

Moreover, comparative tractography studies have shown that during phylogenetic evolution, the AF formed stronger connections in humans compared to other non-human primates. Compared to apes and monkeys, the AF connectivity in humans is stronger posteriorly at the middle and inferior temporal gyrus, and anteriorly at pars opercularis, pars triangularis, and pars orbitalis, (Rilling et al., 2008, see Fig. 5.2). In addition, the human AF reaches the posterior inferior temporal cortex, a key structure of the ventral visual stream (Rilling et al., 2012). This structure is relevant for the integration of visual-motor information in gestures as proposed in Pulvermüller (2018a).

A significant demonstration of how the neuroanatomical evolution of the AF can functionally depict differences between humans and non-human primates comes from a modelling study, in which a “human” and a “monkey” frontotemporal network were trained on novel articulatory-acoustic patterns. Associative learning was present in both networks; however, the human frontotemporal connectivity allowed the formation of circuits with long-lasting reverberating activity, resulting in better verbal working

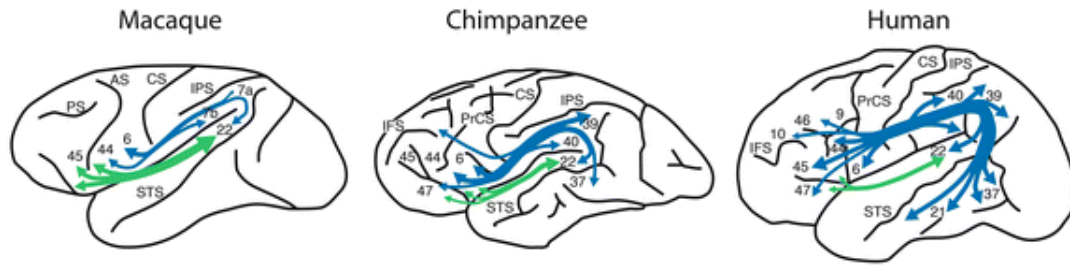


Figure 5.2.: AF dorsal stream connectivity (blue) in macaques, chimpanzees and humans. Adapted from (Friederici, 2017), published under (CC-BY). Original figure from (Rilling et al., 2008).

memory (Schomers et al., 2017). Interestingly, the recruitment of working memory for the retrieval of music tones has been shown to recruit the same sensorimotor network, providing evidence that these circuits do not support only language (Koelsch et al., 2009).

Finally, structural differences of the frontotemporal connectivity are present also in populations with developmental disorders and impairments in language (Catani et al., 2016) and in other non-linguistic domains (Moseley and Pulvermüller, 2018).

5.2.2. APCs: carriers of sound symbolism

Taken together, there is strong evidence that the distinct structural connectivity of the AF in humans permits the generation of action-perception circuits, which carry the memory traces of auditory, visual, and motor knowledge.

As APCs do not only carry language production and perception but also the memory traces of other cognitive abilities (such as music production and processing (Novembre and Keller, 2014)), the integration of action information and the audiovisual by-products of these actions could be supported by these same circuits. The findings of Chapter 4 suggest that the subjects were categorized as “good” and “bad” mappers, as their performance in a classic sound symbolic task and in mapping action sounds to shapes was positively correlated. Subjects who could easily infer sound symbolic associations

inferred equally well action sound-shape associations. Notably, detailed examinations of the audiovisual by-products of our hand actions and of the sound symbolic audiovisual stimuli used in Chapter 4, revealed physical similarities between the two categories. Furthermore, as discussed above, individual differences in the strength of AF can affect the linguistic as well other abilities of humans. On that basis, one can arguably assume that individual differences at the strength of the AF could affect the performance of individuals in action sound-shape mappings, and hence in sound symbolic mappings. As a result, the theoretical proposal that the neurobiological mechanistic grounds of sound symbolic links can be found in the APCs for hand actions in the human brain is supported by the present findings.

APCs for hand actions can also explain the human specificity of sound symbolism. Regarding the findings of Chapter 3, as our closest ancestors, the great apes, lack the genetically determined infrastructure to develop APCs for hand actions, they also lack the ability to store and carry the representations of the perceptual and motor outputs of actions. Thus, they lack the ability to infer or detect audiovisual mappings that sound and look similar to the audiovisual by-products of hand actions. Despite that, one could claim that the lack of sound symbolic congruency detection in great apes resulted from their inability to produce all these phonemic variations present in sound symbolic pseudowords, due to anatomical differences in their vocal apparatus, such as the absence of a descend larynx (Lieberman, 1984), or due to a lack of vocal control over their vocal apparatus. However, the vocal abilities of non-human primates still remain a controversial topic, with studies proposing that monkeys have a speech-ready track (Fitch et al., 2016), that their vocalizations can have some vocal properties similar to human vowels (e.g., F1/F2 formants) (Boë et al., 2017), and that they can voluntarily vocalize (for a discussion, see Perlman, 2017). Even by considering the scenario in which great apes could voluntarily produce various complex vocalizations and imitate the auditory by-products of round or sharp hand actions and hence produce ‘round’ and ‘sharp’ sounding pseudowords, this ability would still not be sufficient to support sound symbolism, as they lack an efficiently developed neuroanatomical connectivity to carry

action-perception circuits for hand actions. Consequently, great apes lack the knowledge of the auditory, visual and motor outputs that offer the basis of sound symbolic mapping. The distinct frontotemporal connectivity of humans could neurobiologically explain the human specificity of sound symbolic ability.

Except of the AF, there are other white matter bundles, shared between human and non-human primates, that can be relevant for the integration of perceptual and motor outputs of hand actions and thus for sound symbolic processing (Bryant et al., 2020). For example, the superior longitudinal fasciculus (SLF), a long white matter tract connecting lateral frontal to lateral parietal regions, whose functions have been linked to visuospatial processing and attention (De Schotten et al., 2011), appears to be of major importance for the coupling of perceptual and motor information during action execution. Moreover, findings from a comparative tractography revealed a unique SLF connectivity with the inferior frontal gyrus (IFG) in humans but not in chimpanzees (Hecht et al., 2015). The presence of this connectivity with the IFG is proposed to be of major importance for the evolution of fine motor control in humans (Hecht et al., 2015).

Finally, another long associative bundle in the human brain and potentially relevant for the emergence of APCs for actions is the inferior fronto-occipital fasciculus (IFOF). This white matter tract spreads from occipital cortex to superior parietal and frontal lobe (Martino et al., 2010), and its frontal terminations partially overlap with the AF and the SFL (Sarubbo et al., 2013). In addition, the surface projections of the IFOF in humans are more extensive than in chimpanzees, reaching the prefrontal cortex. Although the exact role of the IFOF is not well defined, due to its different components (Wu et al., 2016), its function has been related to a series of cognitive abilities such as visual attention (Rollans and Cummine, 2018) and sensorimotor integration (Sarubbo et al., 2013). The functional aspects of IFOF as well as the presence of neuroanatomical differences between humans and non-human primates suggest a potential involvement of IFOF in the emergence of action-perception circuits for hand actions.

5.2.3. Theoretical implications for language evolution

According to embodied semantic theories of language, meaning is grounded in our perceptual and action knowledge and processed in corresponding sensory, motor, and multi-modal cortices in the human brain (Barsalou, 2008; Fischer and Zwaan, 2008; Glenberg and Kaschak, 2002; Glenberg and Gallese, 2012; Pulvermüller, 2012). These theories are in accordance with the above mentioned neurobiological model of APCs that explains how distributed circuits can carry meaning in the human brain (e.g., Pulvermüller et al., 2005; Shebani and Pulvermüller, 2013; Shtyrov et al., 2004). Sound symbolism—namely, the immanent links between meaningless speech sounds and visual shapes—is a great example of how meaning is embodied and includes perceptual and motor representations. According to the findings of Chapter 4, it became evident that sound symbolism is linked to the perceptual and motor outputs of our interactions with the environment. The similarities between the auditory and visual outputs of our hand actions and sound symbolic pseudowords and abstract shapes show the link between sound symbolic mappings and perceptual and motor representations of our hand actions. The sound form of a ‘round’ or ‘sharp’ sounding pseudoword is similar to the auditory output of a round or sharp hand movement. Thanks to these acoustic similarities the meaning of the sound form ‘maluma’ includes information about visual shape and action. These information derive from the associations established between the auditory, motor, and visual outputs of our hand actions. Hence the pseudoword ‘maluma’ is ‘round’ sounding but also round in the visual modality, and round as a motor representation. These findings favor the view that sound symbolic ability can be grounded in our sensory and motor systems—the same systems that support the integration of sensory and motor information available from our interactions with the environment we live in. From an evolutionary perspective (if indeed, the same neuronal circuits support sound symbolic processing and the integration and knowledge of our actions in the human brain) this neurobiological link is of particular interest regarding the emergence of human language.

First, under the neural reuse hypothesis (Anderson, 2010, 2016), it is plausible that

the APCs devoted initially to multimodal integration, such as the matching of ‘round’ and ‘sharp’ action sounds to round and sharp visual prints generated by our hand movements, were recruited later on for supporting higher cognitive abilities in humans, such as language. In other words, humans first used their frontotemporal neuroanatomical architecture under Hebbian learning principles, to store memories between their action knowledge and the output of their actions, and later used the same network to support sound symbolic associations.

Sound symbolic emergence would require also the ability to imitate the auditory products of our actions, via vocal iconicity. Vocal imitation of the sounds of our actions would have been possible if our vocal repertoire and the sounds of our actions showed some physical acoustic similarities. These similarities are highlighted in Chapter 4.

Vocal imitation is easily identified in onomatopoeia, in which there is a direct “translation” of an auditory signal (e.g., the chirp of a bird) to an auditory channel—namely, our voice (Assaneo et al., 2011). However, as onomatopoeia is limited to the expression of auditory meaning, the role of sound symbolism becomes important because it permits the expression of meaning beyond the auditory modality. As shown in Chapter 4, the meaningless speech sounds in sound symbolism express meaning about the roundness or sharpness of abstract shapes linked to our action knowledge. Indeed, beyond the vocal imitation of auditory signals, recent evidence proposes that humans are capable of identifying and producing meaningless speech sounds to communicate meaning for several other non-auditory modalities (Lemaitre et al., 2016; Perlman and Lupyan, 2018). Even if meaningless vocalizations are produced to communicate information about other modalities, it is possible that somehow these modalities are linked to some acoustic information. This scenario would be in accordance with the present findings showing that although sound symbolism initially seems to communicate information about shape, the sound symbolic pseudowords are imitations of the auditory outputs of our hand actions. As a result, additional investigation is needed to understand what such meaningless speech sounds refer to (Lemaitre et al., 2016; Perlman and Lupyan, 2018), and whether they have some links to acoustic events.

Taken together, it is likely that the human ability for sound symbolic communication is a *mélange* of neuroanatomical, anatomical, and cognitive abilities. First, as discussed above, the prerequisites for the emergence of the APCs for hand actions require a neuroanatomical infrastructure. Moreover, the imitation of the sounds produced by different hand actions is not possible by our closest ancestors as they lack vocal imitation abilities (Tomasello, 2010) or at least they have very limited abilities to imitate rich acoustic stimuli (for a discussion, see Perlman, 2017).

Furthermore, for the generation of rich conceptual knowledge and memories under an embodied view, it is necessary to have a rich and flexible interaction with our environment. For that reason, two other factors that could have allowed the emergence of sound symbolic communication in humans are bipedalism and skillful tool-use. On the one hand, bipedalism allowed the freeing of hands and hence their usage for skillful manual activities such as tool use. On the other hand, tool-use could have enriched human sensorimotor interaction with the world's referents (Larsson, 2015), and relevant APCs for hand actions could have emerged. Humans by freeing their hands could interact easier with their natural environment, produce sounds with their manual activities, and later imitate those sounds in order to communicate about them. Lastly, the need to communicate about tool-use to others requires other socio-cognitive abilities, such as the ability to represent other's mental states (i.e., the theory of mind (for a review, see Call and Tomasello, 2008)) and communicative cooperation in the context of joint goals (Hare and Tomasello, 2004; Moll and Tomasello, 2007).

Finally, Köhler claimed that the “maluma-takete” associations could have been present in primitive languages and proposed that sound symbolism emerged from similarities between different modalities, a view similar to the synaesthetic mechanism of Ramachandran and Hubbard (2001). The findings of the present dissertation suggest a different mechanism. Specifically, the association of perceptual and motor aspects of our actions allows the emergence of immanent links between meaningless ‘round’ or ‘sharp’ speech sounds and round or sharp shapes.

5.2.4. From crossmodal correspondences to sound symbolism

Similar to sound symbolism, there are other various immanent links between modality-specific features shared by humans, known as crossmodal correspondences (Spence, 2011), as discussed in Chapter 2. For a long time, the intuitive systematic associations between different modality features and sound symbolism were all considered as expressions of crossmodal correspondences (for a discussion, see Parise, 2016). However sound symbolism is a linguistic phenomenon consisting of meaningless speech sounds and abstract shapes and not just of low-level perceptual properties present in crossmodal correspondences, as described in Chapter 2. Nevertheless, even if these mappings are different from each other, it is possible that they share some mechanisms and belong to a continuum of crossmodal mappings.

Inspection of the pitch-shape correspondence mentioned in Chapter 2, where high-pitched tones fit better to sharp shapes and low-pitched tones to round shapes (O’Boyle and Tarte, 1980), the role of action knowledge and the physical similarities between sound symbolic pseudowords and action sounds could also give a mechanistic explanation for this correspondence. The frequency patterns observed between ‘round’ and ‘sharp’ pseudowords and action sounds were similar (see Chapter 4). Higher frequencies for ‘sharp’ sounding pseudowords and lower frequencies for the ‘round’ ones. As a result higher frequencies are mapped better to sharp visual shapes and lower frequencies to round visual shapes.

Nevertheless, the above chance congruency detection performance for sound symbolic mappings only (see, Chapter 2) suggests that the pseudowords had additional acoustic properties —beyond frequency —that attracted the attention of the subjects and determined their responses. For example, the signal transition was one of these properties, with ‘sharp’ (‘round’) sounding pseudowords having sudden (smooth) transitions. On the other hand, if other acoustic properties of the pseudowords determined the decisions of the subjects, then pitch information was neglected and did not determine their responses for the pitch-spatial position mappings. For this last mapping, it is important

to mention again that the instructions of the task could have likely created a response bias. The instructions guided subjects to match sounds to shapes and not to the spatial position of the shapes. Consequently, any clear conclusions in respect to the pitch-shape mappings cannot be derived.

Although the two “basic” crossmodal mappings presented in Chapter 2 were possibly overshadowed by sound symbolism, the ability to associate modality-specific features that are somehow “compatible” with each other and shared by the general population, could have set the ground for the emergence of sound symbolic mappings (Cuskley and Kirby, 2013). In the case of pitch-shape, this mapping is possibly a “simplified” version of sound symbolism, as the frequencies of the ‘round’ and ‘sharp’ action sounds result in different round or sharp visual outputs. Regarding the pitch-spatial location mapping, despite the absence of shared properties with sound symbolism, this mapping requires also an ability for correlation learning (Parise et al., 2014), similar to the one involved in the formation of APCs in the brain.

Besides, there is already evidence supporting such a crossmodal continuity hypothesis. While great apes can perform crossmodal links between pitch and luminance (i.e., a low-pitched (high) tone mapped to a dark (bright) stimulus) (Ludwig et al., 2011), they cannot infer sound symbolic associations (Margiotoudi et al., 2019). These findings imply that some ability for crossmodal association is present in our close ancestors, and therefore the ability of correlation learning of co-presented audiovisual features. However, as non-human primates lack the necessary neuroanatomical infrastructure to support correlation learning of motor and perceptual information and a human-like rich manual action repertoire in order to strength the sensorimotor representations of these actions, they can not support sound symbolic mapping.

5.3. Limitations and perspectives

This section discusses some limitations of the present studies, and possible research perspectives.

In all the studies presented in this dissertation, the same 2AFC task was adopted. However, the two alternatives present in a forced choice task can come with some limitations. For that reason, there is some criticism in the literature of sound symbolism and of crossmodal correspondences on this type of task (Bentley and Varon, 1933; Dingemanse et al., 2015; Parise, 2016). For instance, the presentation of two versus four alternatives in a learning sound symbolic task, in which subjects had to learn the mappings between pseudowords and shapes revealed that two alternatives facilitated the decision of the subjects on mapping pseudowords to abstract shapes, in contrast to four alternatives (Aveyard, 2012). Nevertheless, the studies of the present dissertation did not focus on the learning effects of sound symbolism, but rather aimed at exploring basic aspects of this effect —namely, its origins and mechanism. For that reason, we had to select a task that would easily capture the intuitive links between meaningless speech sounds and abstract visual shapes, without introducing any additional and demanding task-related factors that could have interfered with these immanent mappings (e.g., speed or response feedback).

In respect to the findings of Chapter 2, the results of the 2AFC task demonstrated the emergence of sound symbolic mappings only, when sound symbolism was co-presented with two crossmodal correspondences. While sound symbolic mappings were detected by the subjects, audiovisual crossmodal mappings that include basic audiovisual properties, such as pitch and spatial position, were not detected. A limitation to this study, particularly for the pitch-spatial position mapping was the instruction. Subjects were instructed to focus on the shape choice and not on shapes' spatial positions. Hence, despite the absence of pitch-spatial position effects when tested with a forced choice task, we cannot exclude the possibility that these mappings would emerge under explicit instructions. In addition, testing the same three mappings under a priming speeded classification task,

similar to the one used for pitch-spatial location mapping (Evans and Treisman, 2009), could lead to different results regarding the interaction of these mappings. Indeed, this type of task requires fast responses, and hence any audiovisual integration could happen almost in an automatic manner. A priming task would also improve our understanding regarding sound symbolism and low-level audiovisual mappings, and whether all these mappings take place in an automatic fashion at a perceptual level or require attention.

All mappings tested in Chapter 2 were limited to the interactions between auditory and visual modality-specific properties. Future studies could also investigate the interaction of pseudoword-shape-type sound symbolic mappings with other sound symbolic types that include other sensory properties, such as tactile information. For example, whether sound symbolic mappings of a pseudoword-shape type are still detected when co-presented with mappings of pseudoword-texture type (i.e., a ‘round’ sounding pseudoword is matched to a soft texture and a ‘sharp’ sounding pseudoword to a rough texture, Etzi et al., 2016). More in-depth investigation on the interaction between different mappings, and particularly their interaction with sound symbolic mappings, could improve our understanding of the relations among modality-specific features present in these immanent associations, and on their common or different origins.

The findings of Chapter 3 provided evidence for the human specificity of sound symbolism, after testing with the same 2AFC healthy humans and great apes. Moreover, a neurobiological explanation followed by these findings proposes that this ability can relate to the distinct neuroanatomical connectivity of the frontotemporal circuit in the human brain. Hence, if sound symbolic ability is indeed related to the human brain’s anatomical connectivity, then language learning should not affect our ability to match meaningless speech sounds to abstract shapes. To provide an answer to this hypothesis, sound symbolism should be tested in a language-trained ape, which will be much more exposed and trained to linguistic material than the apes tested in Chapter 3. To that end, we are currently testing sound symbolism in the language-trained bonobo “Kanzi” (Savage-Rumbaugh et al., 1986, 1993). To sum up, if linguistic competence affects sound symbolic ability, then Kanzi should be able to infer crossmodal mappings, even with a

non-human frontotemporal structural connectivity.

Additionally, it is important to emphasize that the auditory stimuli tested in both species were pseudowords, namely human speech. Although the great apes tested in Chapter 3 are hosted in a research facility in which they interact on a daily basis with humans, and as a result exposed to human speech, testing them with species specific auditory stimuli could validate and consolidate the present findings. This would require the generation of a series of meaningless ‘round’ and ‘sharp’ sounding ape vocalizations. As vocal production is relatively fixed and limited in these species (Seyfarth and Cheney, 2010), the creation of such series remains a challenge.

Furthermore, the present findings on the sound symbolic ability of great apes are discussed in the context of neuroanatomical differences between human and non-human primates. However, the presence of crossmodal mapping between luminance and brightness in the same species (Ludwig et al., 2011) cannot be explained by the same neuroanatomical network. It is therefore important to explore the parallels between crossmodal detection in non-human primate species and neuroanatomical connectivity in relevant brain structures. In order to understand whether sound symbolism belongs to a behavioral and neurobiological continuum of crossmodal correspondences, it is crucial the parallel testing of behavioral and neurobiological factors on crossmodal detection in non-human primates.

Finally, the findings of Chapter 4 suggest a plausible neurobiological mechanistic explanation for the mechanism behind sound symbolic mappings. By a series of behavioral experiments, Chapter 4 provides strong evidence for the importance of action knowledge in sound symbolic processing. However, the findings did not imply a causal relationship. Future studies could explore whether a causal relationship holds between sound symbolism and action-perception knowledge by disrupting the activity of the action-perception circuit and specifically of the motor cortex, using transcranial magnetic stimulation (TMS), while human subjects perform a task on sound symbolic congruency detection. If indeed action knowledge and the by-products of these actions offer the mechanistic

ground for sound symbolic ability, then stimulation of the motor cortex should functionally affect sound symbolic congruency detection performance.

5.4. Conclusion

The findings of the present dissertation contributed to fundamental issues pertaining to our ability to match meaningless ‘round’/‘sharp’ speech sounds to round/sharp abstract shapes. When co-presented with other immanent crossmodal audiovisual mappings, sound symbolic links were validated in a forced choice task, while other audiovisual mappings did not emerge. The rich and complex acoustic information available in sound symbolic pseudowords, beyond pitch, guided the subjects’ attention to sound symbolic mappings. Sound symbolism was found also to be an ability specific to humans. Despite the ability of non-human primates to infer other crossmodal mappings, sound symbolism is present only in humans and can be related to our general linguistic ability, as well as to the distinct neuroanatomical structural connectivity of the human brain. Finally, the mechanism behind sound symbolic links was identified in the knowledge of our hand actions and in the audiovisual by-products of these actions. Sound symbolic stimuli and action sounds and shapes appear to have several physical similarities that support the key role of actions in the mappings of auditory to visual stimuli. From a neurobiological perspective, action-perception circuits for hand actions could explain these findings. Distributed action-perception circuits in the human brain, which ground our experiences in our motor, sensory, and multimodal cortices and which sustain their memory traces, support the human specificity of sound symbolic ability and its mechanistic foundations. The present work brought advancement in the research of sound symbolism by proposing a new theory that explains the mechanism of this human-specific ability and its link to our sensorimotor experiences.

Bibliography

- Adachi, I. and Fujita, K. (2007). Cross-modal representation of human caretakers in squirrel monkeys. *Behavioural Processes*, 74(1):27–32.
- Adachi, I., Kuwahata, H., Fujita, K., Tomonaga, M., and Matsuzawa, T. (2006). Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates*, 47(4):350–354.
- Ahlner, F. and Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38(1/4):298–348.
- Aitchison, J. (2000). *The seeds of speech: Language origin and evolution*. Cambridge University Press.
- Allritz, M., Call, J., and Borkenau, P. (2016). How chimpanzees (pan troglodytes) perform in a modified emotional stroop task. *Animal Cognition*, 19(3):435–449.
- Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33(4):245–266.
- Anderson, M. L. (2016). Précis of after phrenology: neural reuse and the interactive brain. *Behavioral and Brain Sciences*, 39:e120.
- Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., and Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, 63:196–205.
- Assaneo, M. F., Nichols, J. I., and Trevisan, M. A. (2011). The anatomy of onomatopoeia. *PloS one*, 6(12):e2831.

- Aveyard, M. E. (2012). Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory & Cognition*, 40(1):83–92.
- Bar, M. and Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, 17(8):645–648.
- Bard, K. A. (1998). Social-experiential contributions to imitation and emotion in chimpanzees. *Intersubjective Communication and Emotion in Early Ontogeny*, ed. S. Bråten, pages 208–27.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4):577–660.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Ben-Artzi, E. and Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57(8):1151–1162.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.
- Bentley, M. and Varon, E. J. (1933). An accessory study of " phonetic symbolism". *The American Journal of Psychology*, 45(1):76–86.
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80(2):290–311.
- Berlin, B. and O’Neill, J. P. (1981). The pervasiveness of onomatopoeia in aguaruna and huambisa bird names. *Journal of Ethnobiology*, 1(2):238–261.

- Bertamini, M., Palumbo, L., Gheorghes, T. N., and Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, 107(1):154–178.
- Bestor, T. W. (1980). Plato's semantics and plato's" cratylus". *Phronesis*, 25(3):306–330.
- Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., and Bickel, B. (2019). Human sound systems are shaped by post-neolithic changes in bite configuration. *Science*, 363(6432):eaav3218.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39):10818–10823.
- Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y., Rey, A., and Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (papio papio) suggests pre-hominin speech precursors. *PloS one*, 12(1):e0169321.
- Bohn, M., Call, J., and Tomasello, M. (2016). Comprehension of iconic gestures by chimpanzees and human children. *Journal of Experimental Child Psychology*, 142:1–17.
- Bohn, M., Call, J., and Tomasello, M. (2018). Natural reference: A phylo-and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science*, 22(2):e12757.
- Bonetti, L. and Costa, M. (2018). Pitch-verticality and pitch-size cross-modal interactions. *Psychology of Music*, 46(3):340–356.
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., and Spence, C. (2013). “bouba” and “kiki” in namibia? a remote culture make similar shape–sound matches, but different shape–taste matches to westerners. *Cognition*, 126(2):165–172.

- Bresch, E., Kim, Y.-C., Nayak, K., Byrd, D., and Narayanan, S. (2008). Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [exploratory dsp]. *IEEE Signal Processing Magazine*, 25(3):123–132.
- Brickenkamp, R. and Zillmer, E. (1998). *The d2 test of attention*. Seattle: Hogrefe.
- Bross, F. (2018). Cognitive associations between vowel length and object size: A new feature contributing to a bouba/kiki effect. In *Proceedings of the Conference on Phonetics & Phonology*, pages 17–20.
- Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.
- Brown, R. W., Black, A. H., and Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology*, 50(3):388.
- Bryant, K. L., Li, L., and Mars, R. B. (2020). A comprehensive atlas of white matter tracts in the chimpanzee. *bioRxiv*.
- Call, J. and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5):187 – 192.
- Calvert, G., Spence, C., Stein, B. E., et al. (2004). *The handbook of multisensory processes*. MIT press.
- Catani, M. (2009). The connectional anatomy of language: recent contributions from diffusion tensor tractography. In *Diffusion MRI*, pages 403–413. Elsevier.
- Catani, M., Dell’Acqua, F., Budisavljevic, S., Howells, H., Thiebaut de Schotten, M., Froudist-Walsh, S., D’Anna, L., Thompson, A., Sandrone, S., Bullmore, E. T., et al. (2016). Frontal networks in adults with autism spectrum disorder. *Brain*, 139(2):616–630.

- Chen, Y.-C. and Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114(3):389–404.
- Childs, G. T. (1994). Sound symbolism. In *The Oxford Handbook of the Word*. Cambridge University Press Cambridge.
- Chow, H. M. and Ciaramitaro, V. (2019). What makes a shape “baba”? the shape features prioritized in sound–shape correspondence change with development. *Journal of Experimental Child Psychology*, 179:73–89.
- Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5(1):39–62.
- Cuskley, C. and Kirby, S. (2013). Synesthesia, cross-modality, and language evolution. In *Oxford Handbook of Synesthesia*.
- Cuskley, C., Simner, J., and Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, 81(1):119–130.
- Darwin, C. (1888). *The descent of man: and selection in relation to sex*. John Murray, Albemarle Street.
- Davis, R. (1961). The fitness of names to drawings. a cross-cultural study in tanganyika. *British Journal of Psychology*, 52(3):259–268.
- De Schotten, M. T., Dell’Acqua, F., Forkel, S., Simmons, A., Vergani, F., Murphy, D. G., and Catani, M. (2011). A lateralized brain network for visuo-spatial attention. *Nature Precedings*, pages 1–1.
- de Schotten, M. T., Dell’Acqua, F., Valabregue, R., and Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex*, 48(1):82–96.

- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass*, 6(10):654–672.
- Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics*, 3(1).
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10):603–615.
- Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., and Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, 92(2):e117–e133.
- Dolscheid, S., Hunnius, S., Casasanto, D., and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science*, 25(6):1256–1261.
- Driver, J. and Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 2(7):254–262.
- Dubois, J., Hertz-Pannier, L., Cachia, A., Mangin, J., Le Bihan, D., and Dehaene-Lambertz, G. (2009). Structural asymmetries in the infant language and sensori-motor networks. *Cerebral Cortex*, 19(2):414–423.
- D’Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech*, 57(3):367–393.
- Eco, U. et al. (1976). *A theory of semiotics*, volume 217. Indiana University Press.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5):7–7.
- Etzi, R., Spence, C., Zampini, M., and Gallace, A. (2016). When sandpaper is ‘kiki’ and satin is ‘bouba’: an exploration of the associations between words, emotional states,

- and the tactile attributes of everyday materials. *Multisensory Research*, 29(1-3):133–155.
- Evans, K. K. and Treisman, A. (2009). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, 10(1):6–6.
- Fedurek, P., Slocombe, K. E., Hartel, J. A., and Zuberbühler, K. (2015). Chimpanzee lip-smacking facilitates cooperative behaviour. *Scientific Reports*, 5:13460.
- Fischer, M. H. and Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61(6):825–850.
- Fitch, W. T., De Boer, B., Mathur, N., and Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2(12):e1600723.
- Fodor, J. A. (1983). *The modularity of mind*. MIT press.
- Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., and Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science*, 21(5):e12659.
- Fort, M., Martin, A., and Peperkamp, S. (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech*, 58(2):247–266.
- Fort, M., Weiß, A., Martin, A., and Peperkamp, S. (2013). Looking for the bouba-kiki effect in prelexical infants. In *Auditory-Visual Speech Processing (AVSP) 2013*.
- Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2-3):171–195.
- Friederici, A. D. (2017). Evolution of the neural language network. *Psychonomic Bulletin & Review*, 24(1):41–47.
- Fuchs, S. and Perrier, P. (2005). On the complex nature of speech kinematics. *ZAS papers in Linguistics*, 42:137–165.

- Fuster, J. (2015). *The prefrontal cortex*. Academic Press.
- Fuster, J. M. (1999). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. MIT press.
- Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21(11):2047–2072.
- Gallace, A., Boschini, E., and Spence, C. (2011). On the taste of “bouba” and “kiki”: An exploration of word–food associations in neurologically normal participants. *Cognitive Neuroscience*, 2(1):34–46.
- Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). A neuroanatomically grounded hebbian-learning model of attention–language interactions in the human brain. *European Journal of Neuroscience*, 27(2):492–513.
- Glenberg, A. M. and Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, 48(7):905–922.
- Glenberg, A. M. and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3):558–565.
- Grosse, K., Call, J., Carpenter, M., and Tomasello, M. (2015). Differences in the ability of apes and children to instruct others using gestures. *Language Learning and Development*, 11(4):310–330.
- Hamano, S. (1994). Palatalization in Japanese sound symbolism. *Sound symbolism*, pages 148–157.
- Hare, B. and Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, 68(3):571–581.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346.

- Hashiya, K. and Kojima, S. (2001). Acquisition of auditory–visual intermodal matching-to-sample by a chimpanzee (pan troglodytes): comparison with visual—visual intramodal matching. *Animal Cognition*, 4(3-4):231–239.
- Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. J. Wiley; Chapman & Hall.
- Hecht, E. E., Gutman, D. A., Bradley, B. A., Preuss, T. M., and Stout, D. (2015). Virtual dissection and comparative connectivity of the superior longitudinal fasciculus in chimpanzees and humans. *Neuroimage*, 108:124–137.
- Heimbauer, L. A., Beran, M. J., and Owren, M. J. (2011). A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Current Biology*, 21(14):1210–1214.
- Hinton, L., Nichols, J., and Ohala, J. J. (2006). *Sound symbolism*. Cambridge University Press.
- Hockett, C. F. (1959). Animal" languages" and human language. *Human Biology*, 31(1):32–39.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3):88–97.
- Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology*, 109(2):219–238.
- Imai, M. and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, 369(1651):20130298.
- Imai, M., Kita, S., Nagumo, M., and Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1):54–65.
- Irwin, F. W. and Newland, E. (1940). A genetic study of the naming of visual figures. *The Journal of Psychology*, 9(1):3–16.

- Johansson, N., Anikin, A., and Aseyev, N. (2019a). Color sound symbolism in natural languages. *Language and Cognition*, 12(1):1–28.
- Johansson, N. E., Anikin, A., Carling, G., and Holmer, A. (2019b). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*.
- Jones, D. (1922). *An outline of English phonetics*. BG Teubner.
- Jones, J. M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J., and Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36.
- Kantartzis, K., Imai, M., Evans, D., and Kita, S. (2019). Sound symbolism facilitates long-term retention of the semantic representation of novel verbs in three-year-olds. *Languages*, 4(2):21.
- Kantartzis, K., Imai, M., and Kita, S. (2011). Japanese sound-symbolism facilitates word learning in english-speaking children. *Cognitive Science*, 35(3):575–586.
- Katz, J. J. and Fodor, J. A. (1963). The structure of a semantic theory. *Language*, 39(2):170–210.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kita, S. (2008). World-view of protolanguage speakers as inferred from semantics of sound symbolic words: A case of japanese mimetics. In *The origins of language*, pages 25–38. Springer.
- Knoeferle, K., Li, J., Maggioni, E., and Spence, C. (2017). What drives sound symbolism? different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, 7(1):1–11.

- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fmri study. *Human Brain Mapping*, 30(3):859–873.
- Köhler, W. (1929). *Gestalt psychology*, volume 31. Springer.
- Kojima, S., Izumi, A., and Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates*, 44(3):225–230.
- Kojima, S. and Kiritani, S. (1989). Vocal-auditory functions in the chimpanzee: vowel perception. *International Journal of Primatology*, 10(3):199–213.
- Kojima, S., Tatsumi, I., Kiritani, S., and Hirose, H. (1989). Vocal-auditory functions of the chimpanzee: consonant perception. *Human Evolution*, 4(5):403–416.
- Koppensteiner, M., Stephan, P., and Jäschke, J. P. M. (2016). Shaking takete and flowing maluma. non-sense words are associated with motion patterns. *PloS one*, 11(3):e0150610.
- Kovic, V., Plunkett, K., and Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114(1):19–28.
- Ladefoged, P. and Disner, S. F. (2012). *Vowels and Consonants*. John Wiley & Sons.
- Larsson, M. (2015). Tool-use-associated sound in the evolution of language. *Animal Cognition*, 18(5):993–1005.
- Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., and Susini, P. (2016). Vocal imitations of non-vocal sounds. *PloS one*, 11(12).
- Lewkowicz, D. J. and Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on walker et al.(2010). *Psychological Science*, 25(3):832–834.
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.

- Löbner, S. (2013). *Understanding semantics*. Routledge.
- Lockwood, G. and Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6:1246.
- Lockwood, G., Hagoort, P., and Dingemanse, M. (2016). Synthesized size-sound sound symbolism. Philadelphia, PA: Cognitive Science Society.
- López-Barroso, D., Catani, M., Ripollés, P., Dell’Acqua, F., Rodríguez-Fornells, A., and de Diego-Balaguer, R. (2013). Word learning is mediated by the left arcuate fasciculus. *Proceedings of the National Academy of Sciences*, 110(32):13168–13173.
- Lopez-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T. F., and Rodriguez-Fornells, A. (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cerebral Cortex*, 21(12):2742–2750.
- Ludwig, V. U., Adachi, I., and Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, 108(51):20661–20665.
- Löfqvist, A. and Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research*, 40(4):877–893.
- Margiotoudi, K., Allritz, M., Bohn, M., and Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific Reports*, 9(1):1–12.
- Margiotoudi, K. and Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports*, 10(1):1–13.
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3):384.

- Martinez, L. and Matsuzawa, T. (2009). Effect of species-specificity in auditory-visual intermodal matching in a chimpanzee (pan troglodytes) and humans. *Behavioural Processes*, 82(2):160–163.
- Martino, J., Brogna, C., Robles, S. G., Vergani, F., and Duffau, H. (2010). Anatomic dissection of the inferior fronto-occipital fasciculus revisited in the lights of brain stimulation data. *Cortex*, 46(5):691–699.
- Matsuzawa, T. (1990). Form perception and visual acuity in a chimpanzee. *Folia Primatologica*, 55(1):24–32.
- Maurer, D., Pathman, T., and Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental science*, 9(3):316–322.
- McCormick, K., Kim, J., List, S., and Nygaard, L. C. (2015). Sound to meaning mappings in the bouba-kiki effect. *Cognitive Science*, pages 1565–1570.
- McNeill, D. (2006). Gesture: a psycholinguistic approach. *The encyclopedia of language and linguistics*, pages 58–66.
- Melara, R. D. and O’Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116(4):323.
- Mikone, E. (2001). Ideophones in the balto-finnic languages. *Typological Studies in Language*, 44:223–234.
- Miyazaki, M., Hidaka, S., Imai, M., Yeung, H. H., Kantartzis, K., Okada, H., and Kita, S. (2013). The facilitatory role of sound symbolism in infant word learning. pages 3080–3085. Cognitive Science Society.
- Mok, P. P., Li, G., Li, J. J., Ng, H. T., and Cheung, H. (2019). Cross-modal association between vowels and colours: a cross-linguistic perspective. *The Journal of the Acoustical Society of America*, 145(4):2265–2276.

- Moll, H. and Tomasello, M. (2007). Cooperation and human cognition: the vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):639–648.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., and Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651):20130299.
- Mondloch, C. J. and Maurer, D. (2004). Do small white balls squeak? pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2):133–136.
- Moos, A., Smith, R., Miller, S. R., and Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder. *i-Perception*, 5(2):132–142.
- Morris, C. (1946). *Signs, language and behavior*. Prentice-Hall.
- Moseley, R. L. and Pulvermüller, F. (2018). What can autism teach us about the role of sensorimotor systems in higher cognition? new clues from studies on language, action semantics, and abstract emotional concept processing. *Cortex*, 100:149–190.
- Motoki, K., Saito, T., Park, J., Velasco, C., Spence, C., and Sugiura, M. (2020). Tasting names: Systematic investigations of taste-speech sounds associations. *Food Quality and Preference*, 80:103801.
- Mudd, S. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, 66(4):347.
- Munar, E., Gómez-Puerto, G., Call, J., and Nadal, M. (2015). Common visual preference for curved contours in humans and great apes. *PloS one*, 10(11):e0141106.
- Nielsen, A. and Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(2):115.

- Nielsen, A. K. and Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(2):153.
- Nöth, W. (1995). *Handbook of semiotics*. Indiana University Press.
- Novembre, G. and Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Frontiers in Human Neuroscience*, 8:603.
- O'Boyle, M. W. and Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research*, 9(6):535–544.
- Ogden, C. K., Richards, I. A., Malinowski, B., and Crookshank, F. G. (1923). *The meaning of meaning*. Kegan Paul London.
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound symbolism*.
- Ouni, S. (2011). Tongue gestures awareness and pronunciation training. In *Twelfth Annual Conference of the International Speech Communication Association*.
- Ozturk, O., Krehm, M., and Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114(2):173–186.
- Palumbo, L., Ruta, N., and Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PloS one*, 10(10):e0140043.
- Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1-3):7–28.

- Parise, C. V., Knorre, K., and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, 111(16):6104–6108.
- Parise, C. V. and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, 220(3-4):319–333.
- Parr, L. A., Cohen, M., and De Waal, F. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology*, 26(1):73–103.
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976.
- Peirce, C. S. (1960). *Collected papers of charles sanders peirce*, volume 2. Harvard University Press.
- Pejovic, J. and Molnar, M. (2017). The development of spontaneous sound-shape matching in monolingual and bilingual infants during the first year. *Developmental Psychology*, 53(3):581.
- Peña, M., Mehler, J., and Nespors, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, 22(11):1419–1421.
- Perlman, M. (2017). Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies*, 18(3):376–401.
- Perlman, M. and Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific Reports*, 8(1):1–14.
- Perniss, P., Lu, J. C., Morgan, G., and Vigliocco, G. (2018). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science*, 21(2):e12551.

- Perniss, P., Thompson, R., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1:227.
- Perniss, P. and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651):20130300.
- Proops, L., McComb, K., and Reby, D. (2009). Cross-modal individual recognition in domestic horses (*equus caballus*). *Proceedings of the National Academy of Sciences*, 106(3):947–951.
- Pulvermüller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, 22(2):253–279.
- Pulvermüller, F. (2012). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, 25(5):423–459.
- Pulvermüller, F. (2013a). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9):458–470.
- Pulvermüller, F. (2013b). Semantic embodiment, disembodiment or misembodiment? in search of meaning in modules and neuron circuits. *Brain and Language*, 127(1):86–103.
- Pulvermüller, F. (2018a). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, 160:1–44.
- Pulvermüller, F. (2018b). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science*, 10(3):590–620.
- Pulvermüller, F. and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5):351.
- Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21(3):793–797.

- Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia—a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12):3–34.
- Ravignani, A. and Sonnweber, R. (2017). Chimpanzees process structural isomorphisms across sensory modalities. *Cognition*, 161:74–79.
- Rilling, J., Glasser, M. F., Jbabdi, S., Andersson, J., and Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in evolutionary neuroscience*, 3:11.
- Rilling, J. K. (2014). Comparative primate neurobiology and the evolution of brain language systems. *Current Opinion in Neurobiology*, 28:10–14.
- Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., and Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative dti. *Nature Neuroscience*, 11(4):426.
- Rogers, S. K. and Ross, A. S. (1975). A cross-cultural test of the maluma-takete phenomenon. *Perception*, 4(1):105–106.
- Rollans, C. and Cummine, J. (2018). One tract, two tract, old tract, new tract: A pilot study of the structural and functional differentiation of the inferior fronto-occipital fasciculus. *Journal of Neurolinguistics*, 46:122–137.
- Sakamoto, M. and Watanabe, J. (2018). Bouba/kiki in touch: Associations between tactile perceptual qualities and japanese phonemes. *Frontiers in Psychology*, 9:295.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3):225.
- Sarubbo, S., De Benedictis, A., Maldonado, I. L., Basso, G., and Duffau, H. (2013). Frontal terminations for the inferior fronto-occipital fascicle: anatomical dissection, dti study and functional considerations on a multi-component bundle. *Brain Structure and Function*, 218(1):21–37.

- Saussure, F. d. (1959). *Course in general linguistics* (w. baskin, trans.). *New York: Philosophical Library*.
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., and Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58(3):i–252.
- Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., and Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*pan paniscus*). *Journal of Experimental Psychology: General*, 115(3):211.
- Schomers, M. R., Garagnani, M., and Pulvermüller, F. (2017). Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. *Journal of Neuroscience*, 27(11):2693–16.
- Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., and Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31(1):26–35.
- Scott, B. H. and Mishkin, M. (2016). Auditory short-term memory in the primate auditory cortex. *Brain Research*, 1640:264–277.
- Scott, B. H., Mishkin, M., and Yin, P. (2012). Monkeys have a limited form of short-term memory in audition. *Proceedings of the National Academy of Sciences*, 109(30):12237–12241.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–424.
- Seyfarth, R. M. and Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language*, 115(1):92–100.

- Shang, N. and Styles, S. J. (2017). Is a high tone pointy? speakers of different languages match mandarin chinese tones to visual shapes differently. *Frontiers in Psychology*, 8:2139.
- Shatz, C. J. (1992). The developing brain. *Scientific American*, 267(3):60–67.
- Shebani, Z. and Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm-and leg-related action words. *Cortex*, 49(1):222–231.
- Shinohara, K., Yamauchi, N., Kawahara, S., and Tanaka, H. (2016). Takete and maluma in action: A cross-modal relationship between gestures and sounds. *PloS one*, 11(9):e0163525.
- Shtyrov, Y., Hauk, O., and Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European journal of Neuroscience*, 19(4):1083–1092.
- Simner, J., Cuskey, C., and Kirby, S. (2010). What sound does that taste? cross-modal mappings across gustation and audition. *Perception*, 39(4):553–569.
- Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(2):61–70.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4):971–995.
- Spence, C. and Ngo, M. K. (2012). Assessing the shape symbolism of the taste, flavour, and texture of foods and beverages. *Flavour*, 1(1):12.
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hearing Research*, 305:57–73.
- Styles, S. J. and Gawne, L. (2017). When does maluma/takete fail? two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*, 8(4):2041669517724807.

- Team, R. C. et al. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.
- Tomasello, M. (2010). *Origins of human communication*. MIT press.
- Tomasello, R., Garagnani, M., Wennekers, T., and Pulvermüller, F. (2017). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*, 98:111–129.
- Tomonaga, M. and Matsuzawa, T. (1992). Perception of complex geometric figures in chimpanzees (pan troglodytes) and humans (homo sapiens): analyses of visual similarity on the basis of choice reaction time. *Journal of Comparative Psychology*, 106(1):43.
- Walker, P., Bremner, J. G., Lunghi, M., Dolscheid, S., D. Barba, B., and Simion, F. (2018). Newborns are sensitive to the correspondence between auditory pitch and visuospatial elevation. *Developmental Psychobiology*, 60(2):216–223.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., and Johnson, S. P. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21(1):21–25.
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11):483–488.
- Watson, R. L. (2001). A comparison of some southeast asian ideophones with some african ideophones. *Typological Studies in Language*, 44:385–406.
- White, T. D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C. O., Suwa, G., and WoldeGabriel, G. (2009). *Ardipithecus ramidus* and the paleobiology of early hominids. *Science*, 326(5949):64–86.
- Winter, B. (2019). *Sensory linguistics: Language, perception and metaphor*, volume 20. John Benjamins Publishing Company.

- Winter, B., Perlman, M., Perry, L. K., and Lupyan, G. (2017). Which words are most iconic? *Interaction Studies*, 18(3):443–464.
- Wu, Y., Sun, D., Wang, Y., and Wang, Y. (2016). Subcomponents and connectivity of the inferior fronto-occipital fasciculus revealed by diffusion spectrum imaging fiber tracking. *Frontiers in Neuroanatomy*, 10:88.
- Yeatman, J. D., Dougherty, R. F., Rykhlevskaia, E., Sherbondy, A. J., Deutsch, G. K., Wandell, B. A., and Ben-Shachar, M. (2011). Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. *Journal of Cognitive Neuroscience*, 23(11):3304–3317.
- Zlatev, J., Persson, T., and Gärdenfors, P. (2005). Bodily mimesis as “the missing link” in human cognitive evolution. *Lund University Cognitive Studies*, 121:1–45.

A. Appendix Chapter 2

Round	Sharp
bogugu	fezezi
bomodo	feseke
bonolu	fezepi
dolomu	fifike
golulu	kesete
gomodu	kipeki
lomudu	kisezi
ludolu	piteze
mobulu	pizepe
modolu	sitizi
nomunu	tepipi
nobogo	tetipe

Table A.1.: List of trisyllabic pseudowords.

Analysis

In order to explore any further effects on pitch-spatial position congruency performance, we ran a GLMM model with a binomial error structure. The dependent variable was the congruency performance. We included as fixed effects, word type ('round' vs. 'sharp')

sounding) and the pitch of the pseudowords (low vs. high). As random effect, we included intercepts for subject and random slopes for each trial nested within this random effect. We used the likelihood ratio test (LRT) to check if the predictor variables improved the fit of the model; these were calculated by comparing the full model to a reduced model that included all terms except for the fixed effect terms in question.

In addition, in order to explore the congruency performance for the mapping of pitch to shapes, we calculated and compared the congruent responses obtained for the two pitch categories (i.e., high-pitched pseudowords matched to sharp shapes and low-pitched pseudowords to round shapes). Moreover, as previous analysis revealed that three subjects selected more than 80% of the times a round shape, we calculated congruency with and without these subjects.

Results

Comparison of the performance between the full and the reduced model for examining possible effects of word and pitch category on the performance of the subjects for the pitch-spatial location condition revealed no significant results ($\chi^2(2)=1.74, p = 0.41$).

The Wilcoxon signed-rank test against chance, testing congruency for the different categories of pitch-shape mappings revealed that for the low-pitched pseudowords, there was 65.87% congruency and was significantly above chance ($V=294, p < 0.001$). In contrast for the high-pitched pseudowords congruency reached 39.11% and was insignificant ($V=38, p = 0.9$). After the removal of the three subjects, who showed more than 80% preference for round shapes, accuracy for the high-pitched pseudowords increased to 44.16%, but it remained still insignificant ($V=38, p = 0.9$), whereas for the low-pitched pseudowords the accuracy dropped to 61.54% ($V=225, p < 0.01$). Therefore, there was an effect of pitch in the congruency performance of the subjects in pitch-shape mappings (see Fig. A.1 a & b).

Closer examination of these findings, revealed that the higher congruency for the low-

pitched pseudowords in the pitch-shape mappings was not driven by a real map between low-pitched pseudowords and round shapes, but by the mapping of ‘round’ sounding pseudowords to round shapes. As depicted in Figure A.1 b, congruency for low-pitched pseudowords was high only for the ‘round’ sounding and not for the ‘sharp’ sounding pseudowords. If there was a true mapping between low-pitched pseudowords and round shapes, then it should have been present in the ‘sharp’ sounding pseudowords as well. Moreover, we exclude the possibility, that the combination of ‘round’ sounding and low-pitched pseudowords facilitated the mapping of pseudowords to round shapes, as indicated in the Kruskal-Wallis test, evaluating sound symbolic congruency for different pseudoword and pitch categories (see Results 2.4).

	High/Round	High/Sharp	Low/Round
High/Sharp	0.005	-	-
Low/Round	1	0.0002	
Low/Sharp	0.0012	1	0.0008

Table A.2.: Pairwise comparisons with Bonferroni adjusted correction for the congruency performance of the subjects in the sound symbolic condition under the different combinations of pseudoword features.

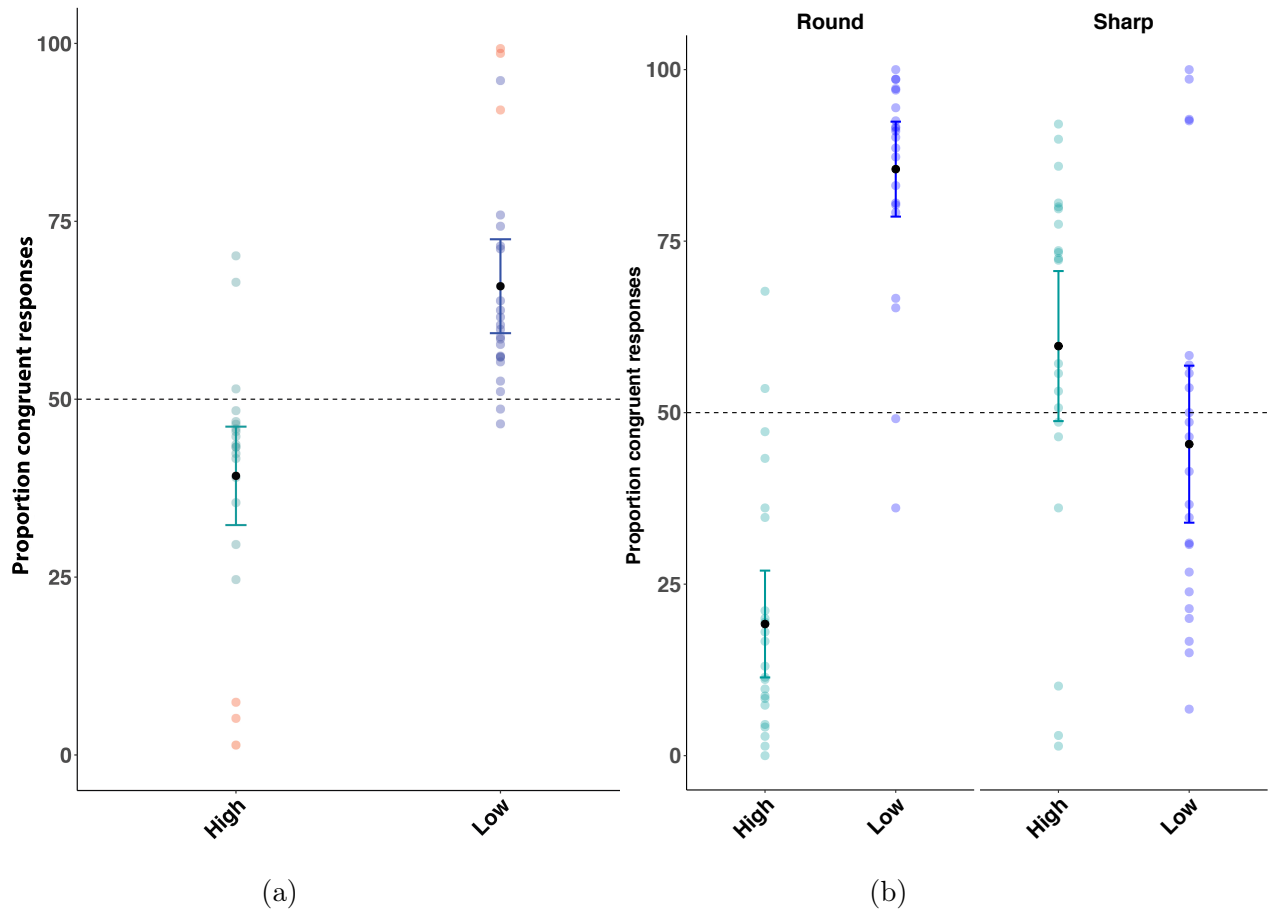


Figure A.1.: a) Proportion of congruent responses for the two pseudoword categories in the pitch-shape condition. Light colored circles indicate congruent responses for each individual for the two pitch categories: high-pitched (green) and low-pitched pseudowords (blue). Red circles indicate the subjects that selected more than 80% of the times the round shapes. Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance. b) Proportion of congruent responses for the pitch-shape condition for the different combinations of pseudoword features (low vs. high-pitched and ‘round’ vs. ‘sharp’-sounding). Light colored circles indicate congruent responses for each individual for the two categories: high-pitched (green) and low-pitched pseudowords (blue). Whiskers show 95% confidence intervals (CIs) and the dashed line at 50% shows chance-level performance.

B. Appendix Chapter 3





















Shapes	Nr	M	S.D	Shapes	Nr	M	S.D
	1	2.36	1.18		11	2.34	0.99
	2	4.79	1.15		12	6.54	1.01
	3	2.31	1.53		13	1.62	0.93
	4	4.73	1.27		14	4.75	1.18
	5	2.26	0.99		15	1.85	0.97
	6	5.68	1.00		16	5.22	1.25
	7	4.93	1.18		17	1.49	1.04
	8	1.80	1.05		18	2.40	1.14
	9	1.58	1.01		19	6.09	0.93
	10	5.01	1.23		20	5.49	1.05

Table B.1.: Ratings of abstract shapes as obtained using an online questionnaire. Each shape selected for the study is listed with a running number, its mean rating (M) on a Likert scale (1-totally sharp; 7-totally round) and its standard deviations (SD).

Sharp	kiki/keke	sisi/sese	fifi/fefe	zizi/zeze	pipi/pepe
Round	nono/nunu	momo/mumu	lolo/lulu	dodo/dudu	gogo/gugu

Table B.2.: Pseudoword stimuli for the two categories ‘sharp’ and ‘round’ sounding used in the experiments with humans and great apes.

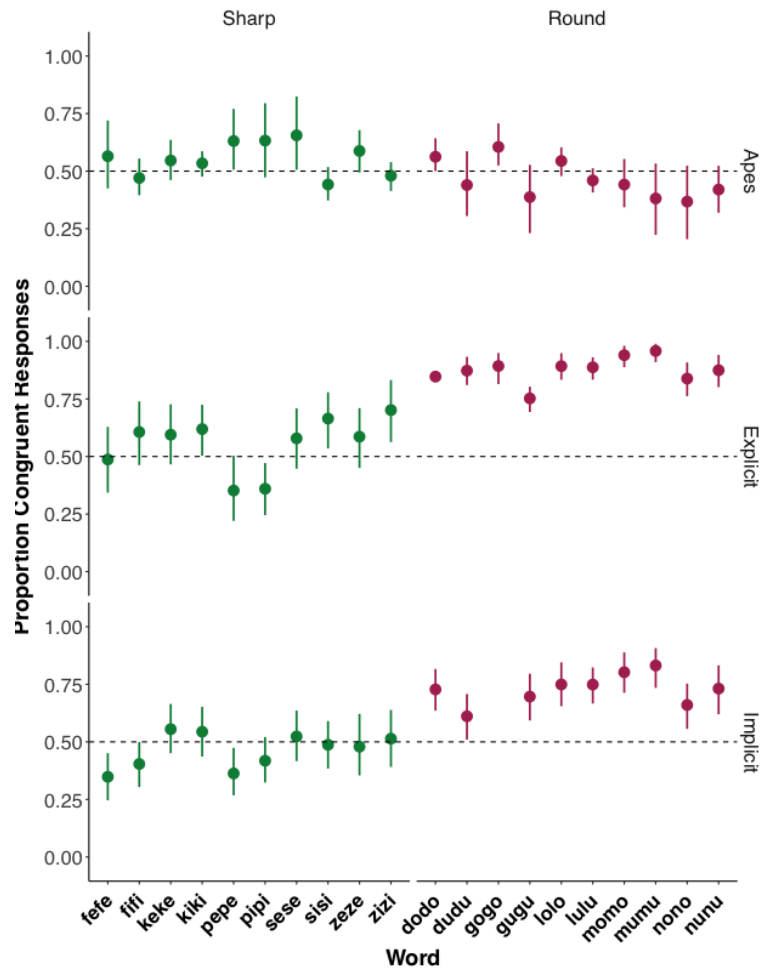


Figure B.1.: Percentage of sound-symbolic congruent responses for each pseudoword obtained from apes and from humans for the explicit and implicit task. Green and maroon circles show the average percentage of congruent responses for each ‘sharp’ and ‘round’ pseudoword separately. The whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance.

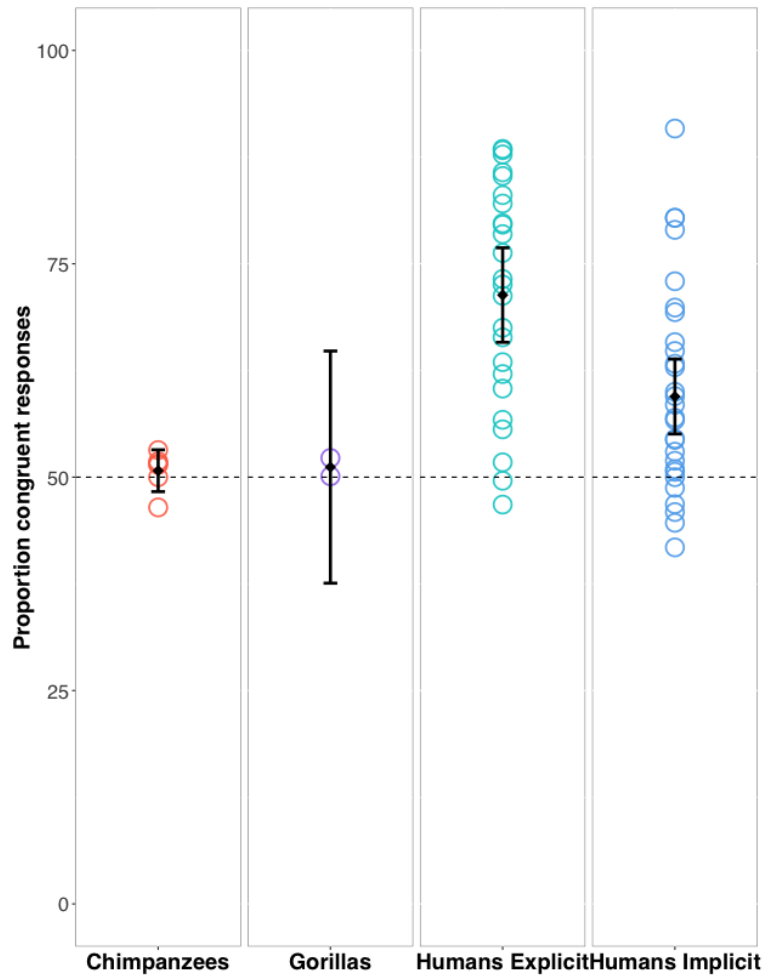


Figure B.2.: Percentage of sound-symbolic congruent responses in chimpanzees, gorillas and in humans tested in the explicit and implicit task, quantified as the proportion of times each individual matched a ‘sharp’ sound to an angular shape or a ‘round’ sound to a curved shape. Orange, purple and cyan and blue circles show the percentage of congruent responses for each. Black diamonds represent the average responses for each species and the whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance level performance.

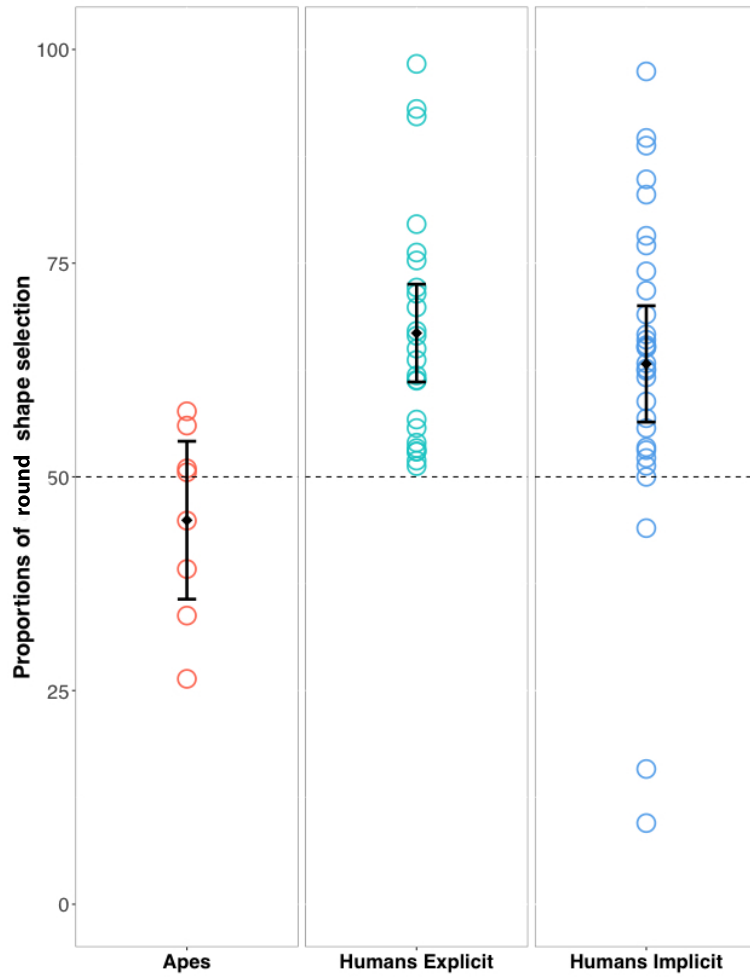


Figure B.3.: Proportion of curved shape selections in apes and in humans from both the explicit and implicit tasks separately. Orange, cyan and blue circles show the proportion of selecting a curved shape for individual chimpanzees and humans for the explicit and task separately. Black diamonds represent the average responses for each species and the whiskers show 95% confidence intervals (CIs). The dashed line at 50% shows chance-level performance.

C. Appendix Chapter 4

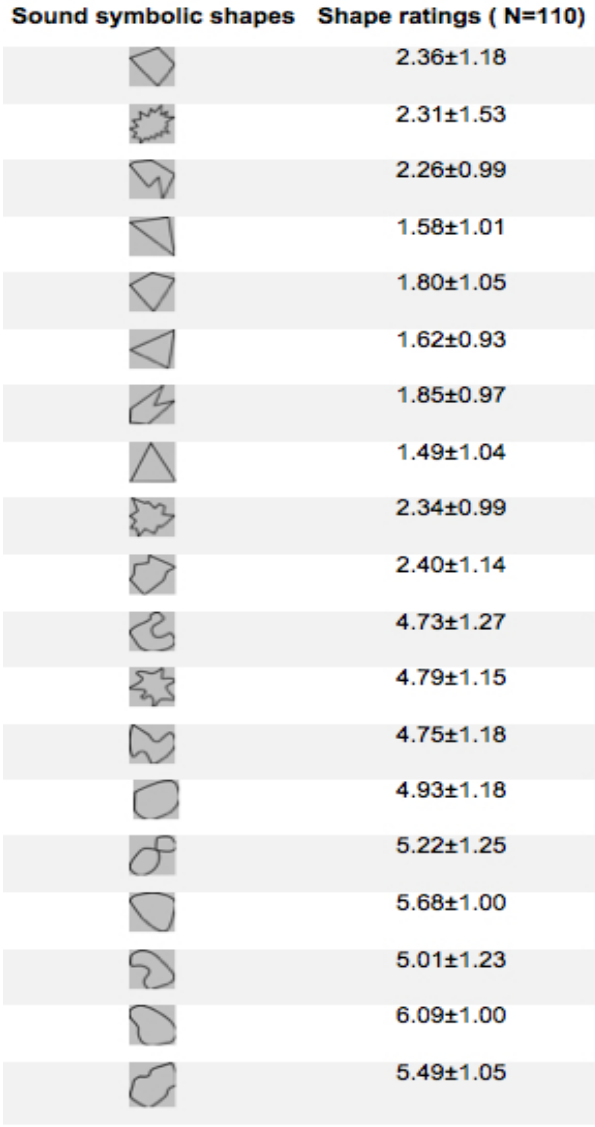


Figure C.1.: Mean ratings (M) and standard deviations (SD) obtained from a a Likert scale (1-totally sharp; 7-totally round) for each abstract sound symbolic shape.











Action shapes	Shape ratings (N=13)	Sound ratings (N=41)
	1.07±0.26	2.19±0.98
	1.15±0.36	2.09±1.04
	1±0	2.24±1.09
	1.15±0.36	2.51±0.86
	2.15±1.79	1.87±0.81
	6.69±0.46	5.60±1.33
	6.76±0.42	5.65±1.45
	6.53±0.63	4.80±1.69
	5.92±0.91	4.85±1.31
	6.78±0.80	5.25±1.60

Figure C.2.: Mean ratings (M) and standard deviations (SD) obtained from a Likert scale (1-totally sharp; 7-totally round) for each action shape and for the sounds produced while drawing these shapes.

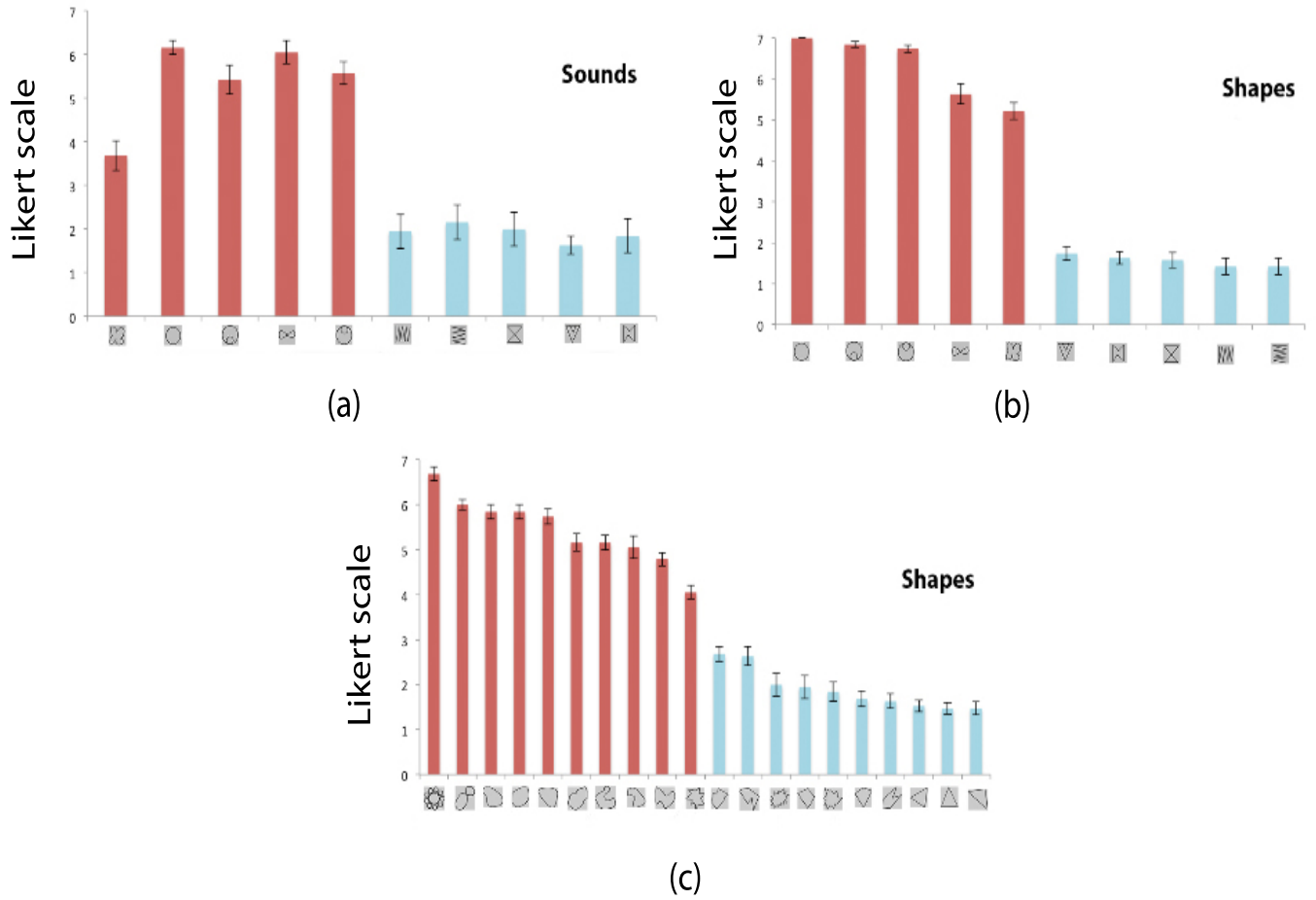


Figure C.3.: Mean ratings (M) and standard errors (SE) obtained from a Likert scale (1-totally sharp; 7-totally round) for each (a) action sound with a two maxima structure (b) action shape (c) and sound symbolic shapes. Red columns represent round stimuli and blue sharp stimuli.

Correlation pairs	Steiger's Z	p-value
SoSyAction vs. SoSyAnimals	1.67	0.04
SoSyAction vs. ActionAnimals	0.55	0.29
SoSyCrossed1 vs. Crossed1Animals	3.38	0.004
SoSyCrossed1 vs. SoSyAnimals	2.62	0.004
ActionCrossed1 vs. ActionAnimals	0.84	0.19
ActionCrossed1 vs. Crossed1Animals	2.74	0.003
SoSyCrossed2 vs. SoSyAnimals	1.94	0.02
SoSyCrossed2 vs. Crossed2Animals	0.92	0.17
Crossed1Crossed2 vs. Crossed1Animals	2.37	0.008
Crossed1Crossed2 vs. Crossed2Animals	0.69	0.24
ActionCrossed2 vs. ActionAnimals	3.40	0.0003
ActionCrossed2 vs. Crossed2Animals	3.6	0.0002

Table C.1.: One-tailed Steiger's z test (Steiger, 1980) was used to compare Spearman's correlation coefficients using the package cocor in R (Diedenhofen, 2016). Correlation pairs are depicted between the SoSy, Action and the Crossed conditions against the control Animal task. Steiger's Z scores are shown in the middle and p-values in the right column. P-values in bold were significant after controlling for multiple comparisons testing with Bonferroni correction (adjusted threshold $p = 0.05/12 = 0.004$).











				
1686ms	1573ms	1428ms	1945ms	1718ms
				
2409ms	2351ms	1521ms	1710ms	2400ms

Table C.2.: The 10 final shapes selected from the hand drawings. Durations required to produce each drawing are annotated below them, corresponding to their action sounds.



Table C.3.: Colored animal pictures used in the control 2AFC condition in Study 1.

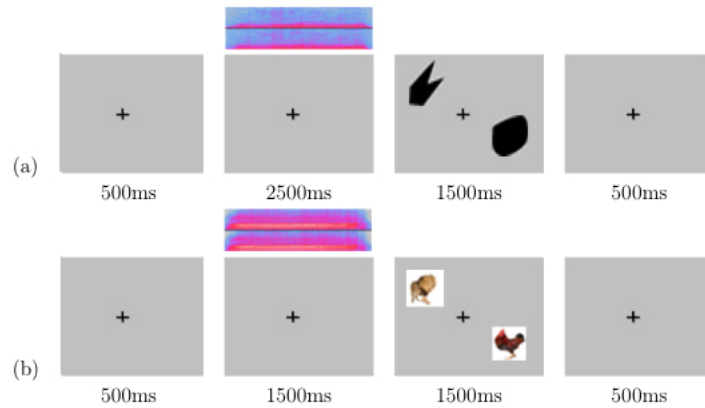


Figure C.4.: Experimental design for the two conditions a) action sound condition, b) animal control condition.

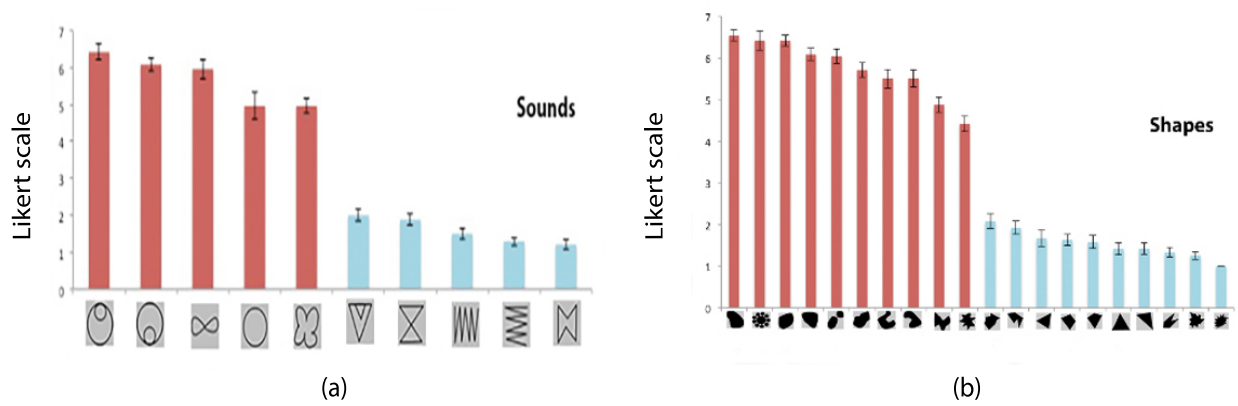


Figure C.5.: Mean ratings (M) and standard errors (SE) obtained from a Likert scale (1-totally sharp; 7-totally round) for each (a) action sound (b) for sound symbolic shapes. Red columns represent round stimuli and blue columns sharp stimuli.



Table C.4.: Blurred animal pictures used in the control 2AFC condition of Study 2.

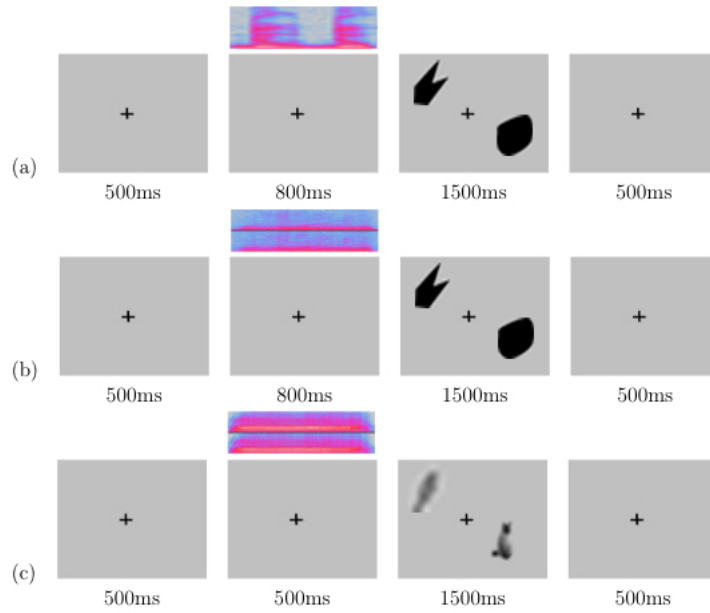


Figure C.6.: Experimental design for the three different conditions a) sound symbolic condition, b) action sound and c) animal control condition for the Study 2.

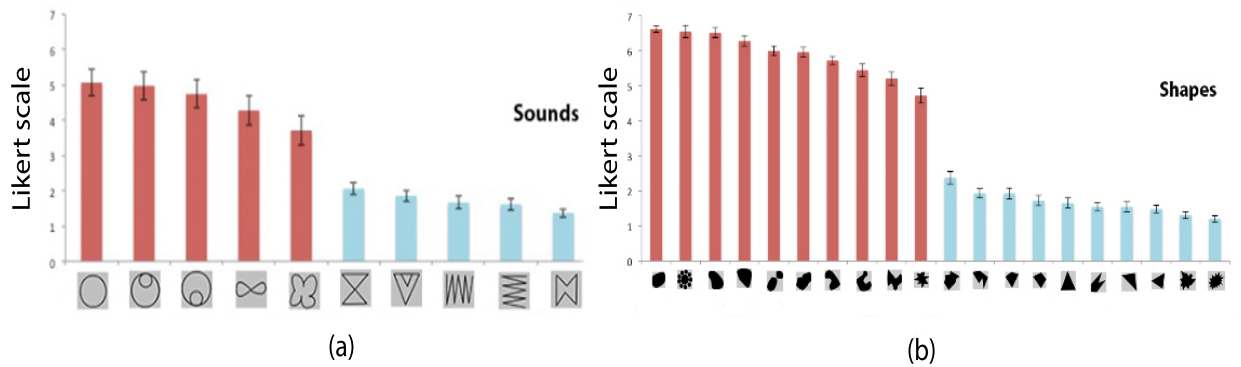


Figure C.7.: Mean ratings (M) and standard errors (SE) obtained from a Likert scale (1-totally sharp; 7-totally round) for each (a) action sound with a duration of 700 ms and (b) for sound symbolic shapes. Red columns represent round stimuli and blue columns sharp stimuli.

List of publications

This dissertation is based on the following publications:

Margiotoudi, K., & Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports*, 10(1), 1-13.

Margiotoudi, K., Allritz, M., Bohn, M., and Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific Reports*, 9(1), 1-12.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst habe und ausschließlich die angegebenen Hilfsmittel verwendet habe. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Konstantina Margiotoudi, Berlin, 2020