LUND UNIVERSITY

Referential iconicity in music and speech within and across sensory modalities

Verónica Giraldo

Supervisor: Prof. Jordan Zlatev

Centre for Language and Literature, Lund University

MA in Language and Linguistics, Cognitive Semiotics

SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

August 2018

ABSTRACT

Musical meaning is multifaceted: both highly sensory and yet often abstract, able to cross cultural boundaries and yet embedded in specific traditions. For the most part it is not denotational (Monelle, 1991). Nevertheless, in "programmatic music", musical themes are intended to refer to worldly objects and events on the basis of iconic (and indexical) grounds. Such non-arbitrariness of the sound-sign (Sonesson, 2013) appears to apply to speech as well, where research has established that the iconicity in question is subtle, but systematic enough to be detectible by both adults and children (Ahlner and Zlatev, 2010; Imai and Kita, 2014). Very often, it operates across sensory modalities, so that for example a sound form like *lulu* is linked to round shapes, while *titti* is associated with sharp and hard ones.

This thesis investigates how referential iconicity in speech operates in relation to music, taking into account different kinds of iconicity, unimodality and cross-modality and finally cultural background. To address these aspects, an experiment in which 21 Swedish and 21 Chinese native speakers had to match musical fragments or spoken word-forms to referents (represented by schematic pictures) was designed. It included two different conditions. In one there were two sound-stimuli and two referents (*more contrastive*). In the other, a single sound-stimulus was to be matched to one of four alternative referents (*less contrastive*).

The results showed that there was no significant difference between the overall results for music and linguistic tasks, indicating that the psychological, interpretive processes involved are not limited to a single cognitive domain, or semiotic system. As expected, the more contrastive condition was easier for both groups, showing that cultural background played little role for making the appropriate cross-modal mapping when the choice was so constrained. Finally, the fact that participants performed significantly better in more-contrastive tasks than less-contrastive, whilst performing above significant chance in both conditions serves as a clear indicator that interpreting referential music in music and speech sounds involves a combination of primary and secondary iconicity (Sonesson 1997), with a considerable role for the latter.

Keywords: Cognitive semiotics, iconicity, ideophones, multimodality, music, music cultures, semiotics, semiotics systems, signs, sound symbolism, primary and secondary iconicity, unimodality.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor Jordan Zlatev for his amazing guidance and patience, not only during the process of writing this thesis but during the past three years of teaching. He has really showed what the joy of teaching looks like, and his encouragement to keep searching and exploring the relationship between language and music has been really inspiring. His pedagogical methods have helped me navigate through the complicated terminology and methodologies, teaching me to question and challenge in a constructive way what is presented to me, which are tools I will most certainly be able to continue using in the future.

A special thank you to Victor Bogren and He Xiaoyan (何晓艳) for helping me with the translations and recordings of the stimuli to Swedish and Chinese, and to Irene Perez Zumalde for assessing me with the Basque stimuli, and for recording the fictive words and ideophones in Basque. Thank you to Maximilian Roszko for all of his help and advice when programming the experiment, and to Peer Christensen and Joost van de Wiejer for their help with the statistical analysis. Thank you to Jonas Mellström for helping me with an extra laptop for conducting the experiment. Also, a big thank you to Georgios Stampoulidis for all of his insights and constant help during these past months. I would also like to thank all of the kind people that participated in my experiment!

I want to specially thank Martin Sundby for being there so unconditionally, for always taking the time to listen to me, to debate and help me clarify ideas, even in moments where you were as stressed with your own thesis. Thank you for always encouraging me to do better and keep working hard. It has been an absolute gift having you by my side during this process. I would also like to thank my mom for always being supportive and patient, and for listening to me, even when you don’t really get what I’m talking about.

Thank you to my dear friends Frida and Clara for their constant support and for taking their time to take the experiment at its early stages and giving me great feedback and finally, thank you to my family in LA that were incredibly encouraging during the last stages of this whole process.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

The links between language and music have been pointed out on numerous occasions, from Rousseau (1781), proposing that that the origins of music and language are interconnected, suggesting that the pace and sounds of music were born along with the syllables of speech. Much more recent studies, such as Patel (2008) have given support for such speculations.

> The central role of music and language in human existence and the fact that both involve complex and meaningful sound sequences naturally invite comparison between the two domains (ibid: 3).

The task of investigating the many possible similarities, as well as differences, between language and music is ample, and insurmountable in only one study. Whilst trying to narrow down this wide umbrella of research possibilities, a particular aspect of music was drawn to my attention: the fact that a large number of musical pieces aim to resemble objects in the world, as for example Vivaldi's renowned *The Four Seasons*, or Saint-Saëns' *Le Carnaval des Animaux*.

This thesis investigates in what ways phrases in (instrumental) music and linguistic sound forms bear resemblance, or iconicity, to things in the world. Given that it might be difficult to conceptualize how music could be referential, the aim here is to investigate how sound iconicity in language operates in relation to music, within the framework of cognitive semiotics. Cognitive semiotics is a field "focused on the multifaceted phenomenon of meaning" (Zlatev, 2015: 1043), having the sign as one of the central kinds of meaning (making) that it studies. As a whole, the sign can be said to consist of three interacting entities: the *representamen,* the *object* and the *interpreter.* The *ground* is the relation between the object and the representamen, or the way in which "a representamen stands for an object" (Ahlner and Zlatev, 2010: 314). C.S. Peirce (2003) proposed a famous classification of signs based on the three different kinds of *ground* between representamen and object, following the interpretation of Sonesson (2007). First, an *iconic sign* is a sign whose motivated ground is based on similarity. Second, *indexical signs* have their ground based on contiguity or part-whole relationship and finally, *symbolic signs* have their ground in convention. In specific signs, like a picture or a gesture, the three kind of ground may combine, in different proportions (Jakobson, 1965).

Keeping in mind the different kinds of semiotic grounds within the iconic sign, Sonesson (1997) made the important subdivision between *primary* and *secondary iconicity*. In *primary iconicity,* the perception of the similarities between the representamen and object allows the interpreter to figure out the meaning of the sign, while in *secondary iconicity,* the interpreter first has to be informed of what something means in order to discern the similarity between representamen and object, or in other words, to establish the iconic ground. Ahlner and Zlatev (2010) further complement on this division, proposing that the interpretation of iconic signs can involve the *combination* of primary or secondary iconicity, analogously to the way different grounds may be involved in the interpretation process, as pointed out above.

In speech, many linguistic signs have clear resemblance to worldly objects or events (Jakobson, 1965; Dingemanese, 2012), and even though this relation is subtle, it is perceivable by both adults and children (Imai and Kita, 2014), even when these resemblances occur across sensory modalities (Ahlner and Zlatev, 2010). Taking into account that meaning in music is multifaceted (Monelle, 1991), in comparing speech with music it is opportune to focus on *programmatic music*, or musical melodies whose aim is to refer to extra-musical elements, or as mentioned earlier, worldly objects and events on the basis of iconic (and indexical) grounds, as some of the examples mentioned above. The Encyclopedia Britannica defines programmatic – or program – music as following

> Instrumental music that carries some extra-musical meaning, some "program" of literary idea, legend, scenic description, or personal drama. It is contrasted with so-called absolute, or abstract, music, in which artistic interest is supposedly confined to abstract constructions in sound. It has been stated that the concept of program music does not represent a genre in itself but rather is present in varying degrees in different works of music (Editors of Encyclopedia Britannica, 1998).

As can be seen from this definition, the understanding of what constitutes program(matic) music is somehow ambiguous. In a general sense, a piece of music can be regarded as programmatic if it involves references to extra-musical reality, as in the examples given above. In a more specific sense, it seems to involve the representation of some narrative or "program", which would need to be known in advance for this to be appreciated, thus amounting to secondary iconicity

From a cognitive semiotic perspective, the understanding of primary and secondary iconicity in music is still an unexplored field. For example, Sonesson (2009: 51) states that "as for iconicity in language and in music, it most of the time seems to be secondary". But then again, there is little empirical research that can corroborate that this is, indeed, the case. Many scholars have been intrigued by this aspect of music (Coker, 1972; Osmond-Smith, 1972; Monelle, 1991), but there are hardly any empirical investigations that delve into this matter. This opens a wide spectrum of research for those interested in further understanding of music, in particular as a semiotic system.

Many have shown that iconicity in speech may work across sensory modalities. A popular paradigm that has been used for language, and which is also used in this thesis, tests the associations between sound-forms and the shapes of objects and has shown that participants can map successfully between sound (e.g. *bouba* or *kiki*) and visual stimuli (shapes with soft or sharp edges). But, can such mappings also be found between music and referents?

On the basis of this brief background, we can then establish the following research questions, which this thesis addresses with the help of an experimental study:

- RQ 1. How does iconicity in music relate to speech iconicity?
- RQ 2. Is it easier for participants to recognize iconicity through unimodal (i.e. sound to sound mappings) rather than cross-modal (i.e. sound to movement) representamen-object mappings?
- RQ 3. Is iconicity in programmatic music and speech primary, secondary, or a combination of both?
- RQ 4. To what extent is referential iconicity in music and speech perceivable by members of different cultures?

The thesis is organized in six chapters. In Chapter 2 I will present the relevant theoretical background and conclude with general hypotheses. In Chapter 3 the methodology, alongside the stimuli selection and choice of participants in the experiment will be presented, finalizing with specific hypotheses. Chapter 4 reports the results and in Chapter 5 a discussion of the results in relation to the general hypotheses is provided. Finally, Chapter 6 summarizes and gives the conclusions of the thesis in relation to the research questions.

CHAPTER 2. THEORETICAL BACKGROUND

2.1 Introduction

Linguists and musicologists have analyzed music from various perspectives (Piaget, 1971; Coker, 1972; Lerdahl and Jackendoff, 1983; Nattiez, 1987, Monelle, 1991), but there is little research concerning referential iconicity in music. Furthermore, as pointed out in the introduction, there is still much more to investigate within the field of iconicity in speech, which will only contribute to those researching within this field (e.g. Ahlner and Zlatev, 2010; Dingemanese, 2012; Ibarretxe-Antuñano, 2017). Cognitive semiotics allows for the integration of various fields and concepts, but its notions of the sign, iconicity and cross-modal iconicity, are the most relevant for current purposes.

In this chapter, the theoretical background needed to understand the framework for this thesis, is presented with the aim of clarifying if (and how) iconicity in both music and speech may operate analogously or differently. I start by presenting the field of cognitive semiotics in more detail. I continue to elaborate on the structure of meaning of both music and language as semiotic systems, followed by the presentation of the phenomenon generally known as "sound symbolism", and the understanding of the concepts of primary and secondary iconicity. I conclude with a summary of the chapter, alongside presenting the general hypotheses of the thesis.

2.2 Cognitive semiotics

Cognitive semiotics may be defined as "the transdisciplinary study of meaning" (Zlatev, 2012: 2). Its aim is to provide new insights on human (and animal) meaning making, by incorporating methods and theories from the cognitive sciences on the one hand, and the humanities and semiotics on the other, with linguistics rather straddling this divide. Furthermore, an important facet to cognitive semiotics is the importance it attributes to empirical research, in combination with conceptual and phenomenological analysis (Zlatev, 2015). In other words, the use of empirical methods strengthens the understanding of complex theoretical concepts, while benefitting from the analysis of conceptual analyses (ibid). This interaction is often formulated in terms of the *conceptual-empirical loop*, shown in Figure 1.

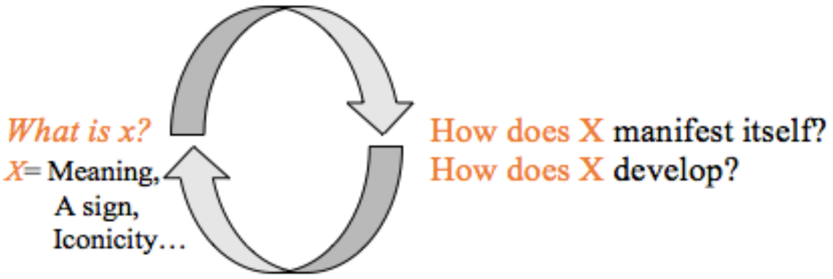


Figure 1. The conceptual-empirical loop (adapted from Zlatev, 2015: 46).

In the understanding of cognitive semiotics as a transdisciplinary field, it is thus important to think of the different methodologies used to achieve this conceptual-empirical loop. A vital tool for this is the field's pluralistic methodology, also known as "methodological triangulation", presented in Table 1 as applied in the present thesis, and elaborated in Chapter 3. Such triangulation implies the integration of methods, by grouping them on the basis of three different *perspectives* on the phenomenon in question: first-person, second-person and third-person (Zlatev 2009, 2012, 2015).

Table 1. Methodological triangulation applied to the phenomenon studied in this thesis. (Adapted from Zlatev 2012: 15)

Perspective	**Methods**	**Applied to (in this study)**
First-person **("subjective")**	* Conceptual analysis * Musical and linguistic intuitions * Introspection	* Perception of iconic signs in both language and music * Intuitions in choosing and analyzing stimuli
Second-Person **("intersubjective")**	* Interviews	*Interaction with participants
Third-Person **("objective")**	*Detached observation * Experimentation	* Analysis of participants' responses to stimuli consisting of music, speech, written words and images

Semiotics as such can be understood as the "interdisciplinary field investigating different systems of meaning-making, such as visual representations, speech and music" (Zlatev et al, 2017:463), where the understanding of the concept of *sign* is vital. According to Peirce's influential understanding of the sign, it may be characterized as follows, involving concepts that were given in Chapter 1.

> [a] sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. … The sign stands for something, its *object*. It stands for that object, not in all respects, but in reference to a sort of idea which I have sometimes called the *ground* of the representamen (Peirce, 2003: 106).

Using and interpreting a sign gives rise to (a kind of) *semiosis,* or meaning-making. Especially within cognitive semiotics, the user/interpreter is construed as a conscious agent who makes and understands the relation between representamen and object. Furthermore, it is necessary to stress that the *object* and the *representamen* are not related in all possible ways, but only in some, and consequentially, result in different kinds of *object-representamen* relationships, called grounds (Sonesson, 2007). The sign, understood in both Peircian and cognitive semiotic terms, can be illustrated as in Figure 2.

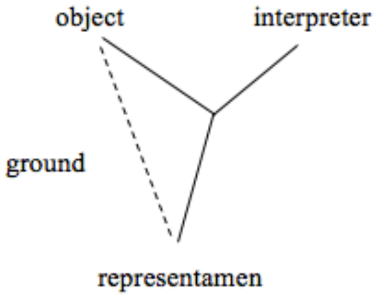


Figure 2. Graphic illustration of the sign and its interacting components (adapted from Ahlner and Zlatev 2010: 314).[1]

As stated in the introduction, different kinds of signs can be classified on the basis of the *predominant* kind of ground between representamen and object, the latter either in the

[1] It may be noted that in language, this model applies above all to what will be described in section 2.3.1, as "categorematic expressions".

world or in imagination. In the *iconic sign,* its ground is that of *similarity.* An example could be a drawing of a cat, as the image itself resembles an actual or imaginary cat (see section 2.4.2 for a discussion of imagistic and diagrammatic iconicity). Such signs can also be found in music, as for example an "imitation" of a bird singing by an orchestral instrument or voice (Monelle, 1991). The second motivated ground is that of *indexicality,* where "the ground is not based on similarity, but on contiguity in time and space" (Ahlner and Zlatev, 2010: 314), as well as part-whole relationships. An example of an indexical sign is a dog's barking, which can be associated with possible threat. The third kind of semiotic ground identified is *symbolicity,* which, unlike the first two, is based solely on general cultural agreement, by "virtue of law" (Peirce 2003: 111). An example of this is a currency sign where its meaning is generated by conventionality, or the "common knowledge shared by the speakers of the respective language" (ibid: 309. See also Zlatev, 2007; Itkonen, 2008). It is important to highlight, again, that signs typically combine all three grounds, or as stated by Jakobson (1965: 26):

> It is not the presence or absence of similarity or contiguity between signans and signatum, nor the habitual [conventional] connection between both constituents underlies the division of signs into icons, indices and symbols, but barely the predominance of one of these factors over the others.[2]

2.3 Structure and meaning of language and music as semiotic systems

Signs do not occur in isolation, but relate to other signs, in more or less complex relations, to form *semiotic systems*, allowing the expression of composite contents, such as stories (Louhema, 2018). Language is the best-studied such system, but gestures and depiction are two other universal systems for human meaning-making, and the three often combine in *polysemiotic communication* (Zlatev, in press). In the present section, I discuss some of the features of the semiotic systems of language and music.

2.3.1 Language

Zlatev (2009: 186) defines language as a "conventional-normative semiotic system for communication and thought". This can take different forms, depending on the media used:

[2] In this citation, the "signans" corresponds to representamen, and the "signatum" to the object.

speech, writing and the body (in the case of the "signed languages"). In this particular thesis, however, I focus mainly on speech.

One way to understand the inter-sign relations of language is in terms of syntagmatic and paradigmatic relations, following structural semiotics (Saussure, 1959 [1916]). *Paradigmatic relations* are the "vertical axis" of language, where signs can be exchanged or substituted without tampering with the grammatical structure. The *syntagmatic* plane "retains language's commitment to the flow of time and depends on the gradual unfolding of linguistic meaning during the 'performance' of a given utterance" (Agawu, 1991: 16). This can be further connected to the concepts of categorematic and syncategorematic expressions. It is necessary to highlight that not all linguistic signs correspond to the definition of the sign provided earlier, as not all represent an object (or signatum). These signs are called *syncategorematic,* and an example of these are prepositions, conjunctions, inflexions, prefixes, etc. (see Bundagaard, 2010). That is, these are morphemes that acquire meaning only when combined with other words. On the other hand, words that clearly stand in a semiotically grounded relation to an object, such as nouns, verbs or adjectives are called *categorematic*. The latter clearly correspond to the definition of the sign presented in section 2.2.

Ferdinand de Saussure was influential in the fields of linguistic structuralism and semiotics. His interpretation of the "linguistic sign" differs from the one here adopted, as he defined it as the linkage between the *signifier,* or the linguistic expression and the *signified,* which is the concept, in relation to other signs in the language. Saussure stressed that the fundamental basis of the relationship between the two is that of *arbitrariness* (Saussure 1959[1916]), usually understood to mean that there is no motivated linkage between the two entities that compose the linguistic sign (see Zlatev, 2014). This notion of the linguistic sign being fundamentally arbitrary has been criticized by many, and particular by Jakobson (1995), as it downplays the co-existence of iconic and indexical grounds in many linguistic expressions, as pointed out above.

In his general theory of human cognitive-semiotic evolution (see Table 2), Donald (1991; 2007) proposed three evolutionary transitions in human cognitive origins. According to this theory, the Mimetic stage, where "[mimetic] skills or mimesis rest on the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic" (Donald, 1991: 171), involving "non-verbal action modelling" (ibid: 218), allowed for shared intention, improved gesture skills, and for the development of representational culture. In other words, the first signs were produced through mimesis, which was further elaborated into the semiotic systems of gesture, depiction, and language (Zlatev, in press).

Table 2. Stages in human cognitive semiotics evolution in Donald's theory (Donald, 2007: 218)

Stage	**Species/Period**	**Novel forms of representation**	**Manifest change**	**Cognitive governance**
EPISODIC	Primate	Complex episodic event-perceptions	Improvised self-awareness and event-sensibility	Episodic and reactive, limited voluntary expressive morphology
MIMETIC (1st transition)	Early hominids, peaking in *H. erectus:* 4M-0.4Mya	Non-verbal action modelling	Revolution in skill, gesture (including vocal), nonverbal communication, shared intention	Mimetic increased variability of custom cultural "archetypes"
MYTHIC (2nd transition)	Sapient humans, peaking in *H. sapiens sapiens:* 0.5 Mya-present	Linguistic modelling	High-speed phonology, oral language, oral social record	Lexical invention, narrative thought, mythic framework of governance
THEORETIC (3rd transition)	Recent sapient cultures.	Extensive external symbolization, both verbal and nonverbal	Formalisms, large scale theoretic artifacts and massive external memory storage	Institutionalized paradigmatic thought and invention

Since language, on this account, is built upon a semiotic system that is completely non-arbitrary, it implies that, up to a considerable degree, language is also a non-arbitrary system, and quite possibly iconicity can be understood as a universal feature of language (Vigliocco, Perniss, and Vinson, 2014). The claim is that iconicity in language is not only an evolutionary relic, but is psychologically real, and in order to evaluate this, it is necessary to do empirical tests, as performed in this thesis.

Further, the fundamentally non-arbitrary nature of language can be reconciled with the notion of it being conventional, once we remember that the three kinds of semiotic grounds intermix (Jakobson, 1965; Zlatev, 2014). Lewis (1969: 76) gives a well-known definition for this *conventionality*:

> A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S is a convention if and only if, in any instance of S among members of P,
>
> (1) everyone conforms to R;
>
> (2) everyone expects everyone else to conform to R;
>
> (3) everyone prefers to conform to R on condition that the others do, since S is a coordination problem and uniform conformity to R is a proper coordination equilibrium in S.

This definition is, however, too broad for the purpose of this thesis, as it can be applied to signs and non-signs (e.g. social conventions). Zlatev (2014: 200) presents a more narrowed-down definition that applies to signs, where he proposes that the term "arbitrary" in Saussure's classical work is ambiguous between "*conventional* (= socially shared) and *unmotivated* (= lacking any iconicity or indexicality between expression and content)". Further, Zlatev proposes that the first sense should be:

> … taken as central, and as the true "design feature" of language, then it is fully possible to combine it with various forms of non-arbitrariness in the first sense, i.e. signs that are based on intermixtures of iconic, indexical and symbolic grounds (ibid: 208).

This discussion supports the understanding that very few, if any, linguistic signs are entirety unmotivated or, in other words, where their ground is conventional only. This conclusion meshes well with theories of "sound symbolism", which will be deferred for section 2.4 below.

2.3.2 Music

> The semiotics of C S Peirce, dating from the turn of the century, seemed relevant to music because Peirce placed iconic and indexical signs alongside the linguistic variety which he called symbolic. This meant that many things, both in nature and culture, could be considered signs by virtue of their similarity to other things, or their habitual association or contiguity with their objects. The 'arbitrary' nature of linguistic signs was difficult to envisage in music.
>
> (Monelle, 1991: 30)

As stated in the introduction, comparisons between linguistic and musical aspects have been explored by many throughout time. Under the understanding that music is systematic in its structure, many researchers have delved into the study of the "syntactic" structure of music, such as Bernstein (1976) and Keiler (1978, 1981) who pioneered this field. One of the best-known works in the perspective is Lerdahl and Jackendoff's (1983) *A Generative Theory of Tonal Music,* where they set out to explore the grammar of music within a generative linguistics theoretical framework. Treating the structure of music, as well as language, in purely syntactic terms, however, misses the obvious fact that both are crucially systems of *meaning*, even if the meaning conveyed in music may differ from meaning in language.

Hence, this section focuses on music as a semiotic system. Taking into account that one of the dominant functions of language is to state "propositions" about the world, it can be understood that the main difference between language and music as semiotic systems is that music has predominantly *non-referential* meaning, while linguistic meaning is both *referential* and *denotational*. These terms are sometimes used synonymously, but, as Saeed (2016) explains, some writers such as Lyons (1977, 1968) and Saeed himself, make a distinction between the two terms, where:

> … to *denote* is used for the relationship between a linguistic expression and the world, while *refer* is used for the action of a speaker out entities in the world. In other words, referring is what speakers do, while denoting is a property of words (Saeed, 2016: 22-23).

In this thesis, I follow this distinction, but mostly use the term *reference*, since as explained in Section 2.2, a sign is never "self-interpreting", but requires an interpreter, i.e. a conscious (human) being. Many argue that that this type of meaning cannot be obtained from music since no specific semantic reference can be ascribed to musical terms (Kivy, 2002, 2007), but as discussed later in this section, there are instances in music that question this claim, namely in the cases of programmatic music, where referential iconicity is present. While it should be acknowledged that each musical motive or fragment relies on context for its signification (Monelle, 1991: 15; Patel, 2008: 327) is can be argued that:

> … lacking specificity of semantic reference is not the same as being utterly devoid of referential power. Instrumental music lacks specific semantic *content,* but it can at times suggest semantic *concepts.* Furthermore, it can do this with some consistency in terms of the concepts activated in the minds of listeners within a culture (Patel, 2008: 328).

Hacohen and Wagner (1997) performed an empirical study set to demonstrate that the leitmotifs in Richard Wagner's *Ring* cycle operas "bear inherent meaning" (ibid: 445). This study consisted of two parts. In the first, they provided a total of 174 listeners with a selection of nine Wagnerian leitmotifs, alongside seven semantic scales. The participants were thus asked to rate each leitmotif on the seven scales, finding that the leitmotifs were systematically arranged in three clusters: a friendly, a violent and a dreary cluster. This provided strong motivations for the second task, where 102 of the participants, based on the same selection of nine leitmotifs, were first asked to imagine the piece as a musical theme in a movie, and then asked to provide a name for the movie that would go along with the music. In this task, no constraints or categories were provided. The results showed that there was an overlapping consistency in the titles given to each leitmotif, demonstrating that studies that concern musical semantics should not only rely on setting where the participants are given constraints and context, but should be also allowed to make free associations.

An important characteristic found in both language and music as semiotic systems and returning to the understanding of meaning in music as non-referential, is the implementation of intra-musical meaning. In other words, music can be referential to other genres and pieces of music or meaning within the musical realm. An example of this are the characteristics that delimit each music genre, or the categorization of a musical piece, according to similarity in their characteristics, very often delimited by time epochs. So, in this sense, a salsa piano player in the 1970s could actually play a romantic piece (dating from the XIX century) adapting the music to the rhythms and styles characteristic of salsa.[3] This can be understood as parallel to *dialogicity* in language, or as the continuous connection in which new statements within all possible levels of expression in language presuppose previous utterances (Bakhtin, 1981). So, in this line of thought, and keeping in mind that, unlike language, music is not always aimed to represent physical worldly objects, does this mean that music lacks the properties for referential iconicity?

As mentioned at the introductory chapter, there is indeed music whose aim is to communicate about objects in the world. This type of music is also known as *programmatic music.* Some examples of this type of music, apart from *Le Carnaval des Animaux,* mentioned earlier are works by Rimsky-Korsakov, known as a programmatic music composer, who created pieces such as *Flight of the Bumblebee* (1889-1900) and *Scheherazade* (1888), Vivaldi's well-known *Four Seasons* (1721), and *Peter and the Wolf* (1936), by Sergei Prokofiev, which is the piece used in the experimental study described in this thesis, and which is further discussed in Chapter 3.

[3] For example, *Sonido Bestial* (1971) by Richie Ray and Bobby Cruz.

A common question posed regarding programmatic music is whether it can convey referential meaning across cultures. The music pieces that were mentioned above belong to Western classical music tradition, and Tien (2015: 1) argues that:

> … across cultures, music seems to have been well and truly embraced, though it is highly contextualized in the individual culture and hence culture-unique i.e. the way in which the people of a culture practice music may greatly differ from those of other cultures.

This leads to the idea of music being a *conventional* semiotic system, not unlike language (see above). The Western classical-music art tradition, encompassing a broad number of musical periods and styles that vary according to their historical period, ranges from the year 500 to the present. A dominant genre originally in Europe, it has expanded world-wide due to colonization and cultural influence. An interesting aspect of the dominance of classical music is the influence it brought to European and non-European folk music. For example, regardless of classical music being remarkably different from a Colombian *bambuco,* its influence is reflected through musical notation and instrumentation.[4]

The use of European-original instruments institutes the same notation found in European classical music, brought to America by Europeans during the colonization period. This type of influence can be further perceived through a large range of folk music genres across the world, such as jazz, salsa, and a number of traditional folk music across Europe and the Americas. Such influence is consistent with Monelle's (1991: 276) statement:

> Music carries a social context: it starts and develops amongst a whole community. The aspects of intonation are tied to a historical, social and cultural period. These define the means of musical expression and the selection and interconnection of musical elements.

At the same time, even though many music cultures around the world have been influenced by European classical music, this does not mean that they are considered as part of classical music, nor does it mean that members of the cultures that create and perform this type of music feel that classical music reflects their musical culture. These music cultures have drastically different rhythms and intonations, which have been shaped by the local people, their culture, and the historical epoch. These are characteristics of the complex process of

[4] The *bambuco* is a traditional music originating from the Andes region in Colombia that has been inspired by the rhythms of waltzes and polskas, and its main instruments are the piano, vocals, guitar, and a series of variations of guitars namely *tiple*, *bandola* and *requinto*.

musical evolution. Further, it is clear that not all musical genres have been influenced by classical music. Some clear examples of this are traditional African rhythms and intonations, Arabic music and Chinese classical music, which are genres that have developed their own instrumentation, notation systems (if any), and intonations.

Especially relevant in the present context is Chinese classical music, which has quite a different approach to music in general, from its origins (Shen, 1991; Tien, 2015), including pitch and harmony. In Western music, tonality is a fundamental arrangement for the composition of most classical music, and most non-classical music alike.[5] Tonality is basically the systematic arrangement of musical pitches or chords, grounded in a hierarchical structure, where different pitches possess larger stabilities than others. This can be reflected in the use of scales, which is a construction of seven tones arranged according to their resonance, where the most stable pitch is called the tonic, and gives rise to different keys. There is a variety of scales, fluctuating on the resonance distance between the tones, but this pattern provides a fundamental structure for Western classical music. This, however, does not quite apply to Chinese classical music, according to Shen (2000: 22):

> For some time, the outside world focused on China's musical scale, thus the pentatonic studies and the pentatonic competitions, which turned up nothing. As it turns out, scale is somewhat of a western concept. The Chinese went by harmony, thus several songs with distinctly different melodic lines may sound totally different to the outside world, but they sound the same to the Chinese, because, as a thorough analyst will tell you, their melodic lines were formed by members of the very same 'series of harmonies' […] Thus the Chinese hears harmonies rather than melodies.

Even though the examples mentioned here are only some out of the many characteristics that differentiate Western classical music and Chinese classical music, this indicates that the Western harmonic and theoretical system cannot be fully applied to Chinese classical music, and that this could even be reflected in the way we perceive and listen to the music itself, as "acoustics is cultural" (Shen, 1991: 2).

[5] Note, however, that not all Western music is necessary tonal, such as the subgenre of *atonal music* that emerged at the beginning of the XX century.

2.4 “Sound symbolism” and iconicity

The understanding that language as a fundamentally motivated system, where the three kinds of semiotic grounds intermix, has been theorized in alternative terms under the heading of “sound symbolism”*,* or “the idea of the existence of a motivated, non-arbitrary relation between the sound patterns and the meaning of words” (Johansson and Zlatev, 2013: 3).[6] An obvious example of this would be *onomatopoeia,* but the phenomenon goes far beyond onomatopoeic expressions in English like *bam, splash, woof* and *meow* (Ahlner and Zlatev, 2010). Such expressions are only the most commonplace example of *ideophones,* defined as:

> … marked words depictive of sensory imagery found in many of the world’s languages […] they are WORDS, that is, conventionalized items with specifiable meaning, as opposed to ‘simple sounds’ (Dingemanse 2012: 654-655).

Ideophones were until recently marginalized given the significant weight of Saussure’s thesis of “the arbitrariness of the linguistic sign” (see section 2.3.1). Nevertheless, studies such as those of Dingemanse (2012) and Ibarretxe (2017) have contributed to the rehabilitation of the study of ideophones in linguistics. In a more recent article, Dingemanse (2018) pointed out that studying ideophones is fundamental topic to the linguistic inquiry of “iconicity and ‘sound symbolism’ in natural language, the typology of property-denoting expressions, and ideologies about what constituted ‘proper’ languages” (ibid: 9).

Dingemanse (2012) provides a review of research on ideophones, covering a large group of languages where this phenomenon is found. He provides not only a thorough description of their form, use and categorizations, but a tentative *implicational hierarchy* for ideophones, as shown in Figure 3, where each category in the hierarchy presupposes the previous one.

SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES

Figure 3. Tentative implicational hierarchy for the constitution of ideophone systems Dingemanse (2012: 663).

[6] It is necessary to make a brief clarification regarding the some of the terminology used here. The term *“*sound symbolism*”* is employed in scare quotes, given the possible confusion with the ground for the *symbolic sign* presented in section 2.2

An example of a sound-to-sound ideophone would be onomatopoetic words, as shown in (1) for English and (2) for Basque (Ibarretxe, 2017).

(1) *Bam*
Crash

(2) *Kluka-kluka* 'to drink in gulps'
Plisti-plasta 'to splash'

Following from the implicational hierarchy in Figure 3 above, if ideophones for movement are found in a language, that language will also have ideophones for sound. Subsequently, if a language has ideophones for visual patterns, it will also have ideophones for both movement and sound, and so on (Dingemanse, 2012). This means that the most common type, are sound-to-sound ideophones, or onomatopoeias. It is also apposite to highlight conventionality as one of the main characteristics of ideophones, which as stated in section 2.3.1 can well co-exist with their iconicity.

The understanding of only the first subcase of ideophones, or sound-to-sound mappings, leads us to the concept of *unimodality*, or mappings that stay within the same sensory modality (in this case, hearing). The rest of the subcases naturally imply that they are *cross-modal*, or mappings that go across sensory modalities, e.g. from sound to movement (kinesthesia) or sound to (visual or tactile) shape (Ahlner and Zlatev, 2010). This aspect is further connected to music, where both *unimodal* and *cross-modal* mappings occur, where a unimodal mapping implies resemblance to the sound of something in the world. A cross-modal mapping could be of the music bearing resemblance to the movement of an object, which can often be achieved by the quick gradual rising and dropping of musical notes. I will delve more into these aspects in section 3.2.1, as these are relevant for the empirical study, and for the second research question (see Chapter 1).

Furthermore, as pointed out by Ibarretxe (2017), a common, though not compulsory, feature in ideophones is that of reduplication, or the "full or partial repetition of the morph" (ibid: 202). This has different functions, in the sense that "we can use reduplication to suggest repetition and the vowel space to suggest different grades of intensity" (Dingemanse, 2012: 663).

This aspect can be related to a possible motivation for some aspects of "sound symbolism", known as the *frequency code,* where high frequencies are said to be associated with small things, whereas low frequencies are associated with big things (Ohala, 1994). Ohala (1997: 2) claims that:

> […] there is abundant literature, some of it experimental or statistical, supporting the idea that across languages there are phonetically natural classes of speech sounds that are systematically associated with expressions of size.

Johansson and Zlatev (2013) performed a study where they investigated different possible motivations for spatial demonstratives in the world's languages, arguing that the vowels and consonants in demonstrative pronouns could have a connection to proximity or distance, with a sample of 101 languages. In English, for example the front, closed vowel [i] in *this,* would map to closeness, whereas the open back vowel [a] in *that* should be more distant. Their findings show that in 56% of their cases their prediction of frequency-vowel correlations was supported, in 22% of the cases showed a reversed pattern and in 22% there was no difference.

But are such correlations functional for language users, rather than just "evolutionary relics"? The popular paradigm of matching non-words to objects (represented visually), suggests a positive answer to this question (Köhler, 1929; Ramachandran and Hubbard, 2000; Ahlner and Zlatev, 2010). Here, participants are typically provided with two shapes (one sharp, one rounded), and told that one was called e.g. *bouba,* and the other e.g. *kiki.* Following this, they are asked to state which of the shapes was *bouba,* and which one was *kiki.* Ramachandran and Hubbard (2000) found that in 95% of the cases people matched the non-word *bouba* to the rounded figure and the sharp one to *kiki*. Ahlner and Zlatev (2010) followed up on this study, by making different combinations and contrasts of vowels and consonants, where made-up words with the vowel /i/ and consonants /p/, /t/, /k/, were expected to have "sharper" features, while made-up words with the vowel /u/ and consonants /m/, /l/, /n/, would have "softer" features. They found that both vowels and consonants contributed to the appropriate mapping to sharp and round figures, and proposed that haptic sense, in the vocal tract, mediated between the visual and auditory modalities. In this way, the authors also linked the discussion of "sound symbolism" with cognitive semiotics, and in particular the notion of iconicity, discussed further below.

But first we must mention a possible downside to such cross-modal association type of research, as pointed out by Dingemanse (2018). Non-words like *bouba* and *kiki* are artificially created, rather than present in a "proper language" (ibid: 9) and may be unnatural and exaggerated. Thus, in order to truly understand "sound symbolism" in language, it would be pertinent to study ideophones, which are clear cases of iconicity that occur in natural language.

Given that the concept of iconicity can often be too broadly defined, it may help to turn once again to Peirce (1974[1931]), who distinguished three different types of iconicity: *imagistic*, *diagrammatic* and *metaphoric* (Devylder, 2018). Here, I will only focus on imagistic and diagrammatic, similarly to Jakobson (1965).

Imagistic iconicity implies a one-to-one pairing (Kwon, 2016; Devylder, 2018) between the representamina and the object. Kwon (2016) further states that in the case of "sound symbolism" imagistic iconicity "uses acoustic signals of speech sounds only to mimic acoustic phenomena" (ibid: 73). Diagrammatic iconicity is defined by Peirce (1974: 2.277) as icons "which represent the relations […] of the parts of one thing by analogous relations in their own parts". Devylder (2018) presents a very straightforward example of a diagrammatic icon, as shown in Figure*s* 4-6. If we are presented with imagistic icons of two crosses, a circle and a line (Figure *4),* they may be taken to resemble, for example, two instances of the letter X, a stick and a ball.

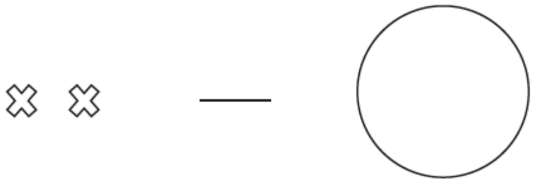


Figure 4. Imagistic iconic signs (Devylder, 2018)

However, if rearranged as shown in Figure *5,* it is clear that the placement and interaction between the signs project what could be understood as a diagrammatic icon of a human face (ibid: 322), given the shapes of each individual sign and the way they are organized in relation to one another.

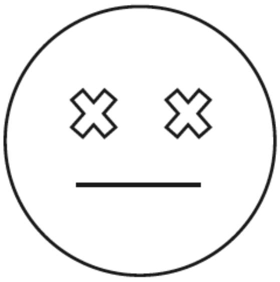


Figure 5. Diagrammatic icon of a human face

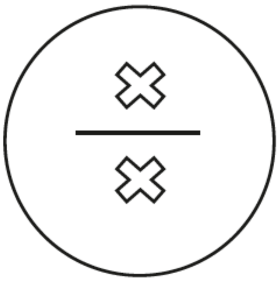


Figure 6. Lost diagrammatic icon of a human face

Obviously, this is an effect created by the interaction between the signs, which means that, if organized differently, as shown in Figure 6 (ibid: 323), the effect will most certainly not be the same. The nature of the diagrammatic icon is not a one-to-one mapping, but consists of a specific arrangement and interaction between the signs that allows for a successful resemblance to the referent.

Sonesson (1997) provided another important subdivision of the iconic sign in terms of *primary* and *secondary iconicity*. The first was defined as the case where the “perception of an iconic ground obtaining between two things is one of the reasons for positing the existence of a sign function joining two things together as expression and content” (Sonesson 1997: 741). In other words, it is the perception of the similarity between *object* and *representamen* that allows for the understanding of the sign. For example, an observer does not need much context to be able to perceive the image presented in Figure 7a as a picture of a lama, given that (s)he is familiar with what a lama in the real world looks like (whether they have seen one or not). *Secondary iconicity,* on the other hand, implies that the “knowledge about the existence of a sign function between two things functioning as expression and content is one of the reasons for the perception of an iconic ground between these same things” (ibid.). Here, the interpreter needs to know what a particular iconic sign represents in order to perceive the similarity in question, as in so-called “droodles”, as in Figure 7b*,* where we can either see a hat or, for those that are still child at heart a boa constrictor digesting an elephant (Saint-Exupéry, 1943).

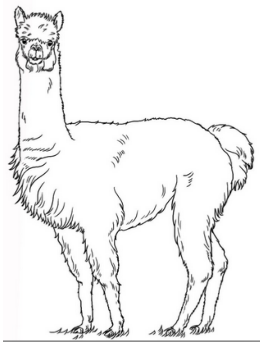


Figure 7a. Example of primary iconicity: the similarity of the image to lama leads to understanding the sign

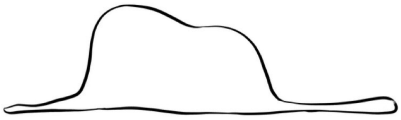


Figure 7b. Example of secondary iconicity: being told that this is "boa having swallowed an elephant", we can see the similarity

In their study of cross-modal iconicity in the context of their experiment in "sound symbolism" Ahlner and Zlatev (2010), propose that to solve the typical "bouba-kiki" task (see above), participants need first secondary iconicity (i.e. to be told that what they are given are two signs), but also primary iconicity, to be able to use the most "natural" matching between sound forms and shapes, in order to decide which is which. Thus, they argue that analogous to the way that the different grounds may combine in a specific sign, the interpretive process may involve a combination of both secondary and primary iconicity. As this would seem to apply to referential iconicity in music as well, this proposal serves as the basis for my third research question (see Chapter 3).

Specifically, Ahlner and Zlatev (2010) describe the combination of primary and secondary iconicity in a sequence of three steps, shown in Figure 8. In Step 1, the instructions inform the participant, or the interpreter, that there is a relationship between the each representamen (R1 and R2) and the objects (O1 and O2). In Step 2, the participant "discerns the composite analogous ground" (ibid: 319) between the objects and the representamina. This is a secondary iconicity element, since the participant knows already there is a sign relation and, on this basis, "looks" for the ground. Finally, in Step 3 the participant posits a specific sign relation (e.g. R1-O1 and R2-O2), thus creating two specific signs, and this is the primary iconicity element. From this we can gather that there is a combination of primary and secondary iconicity in solving the task, with a considerable role for secondary iconicity.

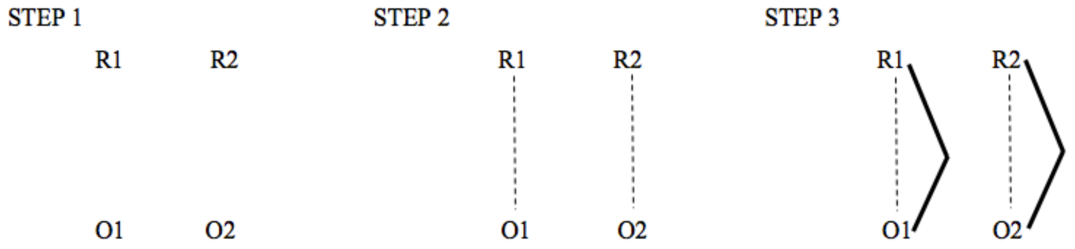


Figure 8. Matching representamina to objects by finding a composite analogous ground in more-contrastive tasks (Ahlner and Zlatev 2010: 319)

2.5 Summary and general hypotheses

In this chapter I presented some of the main characteristics of cognitive semiotics, features of language and music as non-arbitrary semiotic systems (i.e. signs in system-typical interrelations, with different kinds of objects and proportions of semiotic grounds), and finally a review of previous empirical research related to "sound symbolism" and different kinds of iconicity: unimodal and cross-modal, imagistic and diagrammatic, primary and secondary. On the basis of this background, we may state the following hypotheses:

- **Hypothesis 1**: It will be easier for a participant to recognize the referential object (in both speech and music) when representamen and object are in the same modality (unimodal iconicity) than in different modalities (cross-modal iconicity).
- **Hypothesis 2**: In both programmatic music and speech, a combination of primary and secondary iconicity is responsible for succeeding in the task stated in Hypothesis 1.
- **Hypothesis 3**: Given that there is a high degree of conventionality in music, linked to one's culture, there will be differences in how participants belonging to different musical traditions, will solve the task in Hypothesis 1, when the musical stimuli are taken from one of these traditions.

These hypotheses serve as the basis for the empirical study described in the following chapter and will be further specified after providing its design and methodology. I return to them in Chapter 5 again, as a way to structure the general discussion.

CHAPTER 3. METHOD

3.1 Introduction

To address the subject of referential iconicity in speech and music empirically, an experiment was conducted, where Swedish and Chinese participants were asked to match different representamina (musical and linguistic) to a number of objects shown on a computer screen. In line with the use of methodological triangulation (see section 2.2), the methodology for this thesis involves the first-person perspective in the use of intuition (my own, and that of consultants) for choosing and analyzing the materials. Aspects of the second-person perspective are utilized in interviews and interactions with participants and finally, the third-person perspective is used through the experiment itself, by observing and analyzing the participants' responses to the stimuli.

I begin by presenting the materials, followed by the design of the experiment. Ensuing, I present the participants who partook in the experiment followed by the procedure, finishing with a summary of the chapter, and specific hypotheses.

3.2 Materials

The following materials were used in the study, with representamina always in the audial modality, and objects in the visual modality: images or written words denoting the respective objects in the world.

3.2.1 Representamina (audial modality)

3.2.1.1 Music

The music stimuli used for the experiment were six excerpts of the musical piece *Peter and the Wolf*. This piece, composed in 1936 by Sergei Prokofiev, is a "symphonic fairy tale for children" (*Peter and the Wolf,* Wikipedia, March 19, 2018). In this piece, a narrator tells the story of Peter a young boy, who lives with his grandfather in the woods and sneaks out of his yard to play in the meadow, against his grandfather's will. Peter is accompanied by a duck, a bird and a cat, before they encounter a dangerous wolf. As it is a symphonic tale, the narration is complemented by the orchestra. However, in the introduction of the piece, the narrator discloses that each character in the story is represented by a different instrument in the orchestra, followed by a short introductory melody of each character performed by their respective instrument. This piece was chosen for this study as it is an opportune example of programmatic music, with clear and short representations of objects existing in the world.

The musical stimuli consisted of the six introductory melodies corresponding to six of the characters presented in the piece: Hunters, Bird, Cat, Duck, Wolf and Grandfather. Further details of each melody are presented in Table 3. The average duration of the six melodies was 14.6 seconds. Two of the melodies (Hunters, Bird) involved unimodal iconicity given that the music resembles the sounds each character makes; a sound to sound mapping. The others required cross-modal iconicity for their interpretation since there was no clear resemblance between the melodies and the sounds produced by the respective objects (in the semiotic sense). Peter's melody was not used, as its melody is represented by the whole strings section (violins, violas, cellos and double basses) of the orchestra, as opposed to only one instrument, and was used instead for recruitment purposes, this will be further discussed in section 3.4.

Character	**Instrument**	**Pitch Range**	**Duration**	**Iconicity**
Bird	Flute	E4-G6	13 seconds	Unimodal
Duck	Oboe	C4-D5	16 seconds	Cross-modal
Cat	Clarinet	G3-F4	10 seconds	Cross-modal
Hunter	Timpani	E2-C3	7 seconds	Unimodal
Wolf	French Horns	G2-F4	19 seconds	Cross-modal
Grandfather	Bassoon	B1-G3	23 seconds	Cross-modal

Table 3. Details of the musical stimuli

The data was also selected so as to test the "frequency code" motivation for "sound symbolism" (see section 2.4.1), according to which the pitch can reflect the size of an object, where higher pitch can be associated with smaller sized objects, and lower pitch can be associated with larger sized objects. Since each character was represented by a different musical instrument, as shown in Table 3, the pitch range of the latter can be represented using so-called scientific pitch notation (SPN), which combines musical pitch with musical notes (*Scientific Pitch Notation,* Wikipedia, May 05, 2018). A visualization is presented in Figure 9.

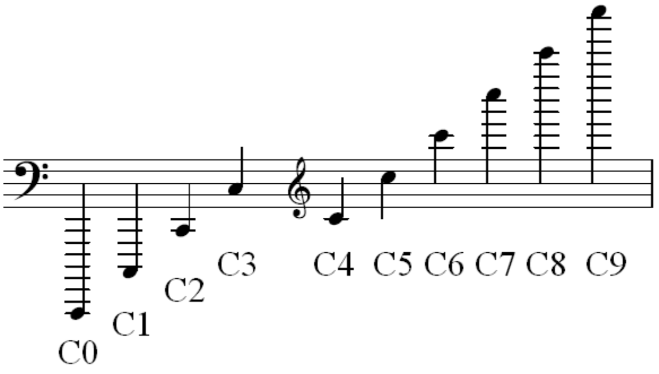


Figure 9. Scientific Pitch Notation in octaves of the note C.

Here we can see the ten possibilities for the note C, where there is a progression from low to high frequencies (from left to right). The distance between each C is of an octave, meaning that it is the same note, but with double its frequency. Between every C in the figure above, there are six notes, namely D, E, F, G, A and B. In Table 3, when referring to, for example, D3, it means that this is the first D placed between C3 and C4. The symbols 𝄢 and 𝄞 are used to indicate the pitch of written notes. The notes marked after the former symbol represent notes with lower pitch, whereas the notes marked after the latter represent notes with higher pitch. The pitch range provided in the third column in Table 3 denotes the pitch range of the instrument in each melody (which I gathered from the music scores of each individual melody), and not the instrument in general. Cells with higher color pigmentation represent instruments with lower frequency, while those with lighter pigmentation represent those instruments that have higher frequencies.

For the warmup task, a 22 second excerpt of the piece *Dance of Dragons* composed by Ramin Djawadi for the popular HBO series Game of Thrones, was used.

3.2.1.2 Ideophones

For the first linguistic task, a set of six words, specifically *ideophones,* in a language that was unfamiliar to all participants (Basque) was used. The ideophones were chosen from Ibarretxe's (2017) wide compilation of Basque ideophones, on the basis of three criteria. The first criterion was that the ideophone had to be reduplicated and denoted an action. The second was a consultation with two Basque native speakers, where they confirmed that the ideophones were unimodal (i.e. resembled the sounds made by the denoted actions), and recognizable (some ideophones in the compilation were outdated or very specific to geographical locations). The third was that the ideophone did not correspond to any word in the languages of the participants, Chinese and Swedish. All of the ideophones were recorded by a native speaker of Basque and are shown in alphabetical order in (3-8).

(3) *Draka-draka* 'horse galloping'

(4) *Grik-grak* 'to crackle'

(5) *Pil-pil* 'sound of boiling water'

(6) *Trinkili-trankala* 'move noisily, with difficulty'

(7) *Zirris-zarraz* 'sound of sawing'

(8) *Zorro-zorro* 'snoring'

As in the music tasks, three foils were employed, namely Swedish and Chinese translations of the English verbs shown in (9-11), see Appendix C.

(9) whisper

(10) splash

(11) groan

3.2.1.3 Fictive words

For the second linguistic task, six two-syllabic CVCV non-words were created. For this, the following criteria provided by Ahlner and Zlatev (2010: 324) was employed: (a) consonants were either voiceless obstruents [tʃ, t, k], or voiced sonorants [m, l, n]; (b) vowels were either front close unrounded [i, e] or back open [a, u].

This provided a basis for creating non-words that sounded either "soft and "round", or "sharp" and "pointy", based on intuition and "the 'synesthetic' properties that have been ascribed to sounds in previous studies" (ibid: 323-324). All of the non-words were audio recorded by the same speaker that recorded the ideophones, and all recordings had a duration of 1 second.

As with the ideophones, it was made sure that these forms did not match actual words in Swedish or Chinese. For example, having [mu] as a first syllable was avoided as it resembled the word *wood* in Chinese. Two L1 Chinese speakers and two L1 Swedish speakers (who did not participate in the main study) were asked to listen to the recordings of all words to make sure none of these had any meaning in either language. A list of all fictive words is provided in Table 4.

Table 4. Spoken stimuli for fictive words

"sharp" and "pointy"	**"soft" and "round"**
[keti]	[nalu]
[kitʃi]	[lulu]
[tike]	[lamu]

Two extra non-words [kling] and [klang] were presented in the warm-up task, where they were to be matched with the two policemen in the well-known Swedish children's books series *Pippi Långstrump.* These were likely to be familiar for the Swedish participants and less so for the Chinese, but since the purpose of the warm-up was to give each participant an understanding of the task, this difference was not judged to be problematic.

3.2.2 Objects (visual modality)

3.2.2.1 Images

The image-objects to be matched with music representamina were generic outlines, all without any coloring, of each respective animal or person. Given that the participant was not to see the representamen of each character more than once (see below), a set of six foils was employed. This group was designed to have a proportionate number of humans and animals. As a result, images of a squirrel, a cow, a pig, a bear, a ballerina and a Mexican luchador, (all following the same characteristics described above) were used. For the warmup, the target image was that of a dragon, and the foils consisted of a platypus, a donkey and a racoon.

All of the objects were found on Google's image search. An example of two of the images employed is shown in Figure 10.

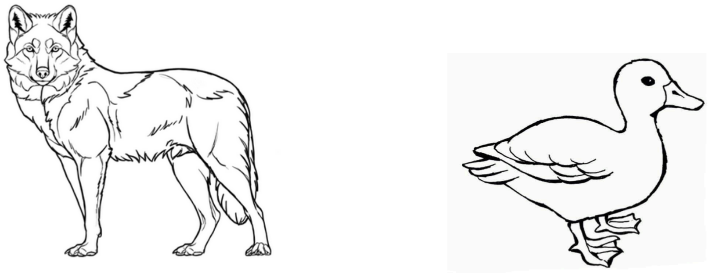


Figure 10. Examples of two of the images chosen to represent their respective characters

The image-objects to be matched with the fictive words described in the previous section consisted of six figures which were hand-drawn and later scanned. Three of the shapes had sharp, angular inflections, whilst the other three had wavy, round contours, as shown in Figure 11. Additionally, two foils were used, consisting of two elemental geometrical figures, a square and a circle.

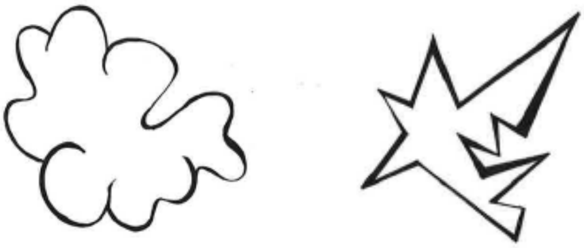


Figure 11. Examples of two of the shapes created for the "sound symbolism" tasks.

The full compilation of the images used for both music and fictive words tasks, including the warmups, is shown in Appendix A.

3.2.2.2 Written words

Since the image-objects could have been culturally and individually biased, in their role of representations of the ultimate objects, half of the trials for the music tasks consisted of written words where, instead of the participant being presented with the generic images described above, they were presented with a word in their respective language, written in a large font. These were used as controls in the two music tasks (see below). The word-objects for the ideophone tasks consisted of the translations of the examples (3-11) to both Chinese and Swedish, performed by native speakers. The full compilation of all written words for both music and ideophone tasks in both Swedish and Chinese is shown in Appendix B.

3.3. Design

Each participant was to perform the following four tasks, using the materials described in the previous section.

T1. Match music representamen with images as objects

T2. Match music representamen with written words as objects

T3. Match fictive word representamen with images as objects

T4. Match ideophone representamen with written words as objects

Each of these tasks had two conditions: a *less-contrastive* and a *more-contrastive*. In the more-contrastive condition, the participant was presented with two representamina, and two objects, which were determined by the content of each task. For example, if the task was a more-contrastive music task with images as objects, then the representamina consisted of two melodies (e.g. ‘wolf’ and ‘duck’) and the objects consisted of the images of the duck and the wolf. If the task was a more-contrastive ideophones with words as objects, the representamina consisted of two ideophones (e.g. *Grik-grak* and *Zorro-zorro*) and the objects consisted of their respective translations to either Swedish or Chinese. Figure 12 illustrates these two previous examples as presented in the experiment.

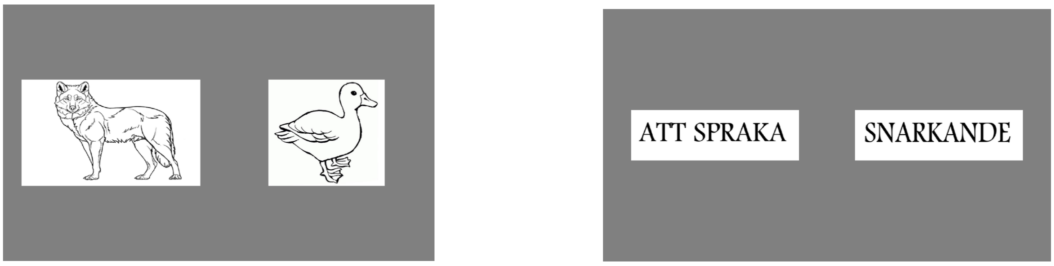


Figure 12. Examples of the more-contrastive condition (T1) and the more-contrastive condition for the ideophone task (T4) presented in Swedish.

In the less-contrastive condition, the participant was presented with one representamen and four objects, which, as in the more-contrastive condition, were determined by the content of each task. In a less-contrastive condition for music with written words as objects (T2), for example, the representamina consisted of one melody (e.g. ‘cat’) and the objects consisted of the word for the cat, which was the target object, and the words for three of the foils (i.e. pig, ballerina and fighter) in either Swedish or Chinese. If the task was fictive words with shapes as objects (T3), the representamina consisted of one fictive word (e.g. [lamu]) and the objects consisted of one round, soft-like shape which was the target object, and the three foils, which consisted of one sharp shape, a square and a circle. Figure 13 below illustrates these examples as presented in the experiment.

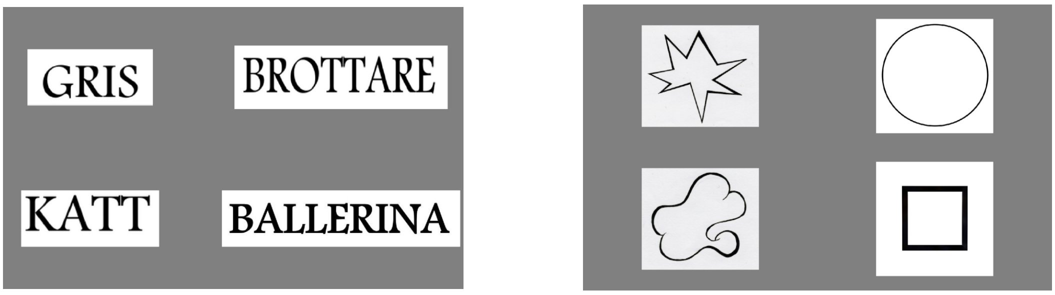


Figure 13. Examples of the less-contrastive condition (T2) presented in Swedish and the less-contrastive condition for fictive words (T3).

Intuitively, the less-contrastive condition is harder than the more-contrastive one. Theoretically, it can be argued that it requires a higher proportion of primary iconicity than secondary iconicity (see section 2.4), as the participant cannot use a simple exclusive inference in making the representamen-object matching (*à la* "this rather sounds like this image, so the other sound must be the other image").

In the more-contrastive condition, the combination of primary and secondary iconicity is as in Figure 8 (Chapter 2). In the less-contrastive condition, the process of semiosis consists of the same three steps, but differs in the number of candidate objects per representamen. This also consist of a combination of primary and secondary iconicity, but here the degree of secondary iconicity is not as high as it is in the more-contrastive condition, given that the participant has more choice possibilities. This process is shown in Figure 14 below.

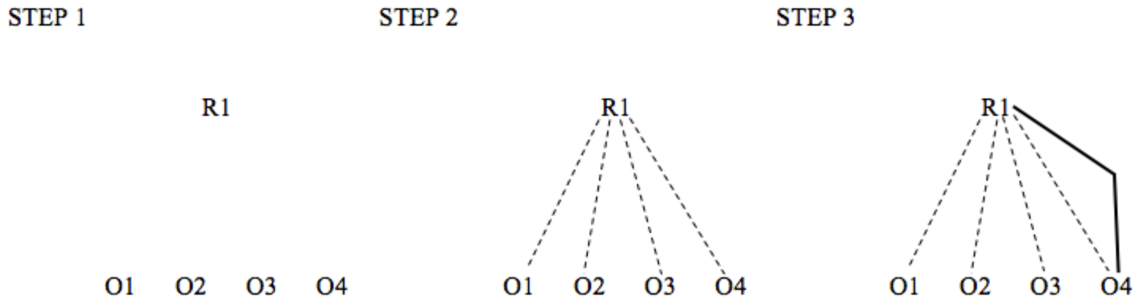


Figure 14. Matching representamina to objects by finding the most congruent iconic ground in the less-contrastive condition (adapted from Ahlner and Zlatev 2010)

The experiment was designed and carried out in PsychoPy. The aim was that the participant was to receive all instructions from the program itself, so as to minimize the interaction between the participant and the researcher during the experiment. All instructions were presented in both written and audial form, in either Swedish or Chinese. These were translated and recorded by the same L1 speakers that produced the translations for the materials described in section 3.2.

Prior to both warmups and the actual experiment, a set of general instructions was presented, explaining that the participant was to match musical melodies and made-up words to objects shown on the screen. The instructions specified that they were to be presented with images in pairs or in groups of four, and that when presented with a pair of objects, they would listen to two melodies or made-up words, and when presented with a group of four objects, they would only listen to one melody or made-up word.

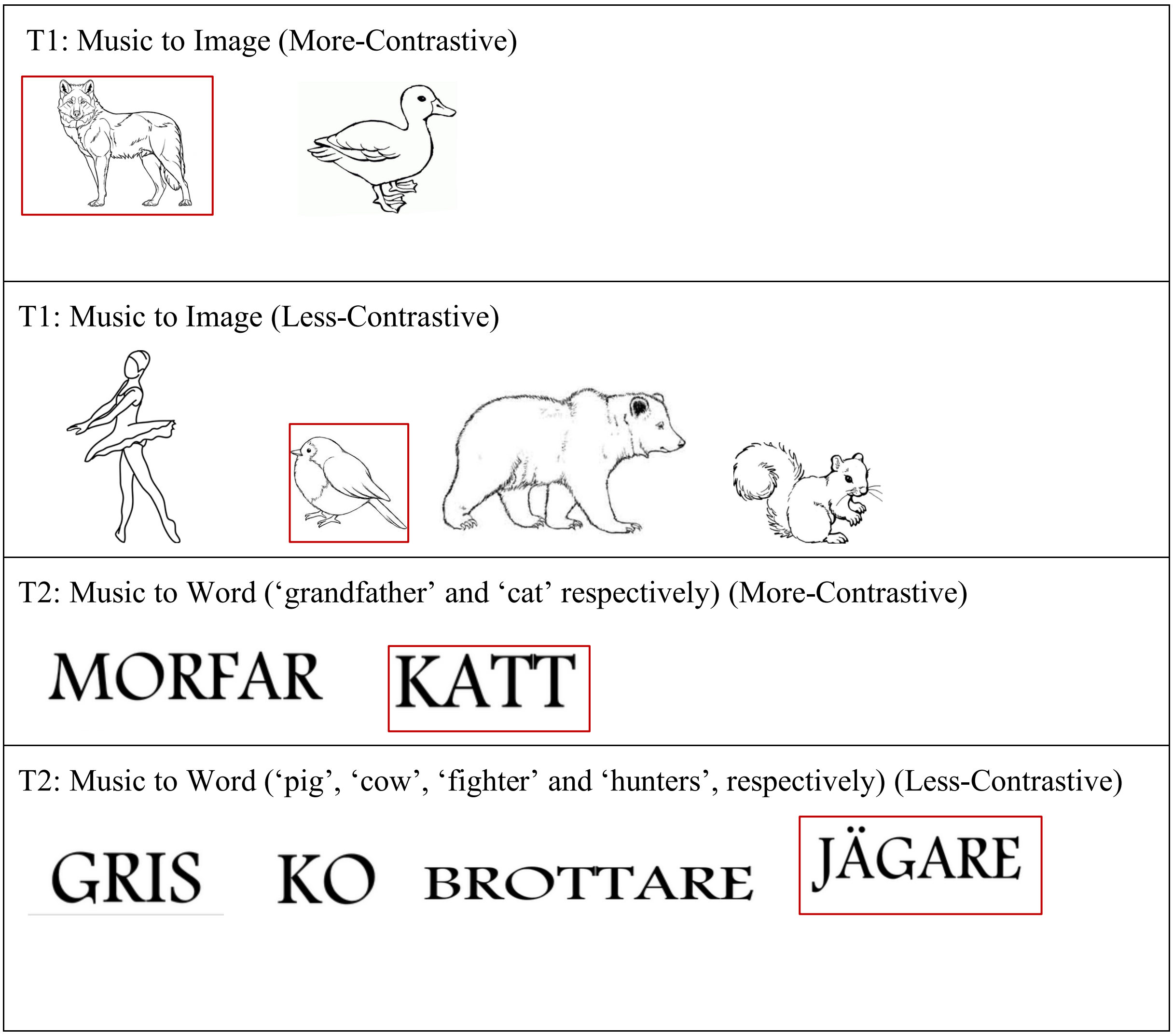
To minimize confusion, a set of instructions came before each task. Before less-contrastive conditions for T1, T2 and T4, the instructions stated that the following melody/word matched *one* of the four objects on the screen. Once the participant had read and listened to all instructions, they were presented with the four objects and the representamina. After the music, or the ideophone was finished playing, the instructions (in only audial format) asked the participant to click on the object that best matched the representamen, which was presented one more time. Once the participant had listened to the music or the ideophone the second time, they were to click on their object of choice. In every task, once the participant made their choice and clicked on the object, the program automatically showed the following task, so they did not get to change their answers.

Before the more-contrastive condition for T1, T2 and T4, the instructions stated that each of the two melodies (or made-up words), matched each of the objects on the screen. Once the participant had read and listened to all instructions, they were presented with the two objects. In these tasks it was imperative to make a clear distinction between both representamina, so for this purpose, the instruction enounced 'recording 1' and 'recording 2' before each representamen was played. After the participant had listened to both melodies, she or he was instructed to click on one of the two representamina, which was then played a second time. In this case, it was taken for granted that, once the participant had clicked on their object of choice, it meant that they had matched the remaining representamina to the remaining object.

For the both conditions in T3, no written instructions were presented. Instead, in the less-contrastive condition, the four objects were shown as the audio instructions informed the participant that one of the figures was called *X* and subsequently asked the participant to click on the figured called *X.* In the more-contrastive condition, the two objects were shown as the audio instructions stated that one figure was called *X* and the other was called *Y,* subsequently asking the participant to click on the image called *Y.* Similarly, it was taken for granted that the clicking on an object called *Y,* meant that they had assigned the name *X* to the remaining figure (i.e. the exclusive inference mentioned above). The position of the objects in each task was randomized every time the experiment was run, as well as the order of the tasks, meaning that no participant would see the stimuli in the same order, thus minimizing chances for associations based on the order the material was presented.

An important aspect that contributed to the design of the experiment was the music stimuli, as it was essential that each participant listened to all six melodies, which was accomplished by presenting the two conditions per type of task. For the fictive words task, they were designed so that if the target figure in the *more-contrastive* condition was one with “soft” shapes, then the target figure in the *less-contrastive* condition would be one with “sharp” shapes.

Three different versions of the experiment were prepared, with different combinations of stimuli and conditions for the different tasks were designed. Figure 15 shows the design of the first version in Swedish, where the target object, is marked with a red square. The visual presentation of all three tasks is presented in Appendix C.


T1: Music to Image (More-Contrastive)
T1: Music to Image (Less-Contrastive)
T2: Music to Word (‘grandfather’ and ‘cat’ respectively) (More-Contrastive)
MORFAR
KATT
T2: Music to Word (‘pig’, ‘cow’, ‘fighter’ and ‘hunters’, respectively) (Less-Contrastive)
GRIS
KO
BROTTARE
JÄGARE

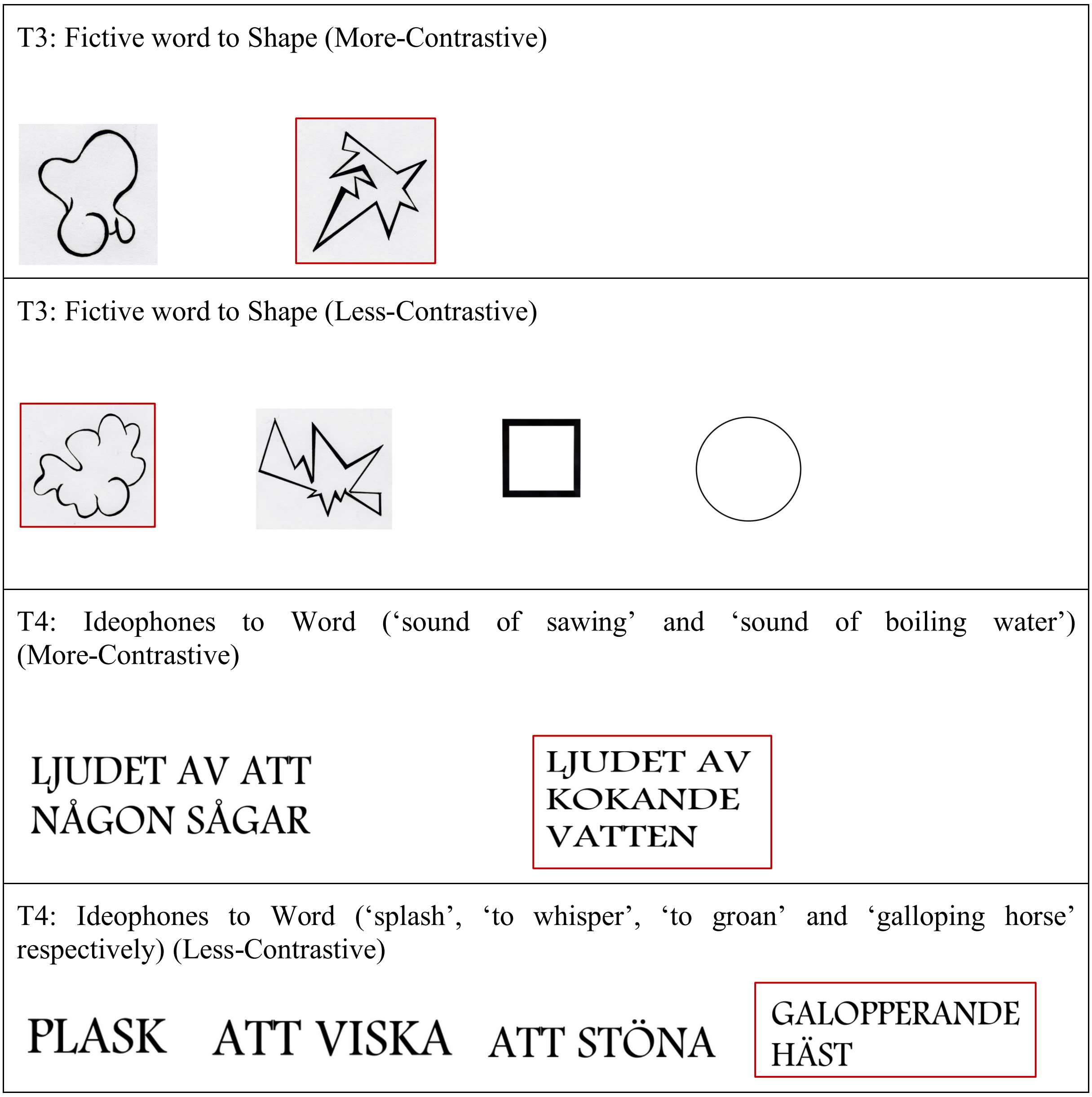

T3: Fictive word to Shape (More-Contrastive)
T3: Fictive word to Shape (Less-Contrastive)
T4: Ideophones to Word (‘sound of sawing’ and ‘sound of boiling water’) (More-Contrastive)
LJUDET AV ATT NÅGON SÅGAR
LJUDET AV KOKANDE VATTEN
T4: Ideophones to Word (‘splash’, ‘to whisper’, ‘to groan’ and ‘galloping horse’ respectively) (Less-Contrastive)
PLASK
ATT VISKA
ATT STÖNA
GALOPPERANDE HÄST

Figure 15. Organization of Version 1 (Swedish)

3.4 Participants

To address the third general hypothesis concerning possible cultural differences in the perception of music, twenty-one L1 Swedish speakers, and twenty-one L1 Chinese speakers were recruited, using online advertising and personal contacts. It was imperative that none of the participant had previously heard the piece *Peter and the Wolf,* as they would thus know the image or word that matched the melody. For this purpose, all of the potential participants were provided with Peter’s melody, which is the most recognizable of all. Those who did not recognize the music were eligible for participation.[7] The only other criterion was that they were native speakers of Swedish or Mandarin Chinese, and that none spoke Basque.

The Chinese group consisted of 21 participants (8 male and 13 female), with an average age of 29. All of them were from mainland China and moved abroad after the age of 18. The Swedish group also consisted of 21 participants, out of which 8 were male and 13 were female, and the average age was 25. All the participants performed the experiment in Sweden, either in Malmö or Lund, and each session took between twenty and thirty minutes. All participants signed a consent form, where they were presented with general instructions and were informed that a part of the session would be audio-recorded. Upon completion, each participant received a cinema ticket as reward for their participation.

3.5 Procedure

Oral interaction with the participants was performed in English. Prior to each session, I asked the participants to state their age, and the languages they spoke (to make sure they did not speak Basque). Participants were then informed that they would start with a warmup session, where they would have the chance to ask any questions or state doubts if necessary. I also clarified that all instructions would be presented in their language. Once the two warm-up tasks were done, I asked again if they had any questions or doubts before continuing. Following this, each participant was presented with the four tasks (in two conditions each) in random order. As stated before, the program provided all instructions, so I did not interact with the participant at all while they were taking the experiment. Instead, my task was to keep track of their associations, for the purpose of the interview that was to follow. The experiment lasted around 8 minutes.

Following this, the participant was interviewed concerning the reasoning behind their choices. The participant was shown a slide-show version of the experiment that they had just run, in their respective language. The procedure here was simple: each slide provided them with the respective representamina to the objects shown. Once they had heard all the representamina one more time, I reminded them of the object they had clicked on and

[7] It is worth mentioning that during the recruitment process, there was a higher amount of Swedes that recognized the music, whilst none of the Chinese (candidate) participants did so.

proceeded to ask why they had chosen that specific one. In the case for the more-contrastive conditions, I asked if they would have matched the remaining representamen to the remaining object, and if so, why. In the end, I asked they participant if they had any questions, and provided them with a brief explanation of the purpose of the experiment. Upon completion, they were rewarded with a cinema ticket.

3.6 Summary and specific hypotheses

This chapter described the methods employed in the empirical study for this thesis. I started by presenting the materials employed, followed by the design of the experiment and the participants that partook in the experiment, finalizing with the procedure.

On this basis, the three general hypotheses from Chapter 2 could be specified as follows:

1. Throughout Tasks 1-4, more-contrastive conditions will show more successful matching than less-contrastive conditions.
2. For the music tasks (T1 and T2), the unimodal conditions (BIRD, HUNTERS), will be matched to its corresponding object more successfully than cross-modal conditions (CAT, DUCK, WOLF, GRANDFATHER).
3. For the music tasks (T1 and T2), Swedish L1 speakers will show more successful matching than L1 Chinese speakers.
4. For the fictive words and ideophones tasks (T3, T4) no difference in the performance between the Swedish and Chinese participants is expected.
5. L1 participants of both languages are expected to perform better in the fictive words tasks (T3) than in the ideophones tasks (T4).

The motivation for the hypotheses was the following:

The first specific hypothesis is based on the assumption that in both programmatic music and speech, a combination of primary and secondary iconicity is responsible for succeeding in the tasks. However, the role of secondary iconicity is greater in in the more-contrastive than the less-contrastive condition, and it has been argued that this is characteristic for both language and music, unlike pictures (Sonesson 2009: 5).

The second hypothesis was motivated by the observation that sound-to-sound mapping characterizes the most common kind of ideophones across languages (Dingemanse, 2012), which could be due to unimodal mappings being more transparent than cross-modal ones.

The third hypothesis specifies the third general hypothesis, by expecting that due to their greater familiarity with Western classical music Swedish participants would find it easier to perceive referential iconicity in the music tasks.

The fourth hypothesis, on the other hand, expects no differences concerning the fictive words tasks (T3), as the type of iconicity presented in these tasks can be understood as reflecting universal "sound symbolism" (Imai and Kita, 2014). Given that the ideophones were in Basque, which was equally unfamiliar to both the Swedish and Chinese participants, no difference was expected here either.

The fifth hypothesis predicted that all participants were expected to perform better in the fictive words tasks (T3) than in the ideophones tasks (T4), given that the first reflects universal human capacities in speech comprehension (Imai and Kita, 2014), while the latter involves word-forms that occur in an actual natural language, and are thus less exaggerated and more conventionalized (Dingemanse, 2018).

CHAPTER 4. RESULTS

4.1 Introduction

This chapter presents the results of the study described in Chapter 3, organized in accordance to the five specific hypotheses presented and motivated at the end of that chapter. I conclude with a short summary.

All of the results were analyzed statistically, where each hypothesis was tested using mixed-effects logistic regression with random intercepts for participant and item. From the 42 participants who took part in the experiment, a total of 336 answers were gathered for all tasks (T1-T4), with 84 answers per task. The overall results obtained for both linguistic and music tasks, are shown in Table 5, with accurate answers (1), and less-accurate answers (0) for all participants.

Table 5. Overall results for Linguistic and Music tasks

Tasks	ANSWER		Percentage
	0	1	
Linguistic	63	105	62.5%
Music	72	96	57.1%

This can be further visualized in Figure 16*,* where the proportions of accurate answers for linguistic and music tasks are presented. A binomial test showed that the proportions of correct answers in music and linguistic tasks were both above chance (0.375), *p*= 0.000 (music tasks); *p*= 0.000 (linguistic tasks). The regression analysis further showed that the effect of condition on performance was non-significant: ***β***= -0.3494, z= -0.674, *p*= 0.5005, indicating that, even though the participants performed slightly better at the linguistic tasks, there was no statistically significant difference between the two kinds of tasks.

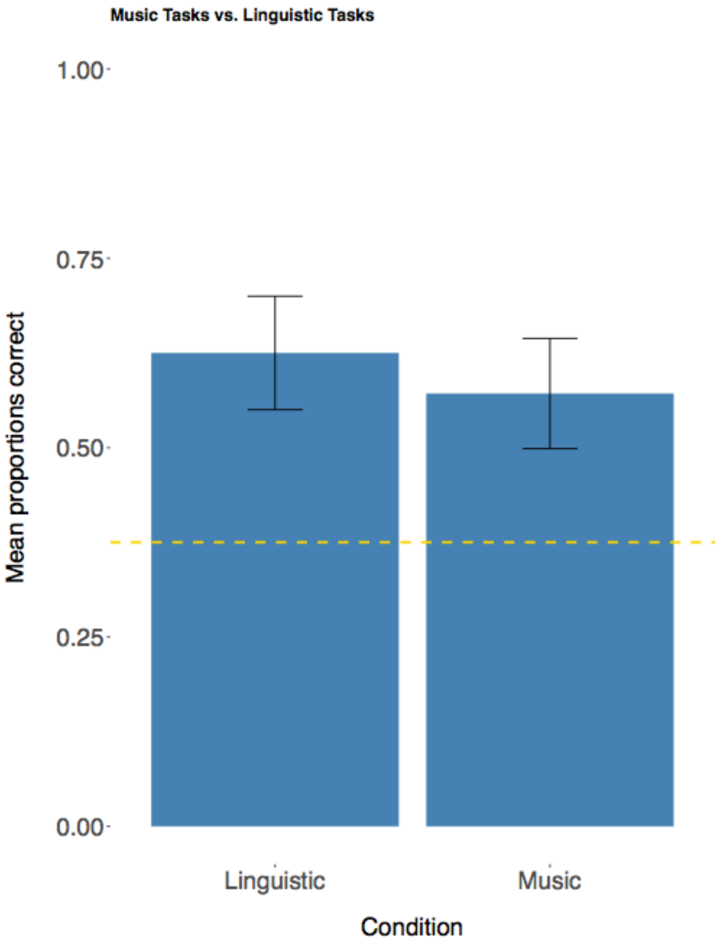


Figure 16. Proportions of accurate answers for both music and linguistic tasks. The golden dotted line represents the chance level.

As stated in Chapter 3, half of the objects used in the music tasks (T1-T2) were generic images, while the other half were written words in either Swedish or Chinese. This was because the image-objects could have been culturally and individually biased, thus being potentially misleading, especially for the Chinese participants (see section 3.2.2.2). The regression analysis showed that despite that the music to image task (T1) gave rise to slightly higher accuracy than the music to word task (T2), (see Table 6), this difference was not significant: *ß*= -0.3013, *z= -0.775, p*= 0.438. Furthermore, the exact binomial test showed that for both tasks accuracy was above chance significance (0.375), p-value = 0.001971 (words), *p*-value = 0.000 (images).

Table 6. Overall results for music tasks: images (T1) and words (T2)

Tasks	ANSWER		Success rate
	0	1	
Images (T1)	33	51	60.7%
Words (T2)	39	45	53.5%

If we look at each language group individually (Tables 7-8), the accuracy difference between the tasks is likewise not significant: *ßImageWordsMusicChinese*= -0.3857*, z*= -0.875*, p*= 0.381. and *ßImageWordsMusicSwedish*= -0.2239*, z*= -0.388*, p*= 0.698. The exact binomial test, however, showed a difference. For the Chinese group, accuracy in the images task was above chance significance (0.375), *p*= 0.003085, but not in the task with words as objects, *p*= 0.06663. For the Swedish group, the exact binomial test showed that accuracy was significantly above chance for both tasks: *p*= 0.001149 for images, *p*= 0.007564 for words.

In sum, this indicates that it was not easier to match musical representamina to linguistic objects than to images, and for the Chinese group, it was even somewhat harder. Since the purpose of the task with words as objects (T2) was to serve as a control for the quality of the images as representations of the real-world referents in T1, we can see that these were apparently not culturally biased toward the Swedish group.

Table 7. Overall results for music tasks: images (T1) and words (T2) for Chinese participants

Tasks	ANSWER		Success rate
	0	1	
Images (T1)	17	25	59.5%
Words (T2)	21	21	50.0%

Table 8. Overall results for music tasks: images (T1) and words (T2) for Swedish participants

Tasks	ANSWER		Success rate
	0	1	
Images	16	26	61.2%
Words	18	24	57.1%

4.2 Results by hypothesis

4.2.1 H1: More-contrastive vs. less-contrastive conditions

In order to test the first hypothesis, concerning higher success rates for the more-contrastive than the less-contrastive condition, the total amount of answers for all participants were counted. Out of the total 336 replies gathered, there were 168 per condition. Table 9 and Figure 17 show considerable differences between the two conditions. The effect of condition on performance was, as expected, highly significant: $ß = 2.0500$, $z = 6.684$, $p= 0.000$, thus supporting the first hypothesis. Furthermore, an exact binomial test indicated that the proportions of correct answers for both more and less-contrastive tasks were greater than chance: for the more contrastive tasks, this would be 0.5, as the participant had a 50-50 chance of making the correct association, where the *p*-value obtained was of $p= 0.000$. For the less-contrastive tasks, the participant had four options, meaning that the chance was 0.25, and $p= 0.000$

Table 9. Overall results for More and Less-Contrastive conditions

Condition	ANSWER		Success rate
	0	1	
Less-Contrastive	99	69	41.1%
More-Contrastive	36	132	78.5%

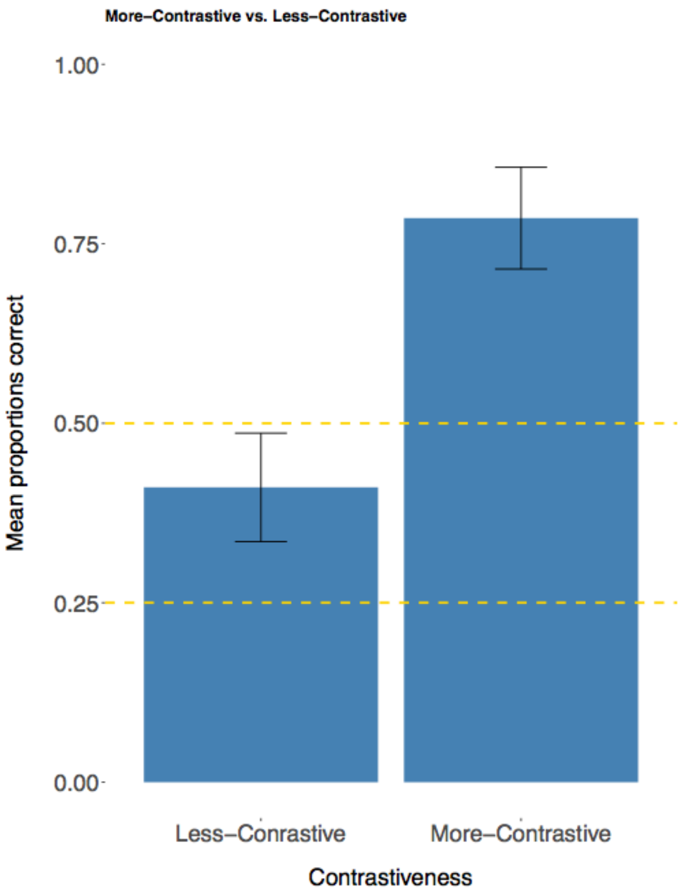


Figure 17. Proportions of accurate answers for less- and more-contrastive conditions for all tasks (T1-T4). The golden dotted line represents the chance levels (50% for more contrastive and 25% for less contrastive conditions).

Looking at the music tasks (T1 and T2) separately, Table 10 shows that, as expected, music tasks obtained much better results in the more-contrastive condition than in the less contrastive condition: *ß*= 2.1052, *z*= 5.650, *p*= 0.000. The exact binomial test showed that both conditions were above chance significance (less-contrastive (0.25): *p*= 0.03248; more-contrastive (0.50): *p*= 0.000).

Table 10. Overall results for more and less-contrastive conditions for music tasks (T1 and T2)

Condition	ANSWER		Success rate
	0	1	
Less-Contrastive	55	29	34.5%
More-Contrastive	17	69	79.7%

The results obtained for the linguistic tasks (T3 and T4), presented in Table 11, were similar, with significant differences between the conditions: *ß*= 1.8896*,* z= 3.391*,* p= 0.000695 and above chance significance (less-contrastive (0.25): p=0.000; more-contrastive (0.5): p=0.000).

Table 11. Overall results for more and less-contrastive conditions for linguistic tasks (T3 and T4)

Condition	ANSWER		Success rate
	0	1	
Less-Contrastive	44	40	47.6%
More-Contrastive	19	65	77.3%

4.2.2 H2: Unimodal vs. cross-modal conditions

Again, the total amount of answers for all participants were counted. Unlike the symmetrical design in H1, where there was a 50-50 division between the conditions, the amount of cross-modal mappings doubled the number of unimodal ones. Given that this hypothesis concerns only musical tasks (T1 and T2), the total amount of replies is of 168. Out of this total, 112 corresponded to cross-modal and 56 to unimodal conditions. Table 12 below shows the total results for both unimodal and cross-modal conditions – accurate (1) and less-accurate (0) – and the percentage of success.

Table 12. Overall results for unimodal and cross-modal conditions

Condition (Mapping)	ANSWER		Success rate
	0	1	
Cross-Modal	43	69	61.6%
Unimodal	29	27	48.2%

Contrary to the prediction stated in this hypothesis, where unimodal mappings were expected to be easier than cross-modal mappings, we can see there that cross-modal stimuli condition had a higher success rate, though the regression analysis showed that the difference between the two conditions was not significant (*ß* = -0.5481, z = -1.519, *p*= 0.1288). However, the exact binomial test showed that proportions of correct answers for the cross-modal condition was above chance (0.375), *p*= 0.000, while those for unimodal conditions were below chance (0.375), *p*=0.06595 (see Figure 18). From this, it can be concluded, against expectations, that the cross-modal condition was in fact easier for participants, and H2 was not supported.

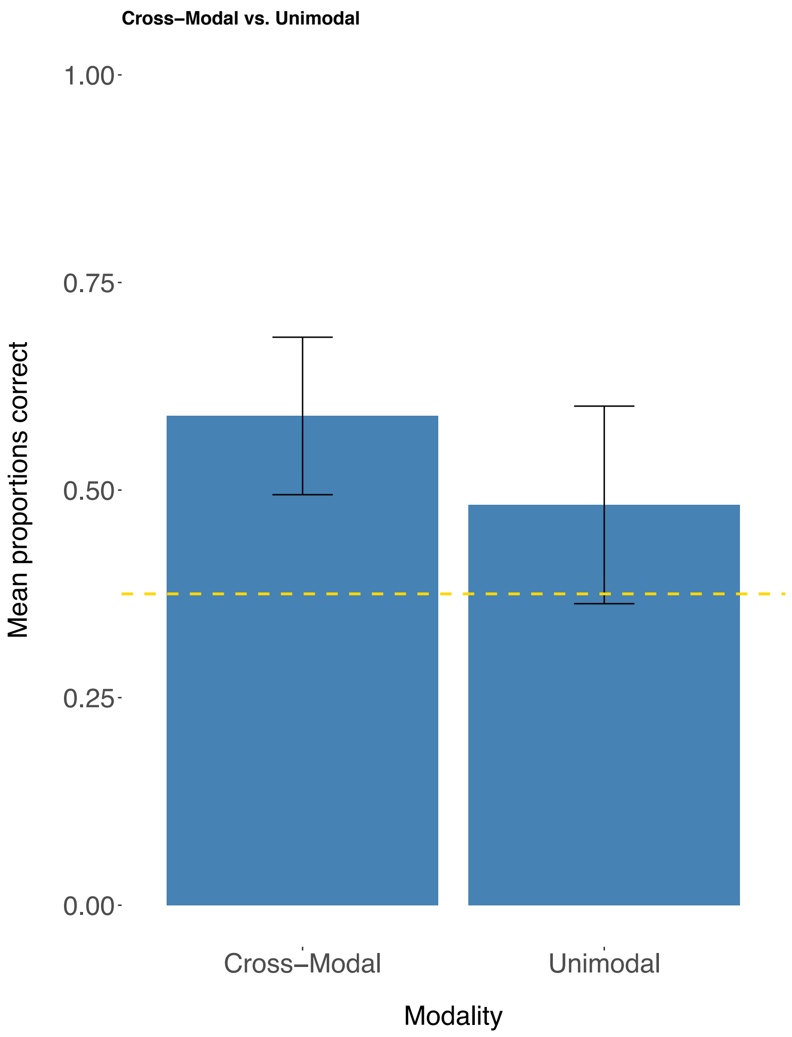


Figure 18. Proportions of accurate answers for unimodal and cross-modal conditions in music tasks (T1-T2). The golden dotted line represents the chance level (0.375).

4.2.3 H3: L1 Swedish speakers vs. L1 Chinese speakers in music tasks

This hypothesis, as the previous one, focuses only on music tasks (T1-T2), where a total of 168 answers were collected. In order to analyze this data, the total amount of answers for the two tasks performed by L1 Swedish speakers was compared to the total amount of answers provided by L1 Chinese speakers, where each group provided 84 responses. The overall results for both language groups are presented in Table 13 and visualized in Figure 19.

Table 13. Overall results for L1 Swedish and L1 Chinese speakers in the music tasks (T1 and T2)

Group	ANSWER		Success rate
	0	1	
Chinese	38	46	54.7%
Swedish	34	50	59.5%

The regression analysis showed that the difference was not significant: $ß$ = 0.1992, z = 0.631, p= 0.528 and the exact binomial test showed that proportions of correct answers for the music tasks for Swedish and Chinese participants was above chance (0.375), p= 0.0009627 (Chinese), p= 0.000 (Swedish). Thus, H3 was not supported.

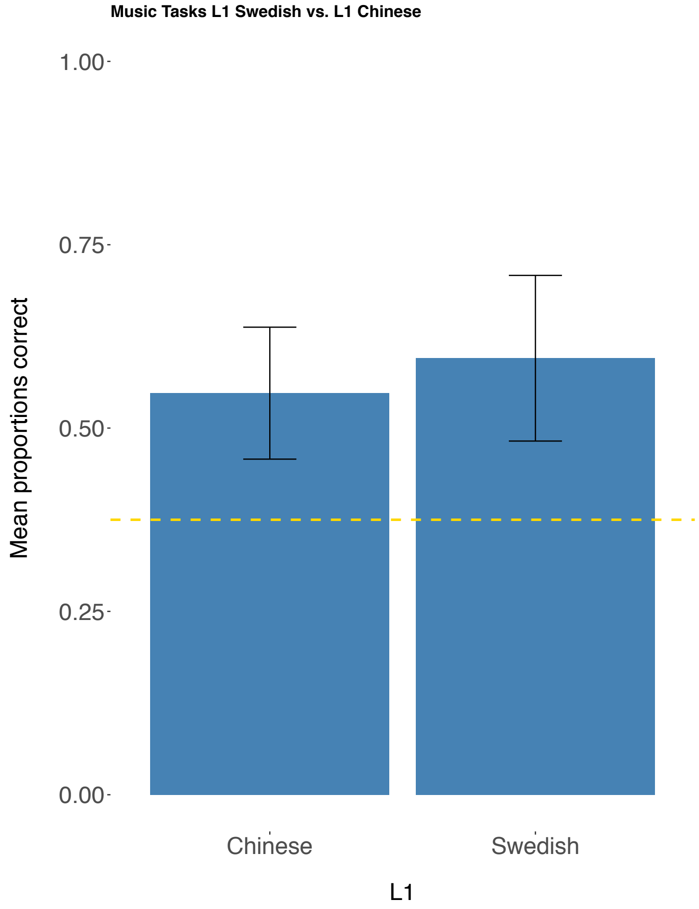


Figure 19. Proportions of accurate answers for L1 Swedish and L1 Chinese speakers in music tasks (T1-T2). The golden dotted line represents the chance level (0.375).

4.2.4 H4: L1 Swedish speakers vs. L1 Chinese speakers (fictive word and ideophones tasks)

For the fictive words linguistic task (T3), a total of 84 answers were gathered. Similar to the previous hypothesis, the total amount of answers provided by participants of each group (42 answers per language group) were gathered and then compared, as shown in Table 14 and Figure 20.

Table 14. Overall results for L1 Swedish and L1 Chinese speakers in the fictive words tasks (T3)

Group	ANSWER		Success rate
	0	1	
Chinese	17	25	59.5%
Swedish	11	31	73.8%

While the L1 Swedish performed somewhat better than L1 Chinese speakers, the regression analysis showed no significant difference between the two groups: *ß* = 0.7139 , z = 1.447, *p*= 0.148. Also, the exact binomial test showed that proportions of correct answers for the fictive words task for both groups was above chance significance (0.375), *p*= 0.003085 (Chinese), *p*= 0.000 (Swedish). Thus, concerning the matching of fictive words, the hypothesis was supported.

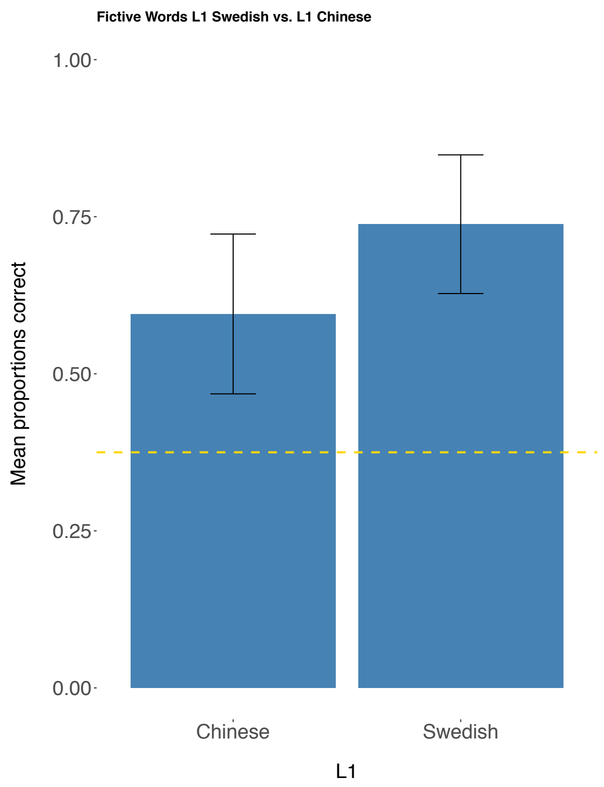


Figure 20. Proportions of Accurate answers for L1 Swedish and L1 Chinese speakers in the fictive words task (T3). The golden dotted line represents the chance level (0.375).

With respect to the ideophones task (T4), where a total of 84 answers were collected (42 per group), there was even less difference between the two language groups, as shown in Table 15 and Figure 21. According to the mixed-effects logistic regression, the difference was not significant (*ß* = 0.2269 , z = 0.362, *p*= 0.717.) and the exact binomial test showed that proportions of correct answers for the ideophones task for Swedish and Chinese participants was above chance (0.375), *p*= 0.007564 (Chinese), *p*= 0.003085 (Swedish). Thus, the hypothesis was supported for this task as well.

Table 15. Overall results for L1 Swedish and L1 Chinese speakers in the ideophones tasks

Group	ANSWER		Success rate
	0	1	
Chinese	18	24	57.1%
Swedish	17	25	59.5%

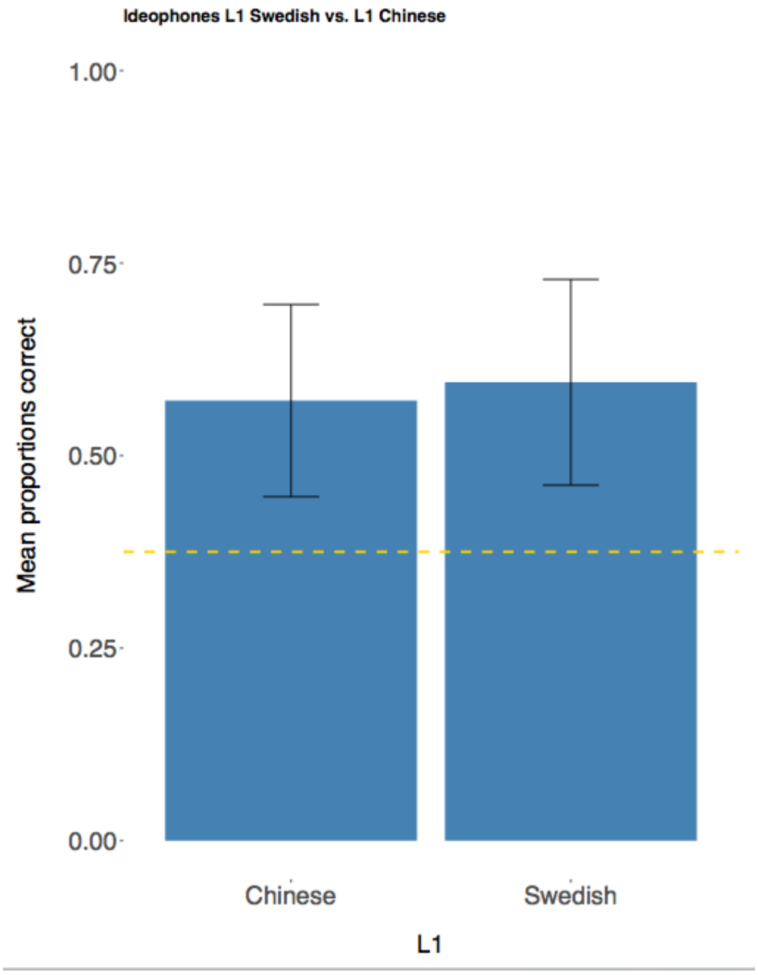


Figure 21. Proportions of Accurate answers for L1 Swedish and L1 Chinese speakers in ideophones tasks (T4). The golden dotted line represents the chance level (0.375).

4.2.5 H5: Fictive words vs. ideophones tasks

Similar to the previous hypothesis, this concerned both linguistic tasks (T3 and T4), with a total of 168 answers, 84 per task, but compared tasks rather than groups. It predicted that participants were to perform better in the fictive words task (T3) than in the ideophones task (T4). The results are shown in Table 16 and Figure 22.

Table 16. Overall results for fictive words (T3) and ideophones (T4) tasks

Task	ANSWER		Success rate
	0	1	
Fictive Words (T3)	28	56	66.6%
Ideophones (T4)	35	49	58.3%

While participants seemed to perform better in the fictive words than in the ideophones task, the regression analysis showed that this difference was not significant (*ß* = -0.6568 , z = -0.700, *p*= 0.484). Furthermore, the exact binomial test showed that proportions of correct answers for the ideophones and fictive words tasks was above chance significance (0.375), *p*= 0.000 (Fictive Words), *p*= 0.000 (Ideophones). Thus, H5 was not supported.

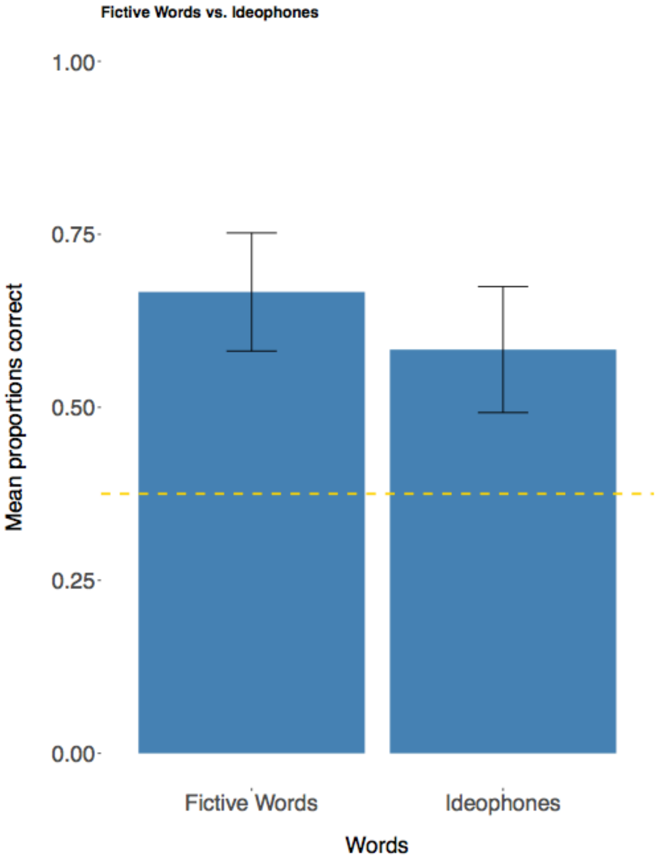


Figure 22. Proportions of accurate answers for fictive words and ideophones tasks (T3-T4). The golden dotted line represents the chance level (0.375).

4.3 Summary

In this chapter, the results and statistical analysis of each of the five hypotheses formulated in Chapter 3 were presented. Out of the five hypotheses, two were supported: H1 concerning the conditions more or less contrastive, and H4 concerning lack of differences between the Swedish and Chinese groups in the linguistic tasks. Importantly, however, the results for even those hypotheses that were not supported yield an important result: referential iconicity played an important role in solving the respective tasks, no matter if it concerned unimodal or cross-modal mappings, and irrespective of cultural group (and thus familiarity with Western classical music).

The results for the first hypothesis, which predicted a statistically significant difference between more-contrastive and less-contrastive conditions showed that, indeed, more-contrastive tasks were more transparent to participants of both language groups. The second hypothesis anticipated that unimodal conditions (for the stimuli BIRD and HUNTERS) were to be easier than cross-modal conditions, for the other four stimuli. However, the results showed that the unimodal conditions were in fact below chance significance, unlike the cross-modal condition, and thus harder for participants. The third hypothesis predicted an advantage for the Swedish group in the music tasks (since they involved fragments from Western classical music), but contrary to the expectations, both groups performed similarly.

Concerning both linguistic tasks (fictive words and ideophones), the fourth hypothesis anticipated no difference between the groups, which was indeed the case. On the other hand, the fifth and last hypothesis predicted that the overall result of the fictive words task were to show a higher success rate than in the ideophone task, which was not the case.

Perhaps most importantly, the results showed that the participants of both language groups solved both music and linguistic tasks equally well, and that there was no difference between the case where objects were represented by images (T1) and by words (T2), indicating that the use of (schematic) images was reliable.

Chapter 5. Discussion

5.1 Introduction

Having discussed the results of the experiment with respect to the five specific hypotheses in the previous chapter, it is now pertinent to move to a higher level of generality in the discussion. Thus, this chapter will be structured in terms of the three general hypotheses that concluded Chapter 2, repeated here again:

- **Hypothesis 1**: It will be easier for a participant to recognize the referential object (in both speech and music) when representamen and object are in the same modality (unimodal iconicity) than in different modalities (cross-modal iconicity).
- **Hypothesis 2**: In both programmatic music and speech, a combination of primary and secondary iconicity is responsible for succeeding in the task stated in Hypothesis 1.
- **Hypothesis 3**: Given that there is a high degree of conventionality in music, linked to one's culture, there will be differences in how participants belonging to different musical traditions, will solve the task in Hypothesis 1, when the musical stimuli are taken from one of these traditions.

The discussion will combine the results from the experiment described in the previous chapter (the third-person method), along with results from the post-experimental interviews, using a combination of first-person and second-person aspects of the methodological triangulation of cognitive semiotics (see section 2.2).

5.2 Hypothesis 1: Unimodal vs. cross-modal mappings

The hypothesis predicted that unimodal mapping (i.e. sound-to-sound) in the music task would be easier for participants from both cultures than cross-modal mappings (e.g. sound-to-movement, sound-to-visual patterns). This derived from the assumption that sound-to-sound mappings are "the simplest kind of semiotic mapping" (Dingemanse, 2012: 663), accounting for their prevalence among ideophones in the world's languages (see section 2.4). This prediction was, however, not supported, given that the results pointed in the opposite direction: the cross-modal music-to-object mappings had a success rate of 60%, whereas unimodal had a success rate of 48.2%. Even though this difference was not statistically significant, the success rate for unimodal matches was below chance significance, indicating that they were in fact harder.

Some methodological issues may have contributed to this result: there were 54 trials with the unimodal condition, as opposed to 112 trials for cross-modal condition, given that only two musical fragments were judged to involve the former (BIRD, HUNTERS), and four the latter. The reason for this unbalance was that the experiment itself was designed around the six characters of *Peter and the Wolf*, and the priority was that each musical fragment was to be the target once in a more-contrastive condition, and once in a less-contrastive condition for each participant (see section 3.3). The fact that only two of the six fragments were (considered as) unimodal was a disproportion that was not thought ahead. Clearly, this is a factor that must be more carefully treated in future research.

An object that was problematic was the filler of the Ballerina, given that every time it was given as an option, participants tended to pick it regardless of the pitch or rhythm of the melody. This bias was probably due to the genre of the music fragments, specifically classical music, which could be seen as standing in an *indexical* association (i.e. spatio-temporal contiguity) with the object in question (see section 2.2).

For additional insight on the nature of the associations made by participants, it is pertinent to turn to the interviews performed after each participant took the experiment, where the interviewer went through their choices and asked them why they had made such associations. These indicated that even when participants made the correct association in the unimodal conditions, the reasoning behind their associations did not usually rely on sound-to-sound mappings. Instead, comments like “this sounds like a small bird flying around” (i.e. mappings from sound to movement) were more common. Mappings from sound to inner feelings and cognitive states were also frequent, where comments suggested that participants associated the HUNTER fragment with “danger” or “war”. Concerning the cross-modal conditions, most of the comments suggested mappings from sound to movement or sound to movement and sound to inner feelings and cognitive states: “this melody [CAT] sounds like how a cat moves”, or “this melody [WOLF] sounds like darkness, or a full moon, or danger”. An interesting point that emerged from the interviews was the fact that associations could be very subjective, pointing to the need to elaborate the first-person perspective method in future studies.

Theoretically, the results of the study suggest that Dingemanse’s (2012) tentative implicational hierarchy (see section 2.4.1) may not apply when it comes to referential iconicity in music, where cross-modally mappings may in fact be both more common, and easier to perceive. Thus, instead of sound-to-sound mappings being “the simplest kind of semiotic mapping” (Dingemanse, 2012: 663), sound-to-movement mappings may be the most transparent ones when it comes to referential iconicity in music.

Furthermore, it is important to highlight that each musical theme is itself composed of different elements that make each theme unique, such as its pitch, rhythm, tonality and tempo. When trying to understand why participants associated unimodal representamina such as the bird’s melody, to a bird’s movement, it is clear that it was not just because of the high pitch of the instrument, but also due to the interplay between the theme’s other elements. In the case of the bird’s melody, these would be its melodic intonation, its major key and its fast tempo and rhythm. The specific mutual relations between the theme’s elements constitute the diagrammatic icon of the bird’s melody.

5.3 Hypothesis 2: Primary and secondary iconicity

As discussed in section 2.4 (and 3.3), the process of interpreting referential iconicity in speech and music was expected to show a combination of both primary iconicity (from ground perception to sign recognition), and secondary iconicity (from sign knowledge to ground recognition). But the roles of each kind of process was expected to vary, depending on the type of task, and possibly on how transparent the representamen-object relationship is.

With respect to the “type of task” factor, the more-contrastive condition operationalized a higher role for secondary iconicity (see Figure 8), as participants were told that they are presented with two signs, and only have to decide which of the two possible mappings is more “natural”. The less-contrastive condition (see Figure 14), on the other hand, was an operationalization of a greater role for primary iconicity, as participants have to choose between four possible objects for a single representamen, and make this decision on the basis of the perception of the “best” iconic ground. The specific hypothesis (H1) corresponding to this, which predicted that participants of both language groups would be more successful in the more-contrastive conditions (for all four tasks) than in less-contrastive conditions, was clearly supported. This is in line with Sonesson’s conjecture that (referential) iconicity in both speech and musical signs is predominantly of the secondary kind. Still, as pointed out in section 4.2.1, even the less-contrastive condition gave rise to success rates that were significantly above chance. This is the clearest indication that the referential iconicity in question is not only secondary, but involves a *combination* with primary iconicity. In this respect, the general hypothesis was also supported.

With respect to degree of transparency, it could have been expected that the unimodal mappings would have been more so, and thus have a higher role for primary iconicity. But as shown with respect to Hypothesis 1, this was not the case, at least with respect to the musical representamina. Another expected difference that could be seen as a reflection of different proportions of primary/secondary iconicity was likewise not supported: that the unconventional fictive words would be more transparent than the more conventionalized ideophones (see section 4.2.5). Thus, this "non-difference" can be interpreted as more or less equal roles for both primary and secondary iconicity in (a) unimodal and cross-modal mappings in music and (b) unconventional and conventional speech forms, at least in the case of the particular stimuli that were used in the experiment.

In sum, the results of the study support the hypothesis that in both speech and music signs, there is a combination of primary and secondary iconicity (Ahlner and Zlatev, 2010), but with a higher degree of the latter (Sonesson, 2009). As pointed out in Chapter 1, this is rather an example of Jakobson's (1965) understanding of the sign itself as not being exclusively based on a single kind of ground, but rather on a combination of grounds, where one is (usually) predominant.

5.4 Hypothesis 3: Cultural differences and conventionality in music

Given the differences between Chinese musical traditions and Western musical traditions emphasized by Shen (1991, 2000) (see section 2.3.2), and considering that the Western classical music is much more prevalent and accessible in Sweden than China, leads to the prediction that Swedish participants were to perform better in the two music tasks (T1 and T2). Interestingly, this hypothesis was not supported, as there was no significant difference between the success rates of the two language groups, and further that both were above chance. In addition, there were no significant differences between the two music tasks (linking to images or words denoting the same referents, respectively), indicating that the images used were not biased towards Swedish culture. This implies that familiarity with musical conventions is not a prerequisite for perceiving referential iconicity in programmatic music, in the broad sense of the term, as briefly discussed in the introduction.

The lack of cultural differences in the music tasks could be due to Prokofiev's specific choice of interpretation of each character in a way that is culture-general. Furthermore, the Chinese speakers that participated in the study could perhaps be seen as non-representative, as all lived in Sweden, and could have been familiarized with the Western musical tradition. Alternatively, it could be the case that despite there being a great deal of culture-specific conventionality, the way we understand and listen to music is not directly linked to each individual's culture, which could point to a universal aspects in music perception. Thus, it is important to highlight that the present findings apply to this *particular* music piece, and in order to choose between one of these interpretations, it would be necessary to investigate a more varied series of programmatic music, with more "mono-cultural" participants.

Further, despite the fact that this general hypothesis concerns the music tasks, it is pertinent to discuss the lack of cultural differences, also found in the linguistic tasks involving fictive words and ideophones. As was stated in section 4.2.4, despite Swedish participants performing better than Chinese participants, the mixed-effects logistic regression analysis showed no significant difference between the two language groups. This finding stands in line with Imai and Kita's (2014: 5) proposal of "universal sound symbolism", but extending this to the Basque ideophones used in the study as well, and not only the fictive words, as there were no significant differences between these tasks:

> There has been an assumption in the literature that sound symbolism is universal; if a certain sound–meaning correspondence is identified by speakers of one language, this should be generalizable to speakers of any other languages. This assumption has been supported by the fact that speakers of many different languages sense Köhler's shape sound symbolism [26] in the same way, as reviewed earlier (for English [32], Japanese [68,69], Himba [27], Kitwonge-Swahili bilinguals [70])[8].

Nevertheless, it is necessary to mention that many Chinese participants seemed to have difficulties with the fictive words task (T3), as they questioned the formulation of the instructions (*One of these figures is called e.g. lulu and the other is called e.g. keti. Click on the image called e.g. lulu*) during the interviews. Their difficulty was in that they could not comprehend why they were to name an abstract shape. In particular, many participants picked the "blobby" figure despite listening to a word with "sharp" consonants and vowels, since it "looked more alive", as one participant stated. Interestingly, even though Swedish participants

[8] [26] Köhler (1929); [32] Thompson and Estes (2011); [68 69] Asano et.al (in review) and Miyazaki et. al (2011); [27] Bremner et. al (2013); [70] Davis (1961).

did not seem as confused by this task, some participants also chose the blobby figure because of its greater degree of *animacy* (known to be a relevant factor in linguistics), as expressed by Yamamoto (2006: 30): "only animal beings can be agents in a normal sense […] the agency concept goes hand in hand with that of animacy, and that both notions are highly significant determiners of mind-style or world view".

5.5 Summary and suggestions for future research

The discussion in this chapter followed the three general hypotheses presented at the end of chapter 2. The first hypothesis predicted that it would be easier for participants to recognize the referential object when representamen and object are in the same modality (unimodal iconicity) than in different modalities (cross-modal iconicity) in music tasks. This hypothesis was not supported, indicating that cross-modal iconicity could be more prevalent in music than in speech, thus reflecting a possible difference between the semiotic systems in question. The results could have been, however, affected by the unbalanced design of the experiment, and in future research, the same amount of unimodal and cross-modal conditions need to be used. In addition, a wider range of programmatic music needs to be explored.

The second hypothesis anticipated that in both programmatic music and speech, a combination of primary and secondary iconicity is responsible for succeeding in each one of the four tasks. The hypothesis was supported, given that more-contrastive conditions were almost twice as successful as less-contrastive conditions. However, both conditions were above chance level, indicating a clear role for primary iconicity. Furthermore, these conclusions applied when analyzing music and linguistic tasks independently, indicating no difference between the semiotic systems in this respect.

The third hypothesis proposed that given a high degree of conventionality in music, linked to one's culture, there would be differences in how participants belonging to different musical traditions would solve the music tasks. The results obtained showed that this was not the case, as participants of both groups performed similarly. This shows that referential iconicity in at least *some* programmatic music is not culture-specific. To generalize these findings, participants who are more "mono-cultural" than the Chinese speakers involved in the study need to be used, as well as a larger programmatic repertoire than *Peter and the Wolf*. It would also be beneficial to look for programmatic music in non-Western music traditions and test fragments of this with, for example, Swedish participants who have never had contact with that type of music.

Finally, regarding the ideophones and fictive words tasks, results showed that there were no differences between either cultural groups or tasks. The first result was expected but the second was not, as ideophones are typically regarded as involving “conventional sound symbolism” (and secondary iconicity), while fictive words such as those used in the study display “universal sound symbolism” (and at least some primary iconicity). Still, as all the ideophones chosen for the study involved unimodal, sound-to-sound mappings, while the fictive words displayed cross-modal iconicity, this could perhaps have been a confound to be avoided in future studies.

Chapter 6. Conclusions

This thesis investigated referential iconicity in speech and programmatic music. The study of non-arbitrariness in speech, often under the label of "sound symbolism" is not new to the fields of linguistics and cognitive semiotics. However, the linkage between iconicity in speech to iconicity in music has not been widely studied, especially not through empirical research.

This thesis was thus conducted within the theoretical framework of cognitive semiotics, where a vital aspect is the emphasis on the use of empirical methods, as this strengthens the understanding of complex theoretical concepts while benefiting from the analysis of such conceptual terms (Zlatev, 2015), or the *conceptual-empirical loop*, discussed in Chapter 2. In order to gather relevant empirical data, an experiment was designed, where participants were to match representamina of either musical or linguistic nature to different objects, represented either by images or words. Combining this with a first-person method in selection of the stimuli, and a first-person and second-person method when interviewing participants about the motivations for their choices, has contributed to the cognitive semiotic study of iconicity in music and speech in a number of ways. In order to close the conceptual-empirical loop, it is pertinent to restate and answer the four research questions presented in the introduction, in Chapter 1.

- RQ1: How does iconicity in music relate to speech iconicity?

As shown at the beginning of Chapter 4, there was no significant difference in the way participants perceived iconicity in speech to the way participants perceived iconicity in this specific piece of programmatic music. This could mean that the psychological processes used for the perception of iconicity in speech and in music are domain-general, meaning that the psychological (interpretive) processes involved are not limited to a single cognitive domain or semiotic system (i.e. language vs. music), but rather that they cut across these. This could be taken as an indication that not only language, but also other semiotic systems like music rest on bodily mimesis (Donald 1991; Zlatev in press). At the same time, a possible difference in referential iconicity in the two systems was indirectly uncovered, since cross-modal mappings were more effective in the case of music, while for speech, unimodal, sound-to-sound mappings have been argued to be more common, and "simpler" (Dingemanse, 2012).

The understanding of music as a semiotic system has not been explored sufficiently in previous research. This does not mean, of course, that there have been no studies that delve into the relationship between music and language (see Coker, 1972; Osmond-Smith, 1972; Lerdahl and Jackendoff, 1983; Monelle, 1991). This thesis has focused on the referential aspects of music, which could be considered as a peripheral property of music but is sufficient to validate the claim that the similarity between music and language lies not only in comparable syntactic structures. This take on the referential aspects of music may be a small, but still very relevant part out of the many possible ways we can investigate the relations between music and language.

The similar results obtained between music and linguistic tasks show that there is indeed a similarity in the way referential iconicity in music works in relation to language. These results will only help to construct a more exhaustive understanding of music as a semiotic system in relation to language and will further help us to steer future research into other possible similarities, or differences, between these two semiotic systems. Finally, and on a different note, future research should analyze music perception from a more phenomenological point of view, with a greater role for first-person methodology. This would add new layers of understanding to this basic question.

- RQ2: Is it easier for participants to recognize iconicity through unimodal rather than cross-modal representamen-object mappings?

From the results obtained from the experiment we could see that recognizing iconicity through unimodal mappings was not easier than through cross-modal mappings in the music tasks. The design of the experiment was not fully advantageous for this particular question, due to the fact that there was not the same amount of unimodal and cross-modal conditions, giving cross-modal results a possible advantage. Nevertheless, the fact that the overall results for the unimodal condition was not above chance significance shows that the recognizing iconicity through unimodal mappings was apparently harder than through cross-modal representamina, at least in the case of music. This could pose a challenge to Dingemanse's implicational hierarchy, as sound-to-sound mappings were not the most transparent ones in music tasks. This could point to a possible reorganization of the hierarchy applied to musical iconicity, where the "simplest kind of semiotic mapping" (Dingemanse, 2012: 663) could be sound to movement. As pointed out above, this would also indicate a difference between the semiotic systems. Clearly, in order to further confirm this, much more research needs to be done. Lastly, given the imbalance in the design, and the magnitude of this study, it would be necessary to perform similar experiments in the future, with a larger variety of programmatic music, and a more balanced experiment.

- RQ3: Is iconicity in programmatic music and speech, primary, secondary, or a combination of both?

Through this thesis we managed to gather that, in line with the proposal of Ahlner and Zlatev (2010), iconicity in programmatic music and speech involves a combination of both primary and secondary iconicity, but with a higher degree of secondary iconicity. This was made clear through the more and less-contrastive conditions employed in the experiment, where the more-contrastive conditions showed higher degrees of secondary iconicity, given that they were only presented with two representamina and two objects. Still, the overall results of the less-contrastive condition showed above chance significance. The positing of a *specific* sign relations between object and representamen on the basis of the perception of the iconic ground between them is a clear indicator of primary iconicity (Ahlner and Zlatev, 2010), and thus supports the proposal that the two kinds of iconicity are not mutually exclusive but may combine in an act of interpretation.

- RQ4: To what extent is referential iconicity in music and speech perceivable by members of different cultures?

Interestingly, no cultural differences were found in the perception of iconicity in either music or speech between the Swedish and Chinese speakers, which could point to universal human cognitive semiotic-capacities. Some differences concerning the linguistic tasks were found in the interviews, with Chinese speakers more often objecting to the “naming” of abstract figures. These, however, did not lead to significant quantitative differences in performance in the experiment. There are, however factors that could have affected these results, such as the fact that all Chinese speakers live in Sweden, they all speak English, and have most likely been exposed to Western music traditions. It would thus be interesting to test monolingual speakers that do not live abroad and have not been exposed to so many different musical cultures.

In sum, the study of referential iconicity in music and speech within and across sensory modalities is an important topic for cognitive semiotics and beyond, because it allows for a more concrete understanding of how members of different cultures understand and interpret music and speech. While there have been a wide range of studies touching upon the subject of iconicity in language, not so many have dealt with iconicity in music, and this thesis contributes to the further understanding of the semiotic and cultural aspects of this phenomenon, across semiotic systems. More generally, this thesis has provided a new angle to the understanding of music as a semiotic system. It is evident that it has only covered a narrow area out of all the possible connections that can be made between music and language, but nevertheless, the significant results obtained can hopefully contribute to further research on meaning making in music.

REFERENCES

Agawu, V. K. (1991) *Playing with signs: A semiotic interpretation of classic music*. Princeton: Princeton University Press

Ahlner, F., and Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies*, 38 (1/4): 298-348.

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., and Thierry, G. In review. *Sound symbolism scaffolds language development in preverbal infants*.

Bakhtin, M.M. (1981). *The dialogic imagination: Four essays.* USA: University of Texas Press.

Bernstein, L. (1976). *The Unanswered Question, Six Talks at Harvard*. Cambridge, Mass: Harvard University Press.

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 makes similar shape– sound matches, but different shape–taste matches to Westerners. *Cognition*, 126: 165–172.

Bundagaard, P.F. (2010). Husserl and language. In S. Gallagher and D. Schmicking (Eds.), *Handbook of phenomenology and cognitive science* (pp. 369-399). Dordrecht: Springer.

Coker, W. (1972). *Music and meaning: A theoretical introduction to musical aesthetics*. New York: Collier Macmillan.

Cuskley, C. (2013). Mappings between linguistic sound and motion. *PJOS,* 5(1): 39-62.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *Br. J. Psychol*, 52: 259–268.

Devylder, S. (2018). Diagrammatic iconicity explains asymmetries in Paamese possessive Constructions. *Cognitive Linguistics,* 29(2): 313-348.

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): 1-30.

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of human culture*. Cambridge, MA: Harvard University Press.

Donald, M. (2001). *A mind so rare: The evolution of human consciousness.* New York: Norton.

Donald, M. (2007). The slow process: A hypothetical cognitive adaptation for distributed cognitive networks. *Journal of Physiology,* 101: 214-222.

Editors. (1998). "Program Music." *Encyclopedia Britannica*. Web. 2 September 2018. www.britannica.com/art/program-music.

HaCohen, R. and Wagner, N. (1997). The communicative force of Wagner's leitmotifs: Complementary relationships between their connotations and denotations. *Music Perception,* 14(4): 445-476.

Ibarretxe-Antuñano, I. (2017). Basque ideophones from a typological perspective. *Canadian Journal of Linguistics,* 62(2):196-220.

Imai, M., and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the royal society B, 369.*

Itkonen, E. (2008). The central role of normativity in language and linguistics. In J. Zlatev, T. Racine and E. Itkonen (Eds.), *The shared mind: perspectives on intersubjectivity* (pp.279-305). Amsterdam: Benjamins.

Jakobson, R. (1965). Quest for the essence of language. *Diogenes,* 13: 21-38.

Johansson, N., and Zlatev, J. (2013). Motivations for Sound Symbolism in Spatial Deixis: A typological study of 101 languages. *PJOS,* 5(1): 3-20.

Keiler, A. R. (1978) Bernstein's the unanswered question and the problem of musical competence. *Musical Quarterly,* 64: 195-222

Keiler, A. R. (1981). Two views of musical semiotics. Some properties of the design and syntax of tonal music. *SML*, 151-168

Kivy, P. (2002). *Introduction to a philosophy of music.* Oxford, UK: Oxford University Press.

Kivy, P. (2007). *Music, language and cognition.* New York: Clarendon Press.

Kwon, N. (2016). Empirically observed iconicity levels of English phonaestemes. *Public Journal of Semiotics,* 7(2): 73-93.

Köhler, W. (1929). *Gestalt psychology*. New York, NY: Liveright.

Lerdahl, F., and Jackendoff, R. (1983). *A Generative theory of tonal music.* Cambridge, Mass: MIT Pres

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Louhema, K. (2018). From Unisemiotic to Polysemiotic Narratives: Translating across semiotic systems. MA Thesis, Lund University.

Lyons, J. (1963). *Structural semantics: An analysis of part of the vocabulary of Plato*. Oxford:

Blackwell.

Lyons, J. (1968). *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics,* 2 vols. Cambridge: Cambridge University Press.

Mei-Pa, C. (1969). *Guide to Chinese music.* Hong Kong: Tai Hwa Printing Factory.

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. In Proc. 35th Annual Conf. of the Cognitive Science Society, Berlin, Germany, 31 July–3 August 2013 (eds M Knauff, M Pauen, N Sebanz, I Wachsmuth), pp. 3080–3085. Austin, TX: Cognitive Science Society.

Monelle, R. (1991). *Linguistics and semiotics in music*. London: Routledge.

Nattiez, J.J. (1987). *Musicologie generale et semiologie*. Paris: Bourgois

Osmond-Smith, D. (1972). The Iconic process in musical communication. *VS, Quaderm di Studi Semiotici*, 31-42.

Ohala, J.J. 1991. The frequency code underlines the sound-symbolic use of voice pitch. In: L. Hinton, J. Nichols and J.J, Ohala. (eds.) *Sound Symbolism.* Cambridge: Cambridge University Press.

Ohala, J. J. 1997. Sound Symbolism. Proc. 4th Seoul International Conference on Linguistics [SICOL] 11-15 Aug 1997.

Patel, A. D. (2008). *Music, language and the brain*. New York: Oxford University Press.

Peirce, Charles Sanders. 1974 [1931]. The Icon, Index, and Symbol. In C. Hartshorne and P. Weiss (eds.), Collected papers of Charles Sanders Peirce. Cambridge, MA: Harvard University Press.

Peirce, C.S. (2003). Basic concepts of Peircean sign theory. In: Gottdiener, M; Boklund-Lagopoulou, K; Lagopoulos, A (eds.), *Semiotics. Vol. 1, Part one. Fundamentals: The constitution of the field.* London: Sage publications, 101-135.

"Peter and the Wolf." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. 22 July 2004. Web. March 19, 2018. en.wikipedia.org/wiki/Peter_and_the_Wolf.

Ramachandran, V.S., and Hubbard, E.M. (2001). Synesthesia- A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, 8: 3–34.

Rousseau, J.J. (1781). *Essai sur l'origine des langues. Où il est parlé de la mélodie et l'imitation musicale*. Saint Amand: Gallimard.

Saint-Exupéry, A. (1943 [1987]). *Le petit prince.* France: Gallimard.

Saeed, J. I. (2016). *Semantics*. Chichester, West Sussex; Malden, MA: Wiley Blackwell.

Saussure, F. (1959[1916]). *Course in General Linguistics*. New York: The Philosophical Library.

"Scientific Pitch Notation." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. 22 July 2004. Web. May 05, 2018. en.wikipedia.org/wiki/Scientific_pitch_notation.

Shen, S. (1991). *Chinese music and orchestration: A primer on principles and practice.* Chicago: Chinese Music Society of North America.

Shen, S. (2000). *China: A journey into its musical art.* Chicago: Chinese Music Society of North America.

Sonesson, G. (1997). The ecological foundations of iconicity. In: Rauch, Irmengard; Carr, Gerald F. (eds.), *Semiotics around the world: Synthesis in diversity.* Proceedings of the Fifth International Congress of the IASS, Berkeley, June 12– 18, 1994. Berlin and New York: Mouton de Gruyter, 739–742.

Sonesson, G. (2007). From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In T. Ziemke, J. Zlatev, & R. Frank (Eds.), *Body, Language and Mind. Vol 1: Embodiment* (pp. 85-128). Berlin: Mouton de Gruyter.

Sonesson, G. (2009). Prolegomena to a general theory of iconicity. Considerations on language, gesture, and pictures. In: Willems, K; De Cuypere, L (eds.), *Naturalness and Iconicity in Language*. Amsterdam: John Benjamins, 47–72.

Sonesson, G. (2013). The natural history of branching: Approaches to phenomenology of Firstness, Secondness and Thirdness. *Signs and Society* 1(2), 297-326.

Thompson, P.D., and Estes Z. (2011). Sound symbolic naming of novel objects is a graded function. *Q. Exp. Psychol*, 64: 2392–2404.

Tien, A. (2015). *The Semantics of Chinese Music: Analysing Selected Chinese Musical Concepts.* Amsterdam: John Benjamins.

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Phil. Trans. R. Soc. B, 369*(1651), 20130292.

Yamamoto, M. (2006). *Agency and impersonality: Their linguistic and cultural manifestations*. Amsterdam: Benjamins

Zlatev, J. (2007). Embodiment, language and mimesis. In: T, Ziemke; J. Zlatev; R. M. Frank (eds.), *Body, Language and Mind*, 1: Embodiment. Berlin: Mouton, 297–337.

Zlatev, J. (2009). The semiotic hierarchy: Life, consciousness, signs and language. *Cognitive Semiotics*, 4: 169-200.

Zlatev, J. (2012). Cognitive semiotics: An emerging field for the transdisciplinary study of Meaning. *The Public Journal of Semiotics,* 4(1): 2-23.

Zlatev, J. (2014). Human uniqueness, bodily mimesis and the evolution of language. *Humana. Mente Journal of Philosophical Studies,* 27: 197-219.

Zlatev, J. (2015). Cognitive semiotics. In Trifonas, P. (ed.), *International handbook of semiotics*, 1043–1067. Dordrecht, Netherlands: Springer.

Zlatev, J. (in press). Mimesis, learning and polysemiotic communication. In M.A. Peters (Ed.) *Encyclopedia of Educational Philosophy and Theory*. Berlin: Springer.

Zlatev, J., Zywiczynski, P., Wacewicz, S., and van de Weijer, J. (2017). Multimodal-first or pantomime-first? Communicating events through pantomime with and without vocalization. *Interaction Studies,* 18(3): 455-479.

MUSICAL REFERENCES

Djawadi, R. (2015). *Dance of Dragons.*

Prokofiev, S. (1936). *Peter and the Wolf.*

Saint-Saëns, C. (1886). *Le Carnaval des Animaux.*

Rimsky-Korsakov, N. (1899-1900). *Flight of the Bumblebee.*

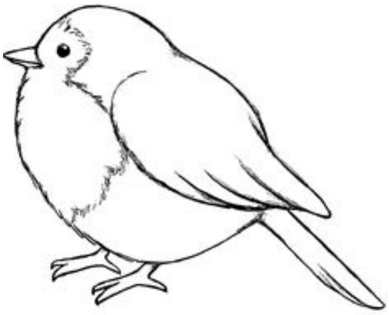
Rimsky-Korsakov, N. (1888). *Scheherazade.*

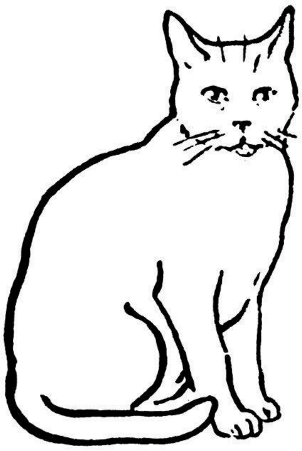
Vivaldi, A. (1721). *Four Seasons.*

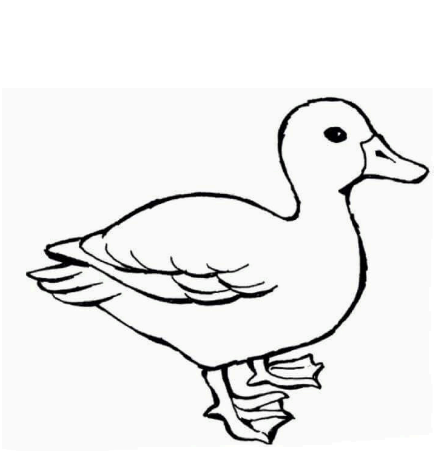
Appendix A. Images

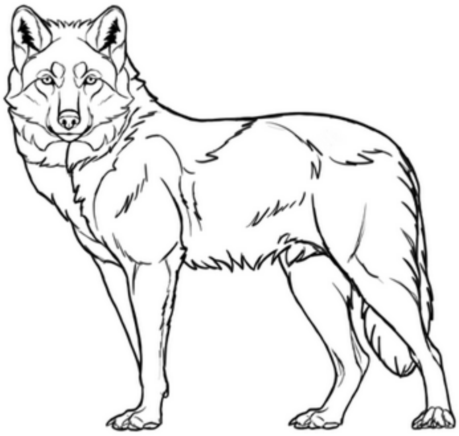
Images Music Tasks

Peter and the Wolf Characters

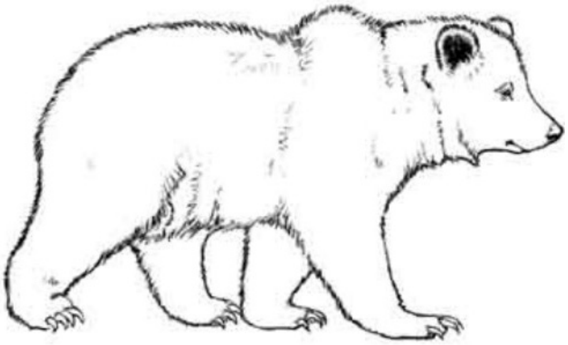


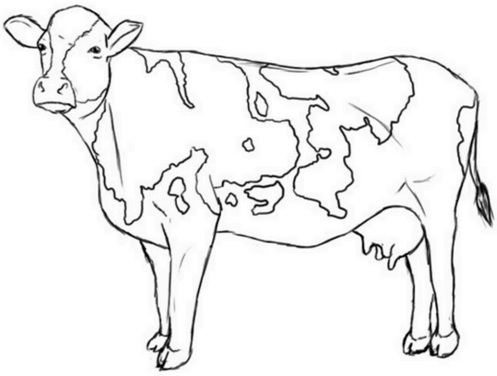


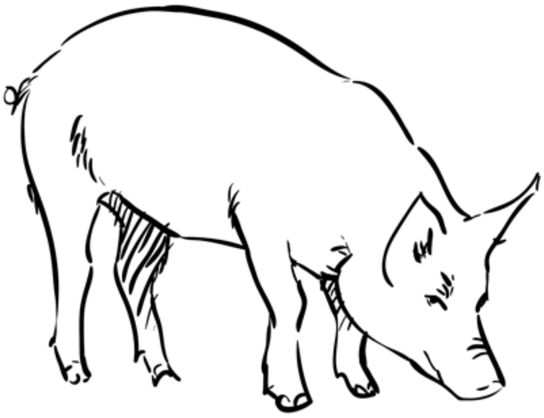




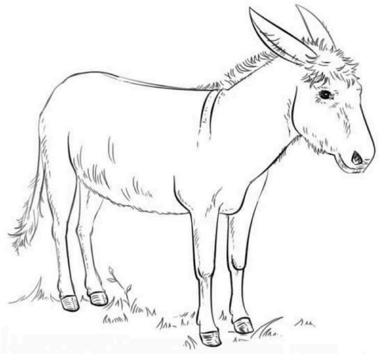
Foils

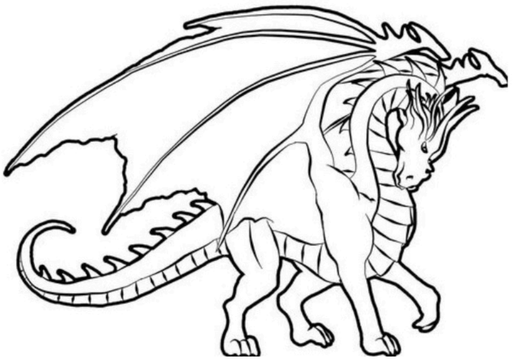


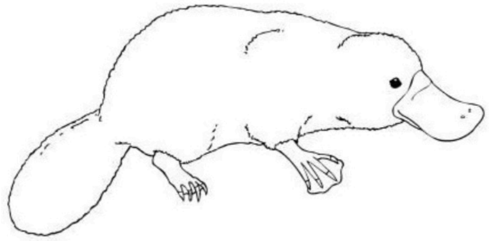




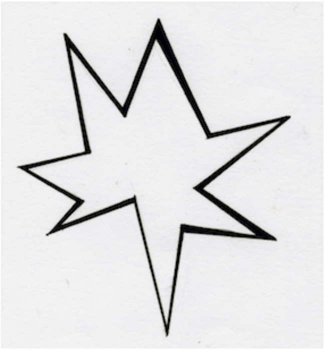
Images Warmup Music Task

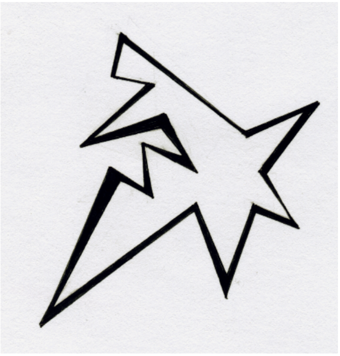


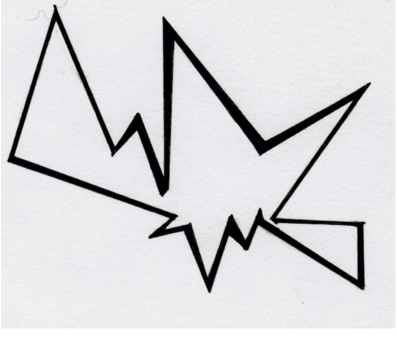


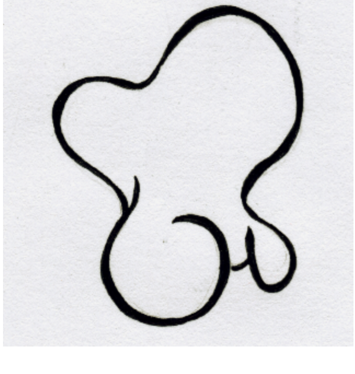


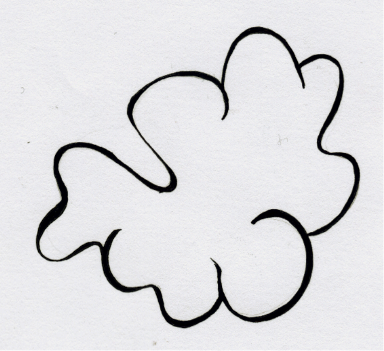
Images Fictive Words Tasks



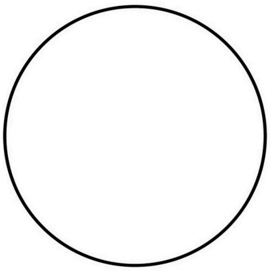


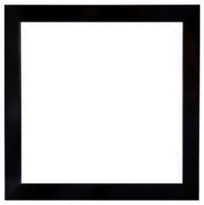






Foils





Images Warmup Fictive Words Task

Appendix B. Written Words

Words Music Tasks

Swedish:
<u>Peter and the Wolf Characters</u>: (Bird, Cat, Duck, Grandfather, Hunter and Wolf respectively)

FÅGEL

KATT

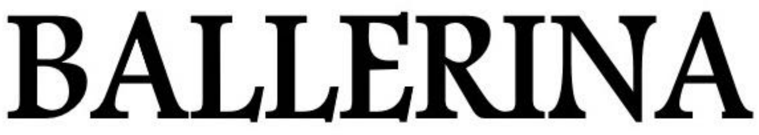
ANKA

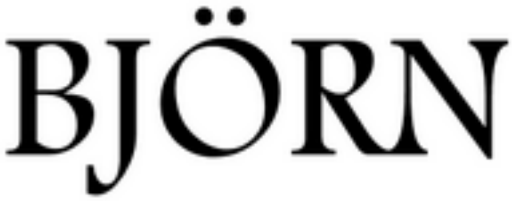
MORFAR

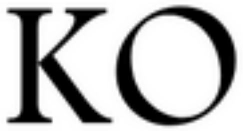
JÄGARE

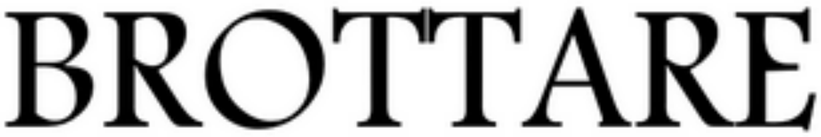
VARG

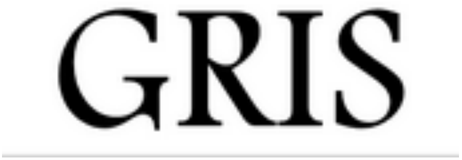
Foils: (Ballerina, Bear, Cow, Fighter, Pig and Squirrel respectively)

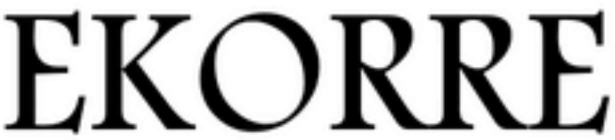

BALLERINA


BJÖRN


KO


BROTTARE


GRIS


EKORRE

Chinese:
Peter and the Wolf Characters: (Bird, Cat, Duck, Grandfather, Hunter and Wolf respectively)

鸟

猫

鸭子

外公

猎人

狼

Foils: (Ballerina, Bear, Cow, Fighter, Pig and Squirrel respectively)

芭蕾女演员

熊

牛

格斗士

猪

松鼠

Ideophones

Swedish: (Pil-pil, Grik-grak, Draka-draka, Trinkili-trankala, Zirris-zarraz and Zorro-zorro respectively)

LJUDET AV
KOKANDE
VATTEN

ATT SPRAKA

GALOPPERANDE
HÄST

RÖRA SIG
HÖGLJUTT MED
SVÅRIGHETER

LJUDET AV ATT
NÅGON SÅGAR

SNARKANDE

Foils: (To groan, to splash and whispering respectively)

ATT STÖNA

PLASK

ATT VISKA

Chinese: (Pil-pil, Grik-grak, Draka-draka, Trinkili-trankala, Zirris-Zarraz and Zorro-Zorro respectively)

烧水声

发出噼啪声

马蹄声

行动困难声

锯木声

呼噜声

Foils: (To groan, to splash and whispering respectively)

呻吟声

泼溅声

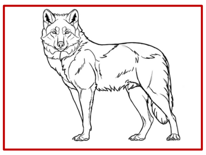
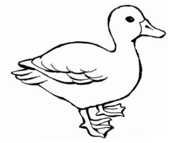
低语声

Appendix C. Organization

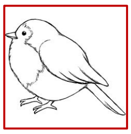
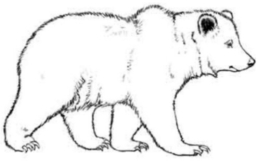
For the three versions of the experiment (in Swedish).

Version 1

T1: Music to Image (More-Contrastive)

T1: Music to Image (Less-Contrastive)

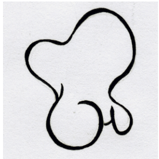
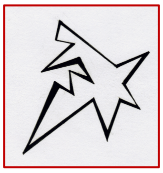
T2: Music to Word (‘grandfather’ and ‘cat’ respectively) (More-Contrastive)

MORFAR KATT

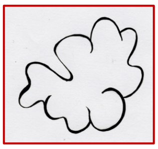
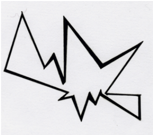
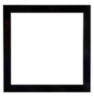
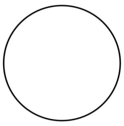
T2: Music to Word (‘pig’, ‘cow’, ‘fighter’ and ‘hunters’, respectively) (Less-Contrastive)

GRIS KO BROTTARE JÄGARE

T3: Fictive word to Shape (More-Contrastive)

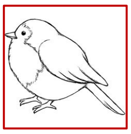
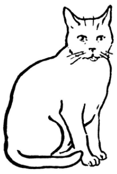
T3: Fictive word to Shape (Less-Contrastive)

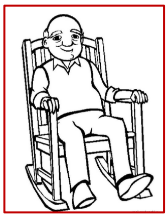
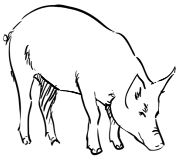
T4: Ideophones to Word ('sound of sawing' and 'sound of boiling water') (More-Contrastive) LJUDET AV ATT NÅGON SÅGAR [✓ LJUDET AV KOKANDE VATTEN]
T4: Ideophones to Word ('splash', 'to whisper', 'to groan' and 'galloping horse' respectively) (Less-Contrastive) PLASK ATT VISKA ATT STÖNA [✓ GALOPPERANDE HÄST]

Version 2

T1: Music to Image (More-Contrastive)

T1: Music to Image (Less-Contrastive)

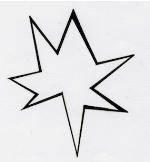
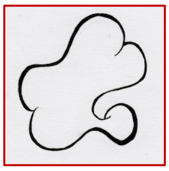
T2: Music to Word (‘hunter’ and ‘wolf’ respectively) (More-Contrastive)

JÄGARE VARG

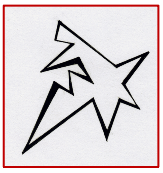
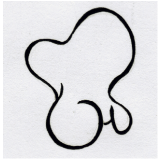
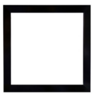
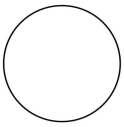
T2: Music to Word (‘cow’, ‘duck’, ‘bear’ and ‘squirrel’ respectively) (Less-Contrastive)

KO ANKA BJÖRN EKORRE

T3: Fictive word to Shape (More-Contrastive)

T3: Fictive word to Shape (Less-Contrastive)

T4: Ideophones to Word (‘galloping horse’ and ‘sound of boiling water’ respectively) (More-Contrastive)

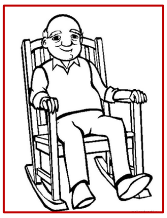
GALOPPERANDE HÄST [✓ RÖRA SIG HÖGLJUTT MED SVÅRIGHETER]

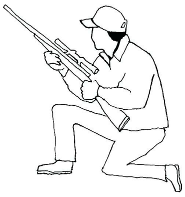
T4: Ideophones to Word (‘splash’, ‘to whisper’, ‘to groan’ and ‘to crackle’ respectively) (Less-Contrastive)

PLASK ATT VISKA ATT STÖNA [✓ ATT SPRAKA]

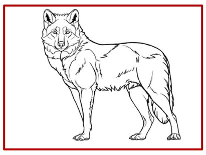
Version 3

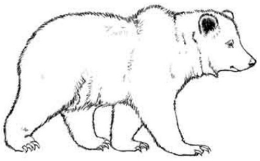
T1: Music to Image (More-Contrastive)

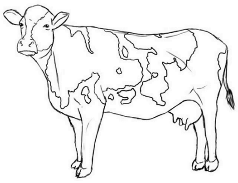




T1: Music to Image (Less-Contrastive)







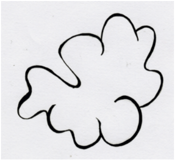
T2: Music to Word (‘duck’ and ‘bird’ respectively) (More-Contrastive)

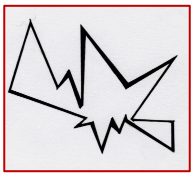
ANKA FÅGEL

T2: Music to Word (‘pig’, ‘cat’, ‘ballerina’ and ‘fighter’ respectively) (Less-Contrastive)

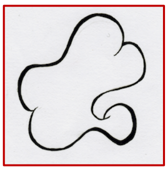
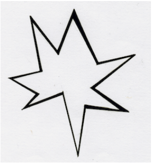
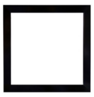
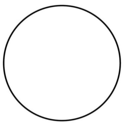
GRIS KATT BALLERINA BROTTARE

T3: Fictive word to Shape (More-Contrastive)





Speech/ Non-Words/ Less-Contrastive (Target: Lamu)

T4: Ideophones to Word ('to crackle' and 'snoring' respectively) (More-Contrastive)

ATT SPRAKA

SNARKANDE

T4: Ideophones to Word ('splash', 'to whisper', 'to groan' and 'sound of sawing' respectively) (Less-Contrastive)

PLASK

ATT VISKA

ATT STÖNA

LJUDET AV ATT NÅGON SÅGAR