Iconicity in ideophones

A cross-linguistic, cross-methodological approach to operationalising iconicity

Bachelor's thesis

Taalwetenschap, Faculty of Arts

Radboud University

Stella Punselie – S1011072

Date: 17 July 2020

Supervisor: Dr Mark Dingemanse

Second reader: Dr Laura Speed

Acknowledgements

This thesis, 'Iconicity in ideophones: a cross-linguistic, cross-methodological approach to operationalising iconicity', was written as part of the Bachelor's programme in linguistics (Taalwetenschap) at Radboud University. This thesis would not have been the same without the support of a number of people.

Firstly, many thanks to my supervisor, Mark Dingemanse, for his guidance in the past few months, for sharing his data on ideophones with me, for our interesting conversations, and for generally being such an inspiring researcher to work with.

I would also like to thank Laura Speed for taking the time to be the second reader of this thesis. Additionally, I am grateful to everyone who participated in my research, for without them, there would have been no results to report.

Finally, my interest in ideophones was first sparked when I took a course on language typology at Leiden University in 2019. For this, I would like to thank Victoria Nyst, who taught this course. Had she not introduced me to the fascinating world of ideophones, this thesis would have turned out very differently.

Stella Punselie
Nijmegen, 17 July 2020

Table of contents

Abstract

Ideophones, or marked words depicting sensory imagery, often make use of iconic mappings that connect aspects of form and aspects of meaning through perceptual or motor analogies. This is why ideophones form an excellent resource for the study of iconicity in natural language. This thesis investigates how Dutch native speakers rate the iconicity of 240 ideophones from five languages unknown to them: Japanese, Korean, Ewe (a Kwa language spoken in Ghana and Togo), Siwu (a Na-Togo language spoken in Ghana), and Semai (a Mon-Khmer language spoken in Malaysia). These ideophones have been divided into different semantic categories related to sound, motion, texture, shape and colour/visual appearance. It is shown that all categories differ significantly from one another in terms of iconicity ratings, with ideophones in the domain of sound being rated as the most iconic and ideophones referring to colour/visual appearance as the least iconic. Furthermore, the iconicity ratings are compared with the results of two forced-choice experiments involving the guessing of meanings of the same set of ideophones. There is a strong positive correlation between the guessing performances and iconicity ratings: in general, ideophones that are given high ratings in the current study have also been guessed above chance in the previous experiments. This implies that people make use of similar intuitions about the concept of iconicity in these two methods. Humans are apparently quite adept at distinguishing the more iconic words from the less iconic, even in languages unknown to them, which supports the idea that iconicity is a universal feature of language. The results of this study form a solid basis for further research about which aspects of form and meaning exactly contribute to iconicity, and if the iconicity of words can be predicted in this way. Combining these different operationalisations can ultimately lead to a better understanding of the concept of iconicity and how sensory information is incorporated in human language.

1. Introduction

While human languages contain many arbitrary words, a substantial part of our vocabularies can also be considered iconic, with aspects of form resembling aspects of meaning. Iconic words appear all around the world, even more so in languages that are rich in a special type of depictive words named 'ideophones'. An example of a highly iconic ideophone is the Japanese *hisohiso* 'whispering', as its form resembles its meaning in a number of ways. Firstly, the form and meaning are both in the modality of sound. Secondly, the form features reduplication, which resembles the repetitive and/or ongoing aspect of whispering. Finally, the form contains the vowel [i], often associated with small and light things (e.g., Westermann, 1927; 1937; Ohala, 1994) and resembling the lightness of whispers.

Looking at structural resemblance-based relations between form and meaning like in the example above can help us understand what properties make some words more iconic than others. But to this end we also need to know which words are in fact perceived as iconic by language users. One way to do this is by letting people guess the meanings of possibly iconic words like ideophones: the more accurate people are at guessing their meanings, the more iconic they are presumed to be. Another approach is by looking at how people directly rate the iconicity of a given ideophone when presented with their (spoken) form and translation.

Iconcitiy is not a strict "monolithic construct" (Motamedi, Little, Nielsen, & Sulik, 2019:9), and thus we cannot simply assume that these different methods will measure the exact same concept. This is why combining these three operationalisations – i.e., iconic mappings, guessability, and iconicity ratings – can ultimately improve our understanding of the concept of iconicity, by enabling us to predict which words are more iconic than others and explain why that is the case. While comparing all three operationalisations is outside the scope of this thesis, I will make a start by collecting iconicity ratings from a linguistically diverse set of ideophones, as well as comparing these ratings to how well the same ideophones were guessed in two previous studies. This will show if high guessability can indeed be trusted to reflect high iconicity, and thus forms a solid basis for predicting iconicity.

The rest of this introduction is constructed as follows. §1.1 serves to define the two key concepts of this paper, i.e., iconicity and ideophones. §1.2 presents an overview of the research that has been conducted on iconicity in ideophones. Finally, the goals and hypotheses of the current study are described in §1.3.

1.1 Defining the terms

1.1.1 Iconicity

The term *iconicity*, originated by Charles S. Peirce in his theory of signs, essentially involves a resemblance-based relation between certain content and the way this content is expressed. Iconic relations can be found in many things in the world around us, whether it is a traffic sign, a drawing, or a linguistic sign. Here, I define iconicity in language as a resemblance-based mapping or perceptuomotor analogy between aspects of linguistic form and meaning (cf. Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Emmorey, 2014; Perniss, Thompson, & Vigliocco 2010; Perniss & Vigliocco, 2014). Apart from Peircean semiotics, this definition also relates to structure-mapping theory, which describes how analogies are characterised by the relations between objects rather than by the attributes of objects themselves (Gentner, 1983; Gentner & Markman, 1997).

Iconicity is present in signed as well as spoken languages. Iconic signs in sign languages are highly frequent (Caselli, Sevcikova Sehyr, Cohen-Goldberg, & Emmorey, 2017) and possibly more transparent than in spoken language (Perlman, Little, Thompson, & Thompson, 2018). It is therefore not surprising that studies on sign languages have contributed to the acceptance of iconicity as a general property of human language, complementing its more arbitrary aspects (Lockwood & Dingemanse, 2015; Perniss et al., 2010; Perniss and Vigliocco, 2014). While iconic relations have also been described for grammatical structures in signed (e.g., Meir, Padden, Aronoff, & Sandler, 2013) and spoken languages (e.g., Diessel, 2008; Givón, 1985; Haiman, 1980; 2008; but see Haspelmath, 2008), the focus in this thesis is on lexical iconicity in spoken language.

The term *iconicity* often co-occurs with the term *sound symbolism*. Some see iconicity as an umbrella term for all form-meaning resemblances, and sound symbolism as a specific type of iconicity for spoken languages that can be divided into various subtypes (Hinton, Nichols, & Ohala, 1994; Lockwood & Dingemanse, 2015). Other studies describe sound symbolism simply as a non-arbitrary relationship between sound and meaning (Imai, Kita, Nagumo, & Okada, 2008) and including iconicity (Lockwood, Hagoort, & Dingemanse, 2016). For want of a uniform definition of sound symbolism, I will here refrain from using this term and instead choose to write about iconicity.

Iconicity goes further than simple imitation. Therefore, researchers have made distinctions between various types of iconicity. Firstly, there is the most straightforward, noticeable type of iconicity, namely *imagic iconicity* (Peirce, 1932 [1974]; Haiman, 1985; Johansson, Anikin, Carling, & Holmer, 2020), which is when a linguistic form directly imitates a real-world sound. This type is also referred to as *imitative sound symbolism* (Hinton et al., 1994) or *absolute iconicity* (Dingemanse et al., 2015) and includes onomatopoeia. Examples from English are *meow* for the sound a cat makes, or *bang* for the sound of an explosion.

Imagic iconicity is unimodal: the meaning is in the domain of sound, and so is the spoken form. But iconicity can also occur across sensory modalities; this is called *diagrammatic iconicity* (Dingemanse, 2011; Johansson et al., 2020). One type of diagrammatic iconicity is *Gestalt iconicity* (Dingemanse, 2011; Johansson et al., 2020). This refers to a resemblance between the structure of a linguistic form and the structure of the event designated by this form. An example of Gestalt iconicity is found in reduplication (Johansson et al., 2020), as reduplicated forms are often associated with repeated events, such as the Swahili *piga-piga* 'to strike repeatedly' (cf. *piga* 'to strike') (Lodhi, 2002:8).

The third type of iconicity, *relative iconicity*, refers to a contrast between linguistic forms which is associated with a contrast between their meanings. This idea is present in Ohala's (1994) frequency code (Johansson et al., 2020). This theory assumes that pitch is related to body size, so that sounds with a high frequency (produced by small animals) are associated with small things and low-frequency sounds (produced by large animals) are linked to large things. Relative iconicity is also found in the lexicon, such as the Japanese *ton ton* for 'light rhythmic sounds' versus *don don* for 'heavy rhythmic sounds' (Hamano, 1986:109), where the voiceless versus voiced stops represent the contrast in heaviness between the two meanings.

1.1.2 Ideophones

Ideophones are a type of conventionalised words that are often viewed as iconic, and thus constitute a particularly rich environment for studying iconicity in natural spoken language. They are words that convey expressive, sensory meanings by means of 'dramaturgic depiction' (Ibarretxe-Antuñano, 2017), forming a particularly interesting incorporation of sensory perception in human language.

Ideophones have been studied since the 19th century (for a historical overview, see Dingemanse, 2018), at first mostly in African languages. But they went by many different names, like *Lautbilder* 'sound pictures', later called 'picture words' (e.g. Westermann, 1927, 1937; Wundt, 1920, as cited in Dingemanse, 2018), or 'echoisms' (Smith, 1920; as cited in Dingemanse, 2018). The term *ideophone* was introduced by Doke (1935) in his description of Bantu languages and gradually spread to other parts of the world, and it is the most common term at the moment. Two other commonly used terms are *mimetic* and *expressive*, for Japanese and South-East Asian languages respectively; while these words may behave somewhat differently, for example grammatically speaking, it is assumed they have enough in common for cross-linguistic comparison (e.g., Dingemanse, 2018, Samarin, 1970). Therefore, I follow many other studies (e.g., Ahlner & Zlatev, 2010; Akita & Dingemanse, 2019; Voeltz & Kilian-Hatz, 2001) in using *ideophone* as a blanket term including mimetics and expressives.

Over time, ideophones have mostly been described according to by Doke's (1935) long-standing definition. Since this definition is limited to Bantu languages, a recent definition was proposed by Dingemanse (2011): *marked words that depict sensory imagery*. As this definition is general enough to be applied cross-linguistically while leaving room for language-specific details, I choose to adopt this definition here. Ideophones are described as 'marked' as they stand out by their often remarkable phonotactics, expressive morphology and syntactic independence (e.g., Akita & Dingemanse, 2019; Childs, 1994; Dingemanse, 2012). They are often considered to constitute their own lexical word class, often an open and productive one, but how they fall into the different grammatical word classes depends on the language (Ameka, 2001; Childs, 1994). In Ewe, a Kwa language of Ghana and Togo, ideophones occur in many different word classes (Ameka, 2001), while in Pastaza Quechua of Ecuador and in Emai, an Edo language of Nigeria, they are seen as a type of adverbs (Nuckolls, 2001; Schaefer, 2001, respectively).

A well-known type of ideophones is onomatopoeia, but ideophones can cover a much broader range of meanings than just sounds, including manner, colour, smell and state (Doke, 1935:118). In fact, the meanings or 'depictions' have been said to have a hierarchical structure. A recent example is the implicational hierarchy by Dingemanse (2012): SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES. This hierarchy implies that if a language has ideophones, it will always have ideophones that depict sound (or onomatopoeia). Moreover, if there are ideophones depicting motion in a given languages, there will also be ideophones depicting sound, and so forth. This hierarchy regards sound as the 'basic' category, which is in line with the idea of direct iconicity as explained in section 1.1.1.

As with iconicity, ideophones have not always been granted the attention they deserve. Ideophones are very widespread (e.g., Kilian-Hatz, 2001; Nuckolls, 2004), but because they are "conspicuously undeveloped and poorly structured in the languages of Europe" (Diffloth, 1972:440), the historical Eurocentric focus (Perniss et al., 2010; Ibarretxe-Antuñano, 2017) made that ideophones were not seen as 'real' or conventional words but rather as "playthings, not the tools of language" (Müller, 1895; as cited in Dingemanse, 2011:2). However, an increasing focus on non-Indo-European languages and cross-linguistic perspectives have brought a renewed interest in the study of ideophones across the globe (Dingemanse, 2012).

Finally, it is important to note that, while ideophones are known for their many iconic elements, they do not simply equal 'iconic words'; iconicity comes in degrees, and some ideophones are more iconic than others. What exactly underlies this is the subject of ongoing research.

1.2 Iconicity in ideophones: an overview

Iconicity in general, and in ideophones specifically, has been studied in various ways. I here highlight some of the important works from three main research types, which I call 'descriptive approaches', 'behavioural experiments', and 'rating studies'.

1.2.1 Descriptive approaches

From the first half of the 20th century, researchers started to systematically study the resemblance between aspects of form and aspects of meaning in various languages. Jespersen (1922 [1933]), for example, investigated the general tendency of words containing the vowel [i] to indicate small or weak things. This he illustrated with long lists of words from a handful of languages, mostly Indo-European, but also including Japanese, Chinese and Greenlandic. Westermann (1927; 1937) used the rich inventories of ideophones in a number of West-African languages to show how features like tone, vowel quality and consonant voicing can systematically map onto certain aspects of meaning. For instance, low tone, voiced consonants and 'dark' vowels ([o ɔ u]) are often associated ideophones with big, heavy, dark and/or soft meanings. Conversely, light, hard and small things are often expressed by ideophones with high tone, voiceless consonants and "light" vowels ([e ɛ i]).

Diffloth (1972), in his description of ideophones in Semai and Korean, also acknowledges the existence of 'non-arbitrary' form-meaning relationships within and across modalities (he calls this acoustic and articulatory symbolism, respectively). However, he argues that this is difficult to recognize because of the articulatory complexity, with each articulatory aspect being a potential basis for this symbolism.

More recent studies have tried to break down this complexity from cross-linguistic perspectives by systematically analysing the articulatory features underlying sound-meaning associations. Blasi et al. (2016), for example, managed to do this with a list of frequent words from for over 6,000 languages. Johansson et al. (2020) conducted a similar study, but with a finer phonetic granularity, capturing certain contrasts like consonant voicing that were absent in Blasi et al. (2016). However, these studies mostly focused on statistical relationships rather than perceptuomotor analogies between form and meaning, making them more about systematicity than iconicity (Dingemanse et al., 2015). Thompson and Do (2019) aim to clearly distinguish iconicity from systematicity. They developed a methodology to explain the phonological structure of ideophones and capture their uniquely iconic features, which they propose are derived from oral articulatory gesture. Although this method has so far only been illustrated for ideophones from one language (i.e., Chaoyang, a Southern Min language of China), it has potential to also be applied cross-linguistically.

1.2.2 Behavioural experiments

Findings from descriptive approaches as mentioned above have been used to support or inspire behavioural experiments, one goal of which has been to find out if humans are actually sensitive to these widely assumed form-meaning resemblances. Various studies have related acoustic properties of vowels and/or consonants to size (Sapir, 1929; Newman, 1933), shape (the *bouba-kiki* paradigm, Köhler, 1947; Ramachandran & Hubbard, 2001) brightness (Newman, 1933) and motion (Cuskley, 2013) or a combination of properties, as in Fischer-Jørgensen (1978), where participants had to match Westermann's (1927; 1937) vowel categories to Danish adjective pairs.

What many of these experiments have in common is that they involve forced-choice paradigms where participants are asked to choose their answer from a number of alternatives, e.g., when matching a number of (non-)words with meanings they consider the most suitable. Participants in these forced-choice experiments overwhelmingly performed above chance level, suggesting that humans are indeed sensitive to iconicity. This has even been taken a step further by showing that iconic words may facilitate aspects of language learning (e.g., Imai et al., 2008; Maurer, Pathman, & Mondloch, 2006; see Nielsen & Dingemanse, 2020 for a critical review). However, there is the issue of validity, since most of these studies used nonwords which cannot be assumed to reflect natural language. The construction of what researchers assume to be 'iconic' words thus constitutes a possible source of experimenter bias. How can we know if iconicity also plays a role in natural language?

This is where ideophones prove very useful, as they provide a 'natural laboratory' for the study of iconicity in language (Dingemanse et al., 2015). Various forced-choice experiments have been carried out with ideophones, mostly from Japanese. Many of these studies show that people with no knowledge of Japanese can in fact accurately guess the meaning of ideophones at above-chance level (e.g., Oda, 2001; Lockwood, Dingemanse, & Hagoort, 2016a), or have similar perceptions of their semantic dimensions as native speakers (Iwasaki, Vinson, & Vigliocco, 2007).

Lockwood et al. (2016a) showed that Japanese ideophones are easier to learn for Dutch speakers when paired with their correct Dutch translation than with their opposite translation, whereas this learning effect is absent for non-ideophonic adjectives. Moreover, the participants could guess the Japanese ideophones' correct meanings at 72% accuracy, although it should be noted that contrasts between meanings were maximised by providing the opposite meaning as the only alternative. Lockwood, Hagoort, and Dingemanse (2016b) replicated this experiment with a prosodically and linguistically more homogenous set of Japanese ideophones and extended it by also looking at participants' brain activity. The findings closely replicated Lockwood et al. (2016a); additionally, they found notable individual differences in sensitivity to iconic sound-meaning mappings.

Dingemanse, Schuerman, Reinisch, Tufvesson, and Mitterer (2016) investigated Dutch speakers' ability to guess the meanings of ideophones from different semantic categories (sound, motion, shape, visual appearance, and texture) from a total of five languages: Ewe (Kwa, Niger-Congo), Japanese, Korean, Semai (Mon-Khmer, Austroasiatic), and Siwu (Na-Togo, Niger-Congo). This was also a binary forced-choice task, but in contrast to Lockwood et al. (2016a; 2016b) the alternative meaning was not the opposite of the correct option, but a related meaning from the same semantic domain. This may have contributed to the fact that overall success rates were slightly lower than in the previously mentioned experiments, but ideophones referring to sounds were rated significantly higher than ideophones from the other categories. Dingemanse et al. also looked at the possible role of prosody in guessability, concluding that segmental and suprasegmental (i.e., prosodic) information are equally important in guessing the meanings of ideophones.

These past experiments with ideophones have shown that people are also sensitive to iconic form-meaning mappings in natural language, but this sensitivity is not as high as has been claimed on the basis of maximally contrastive (nonword) experiments (Dingemanse et al., 2016).

1.2.3 Rating studies

While the forced-choice paradigms illustrated above are very popular in iconicity research, it is important to not limit ourselves to this type of research, as it primarily reflects people's unconscious reliance on iconicity. A way of determining if people can also consciously make judgements of iconicity is by having them rate iconicity on a scale. Such rating experiments are already common in sign language research (e.g., Caselli et al., 2017; Vinson, Cormier, Denmark, Schembri, & Vigliocco, 2008), and their popularity in spoken language research is increasing (Motamedi et al., 2019).

Iconicity rating experiments for spoken language essentially revolve around the idea that laypeople have consistent intuitions about the form-meaning mappings that researchers would call 'iconic'. Perry, Perlman, and Lupyan (2015), for example, used circa 600 words from different parts of speech in English and Spanish and asked native speakers to rate them on a scale of -5 ('anti-iconic') to +5 ('highly iconic'). They found that interjections and onomatopoeia were rated as significantly more iconic than other parts of speech. Additionally, adjectives (which often carry sensory meanings) and English verbs (which often express manners of motion) were also rated as more iconic than nouns, function words and Spanish verbs (which do not express manners of motion).

In a similar study, using the same iconicity scale as Perry et al. (2015), Winter, Perlman, Perry, and Lupyan (2017) collected iconicity ratings for an even larger set of 3,001 English words. They also related the ratings to the words' sensory properties and found that words in the auditory and tactile domains were rated as the more iconic than words in the domain of vision, taste and smell.

While Perry et al. (2015) and Winter et al. (2017) do pay attention to sensory meanings, the languages in question – Spanish and English, both Indo-European – are still known for their largely arbitrary vocabularies (Perry et al., 2015; Vigliocco, Perniss, & Vinson, 2014). It would seem logical to also carry out rating experiments for languages with rich ideophone systems, given their highly sensory meanings and often iconic forms. However, this has so far only been done by Kwon (2018), who asked native speakers of Korean to rate 170 Korean ideophones on a scale from 1 ('not iconic at all') to 7 ('highly iconic'), and concluded that onomatopoeic ideophones were found more iconic than ideophones with cross-modal or both onomatopoeic and cross-modal meanings.

1.3 Current research

Due to the small number of rating studies that have been carried out for spoken language so far, quite little is known about if and how people actively perceive iconicity in spoken language. Moreover, in the rating studies described in section 1.2.3, all raters were native speakers. To get an idea of how iconicity in a language is perceived by its own speakers is certainly informative when linking it to, say, age of acquisition, as in Perry et al. (2015). However, if one wishes to investigate iconicity as a cross-linguistic phenomenon, it would also be interesting to have naïve raters who are unfamiliar with the language in question. This way, we can learn more about which parts of iconic relations are language-specific and which parts are possibly universal (Dingemanse et al., 2015; Thompson & Do, 2019).

The current study is a rating study with a cross-linguistic focus. Because of their natural degree of expressiveness and iconicity, it exclusively involves ideophones from a varied set of languages. The ratings are collected via an online survey platform, from native Dutch speakers who do not speak or understand any of the languages in question. I use the same ideophones as in the experiments by Dingemanse et al. (2016) and Lockwood et al. (2016b), which come from a total of five languages (Ewe, Japanese, Korean, Semai and Siwu), and belonging to different semantic categories, including sound, motion, visual appearance, shape and texture.

One advantage of using the ideophones from previous studies is that they have already been recorded, transcribed and translated to English and Dutch, which is very efficient, given the limited scope for this research. A second advantage is the possibility to directly compare the results from the current study with the results from these previous experiments. The experiments in Dingemanse et al. (2016) and Lockwood et al. (2016b) are mostly based on an indirect notion of iconicity, with the idea that easily guessable words equal iconic words. The current experiment takes a more direct approach, as participants are explicitly asked how iconic they find these words. Using the same dataset for two different methods of operationalising iconicity has – to my knowledge – never been done before for spoken language (it has for American Sign Language; see Sevcikova Sehyr and Emmorey, 2019). Seeing where the two methods overlap or deviate may help us to better understand what exact aspects of iconicity these approaches tap into.

1.3.1 Research question and hypotheses

The main question in this study consists of two parts: (1) how do Dutch speakers rate the iconicity of ideophones from five foreign languages with regard to semantic categories, and (2) to what extent do these ratings correlate with measures of guessability collected for the same ideophones?

With regard to (1), I expect that the ideophones in the domain of sound will be rated as more iconic than ideophones in the other semantic categories. This is based on previous experimental findings (Dingemanse et al., 2016; Kwon, 2018; Perry et al., 2015; Winter et al., 2017) as well as on the intuition that it is easier to map a sound to a meaning when there is direct, unimodal iconicity. In addition, this is the type of iconicity that participants will be most familiar with, as the Dutch language is not rich in ideophones apart from onomatopoeia.

Regarding (2), I expect to find the same patterns of results in this rating study as in the guessing experiments by Dingemanse et al. (2016) and Lockwood et al. (2016b). This is based on Sevcikova Sehyr & Emmorey (2019), who also found a strong correlation between a rating and comprehension task by non-signers in American Sign Language, and on the lack of evidence that this would be any different for spoken language. I thus expect a positive correlation between the two measures, i.e., the ideophones of which the meanings were guessed with high accuracy in the guessing studies will also be given high iconicity ratings in the current study.

2. Method

2.1 Participants

In total, 78 native Dutch speakers participated in this study. They were recruited mostly via online platforms. Before starting the experiment, all participants agreed to their data being stored and processed for research purposes in line with Radboud University's data management regulations. Two participants turned out to have knowledge of one or more of the languages in the survey (i.e., Japanese and Korean). I decided to exclude these participants, in line with Dingemanse et al. (2016), where participants also had no knowledge of the languages included in the study.

For the remaining 76 participants, I calculated person-total correlations (Curran, 2016; Donlon & Fischer, 1968; Motamedi et al., 2019) with a script made available by Motamedi et al. (2019). Person-total correlations can identify careless responses in online data collection by calculating "how consistent any given person is, relative to the expected patterns generated by all other persons" (Curran, 2016:12).

One participant showed a negative person-total correlation, which according to Curran (2016) indicates a careless responder rendering invalid data. Inspection of this participant's data showed they had indeed given the same answer to every question. I thus decided to remove this participant from the dataset, resulting in a total of 75 participants to be included in the analysis. These 75 participants (64% female, 35% male, 1% other) are between 19 and 77 years old (*M* = 37;9, *SD* = 17;6).

2.2 Materials

The materials in this study consist of 240 ideophones from five non-Indo-European languages with a sizeable class of ideophones (or expressives, or mimetics): Ewe (Kwa, Niger-Congo), Japanese, Korean, Semai (Aslian, Austroasiatic) and Siwu (Na-Togo, Niger-Congo).

The ideophones come from two previous studies: 38 Japanese ideophones come from Lockwood et al. (2016b); the other 202 ideophones are from all five languages and were used in Dingemanse et al. (2016b). In the following, these two studies will be referred to with the name of the journal they were published in, which is 'Collabra' for Lockwood et al. (2016b) and 'Language' Dingemanse et al. (2016).

The ideophones in each language belong to different semantic categories, roughly corresponding to the categories in Dingemanse's (2012) implicational hierarchy. For all languages there are ideophones in the domains of SOUND, MOTION, SHAPE, VISUAL APPEARANCE and TEXTURE. Additionally, a number of Japanese ideophones from the Collabra study do not fit these categories as they refer to e.g. other sensory perceptions or cognitive states. I have therefore categorised these ideophones as OTHER. Examples of ideophones for each semantic category and language are presented in Table 1.

For my survey, I used existing audio recordings of the ideophones, which were also used in the Language and Collabra studies, and were made available to me by Mark Dingemanse. All ideophones are spoken by a native speaker of the language in question. In the case of Semai and Siwu, the ideophones were extracted directly from field recordings.

The ideophones were quasi-randomly divided over four lists of 60, where the number of ideophones per category and language was kept as equal as possible. A full overview of the number of items per category and language for all four lists is included in Appendix A.

Table 1. Examples of ideophones from each language and semantic category.

Language	Semantic category					
	SOUND	MOTION	SHAPE	COLOUR/VISUAL	TEXTURE	OTHER
Ewe	*kpa* 'sound of a slap'	*gblɔdɔɔ* 'walking slowly, weakly'	*goroo* 'round'	*mlámlá* 'glittering'	*tsinitsini* 'smooth'	
Japanese	*hisohiso* 'whispering'	*korokoro* 'rolling'	*perapera* 'thin'	*mero-mero* 'blurred'	*nurunuru* 'slimy'	*furafura* 'dizzy'
Korean	*kkilkkil* 'giggling'	*tchuktchuk* 'continuously expanding'	*kkobulkkobul* 'chubby'	*ch'orongch'orong* 'shining eyes'	*pasakpasak* 'crispy, easily breaking'	
Semai	*korrr* 'sound of pouring water'	*gjɛlgjolgjɛlgjol* 'swinging back and forth'	*slaɟɒɒr* 'straight'	*bjuukbjɛɛk* 'bright-coloured'	*grigelrigel* 'surface with ridges'	
Siwu	*tòlontòlontòlon* 'sound of water drops'	*gbadara-gbadara* 'a drunkard's wobbling gait'	*tagbaraa* 'long'	*fututu* 'purely white'	*fiɛfiɛ* 'silky'	

2.3 Procedure

The rating experiment was conducted via the online survey platform Qualtrics (https://www.qualtrics.com). The survey started with an introduction and some background questions about the participant's age, gender and knowledge of non-Indo-European languages. Then, the term 'iconicity' was carefully explained to the participant in lay terms, including a few examples from Dutch. These instructions were partly based on previous iconicity rating experiments (Kwon, 2018; Perry et al., 2015; Winter et al., 2017). However, I also strove to optimise these instructions by making them short yet clear, and by not defining iconicity as 'when a word sounds like what it means' but rather as 'when a word and its meaning resemble one another'. This way, participants were prompted to look for resemblance in both directions. The full instructions are included in Appendix B.

After a practice item, one of the four lists containing 60 items was presented to the participant. Items were presented in randomized order so that a possible fatigue was minimised.

The lists were set to be evenly presented among the participants. This resulted in a total of 18 to 20 raters for each list, thus also for each ideophone, which is nearly twice the advised minimum of 10 raters per ideophone (Motamedi et al., 2019).

For each item, the participant listened to the audio recording of an ideophone – which could be played multiple times – and read its Dutch translation. They were then asked to rate the iconicity of this ideophone on a scale of one to five, where one stood for ‘not iconic at all’ and five for ‘highly iconic’. In contrast to Perry et al. (2015), I decided not to include a scale below zero for ‘anti-iconicity’, because in their study, the negative part of the scale was used less frequently as well as less consistently (Motamedi et al., 2019:10). The scores were given on a slider and rounded to one decimal. The experiment took around 10 to 15 minutes to complete.

2.4 Analyses

All analyses were conducted with R (R Core Team, 2020) and the R packages “emmeans” (Lenth, 2020), “car” (Fox & Weisberg, 2019), , “dplyr” (Wickham, Francois, Henry, & Müller, 2020), “gghalves” (Tiedemann, 2020), “ggplot2” (Wickham, 2016), “ggthemes” (Arnold, 2019), “plyr” (Wickham, 2011), “readxl” (Wickham & Bryan, 2019), and “viridis” (Garnier, 2018).

First, I examined how the raw ratings fell along the scale and whether they were rated consistently by looking at minimum and maximum values as well as means and standard deviations for each ideophone. I then looked at the distribution of the mean ratings for each ideophone, within and between the different semantic categories. The category OTHER, however, is quite a vague residual category that does not reflect a clear semantic domain. Additionally, there are only 17 ideophones in this category (as opposed to around 40 items for the other categories), which all come from the Collabra set of Japanese ideophones. To make the results more interpretable, I decided to exclude this category from the analyses.

To verify if the observed differences between the rest of the categories were significant, I conducted a linear regression analysis with category as the independent variable or predictor and the iconicity ratings as the dependent variable. With pairwise comparisons in the “emmeans” package (Lenth, 2020) I zoomed in on which categories exactly differed from one another.

Even though I had no *a priori* hypotheses about an effect of language on the ratings, I also examined how the ratings were distributed between the five languages and constructed a linear model with language as the predictor and iconicity ratings as the dependent variable. As the goal of these regression analyses was to confirm rather than to explore, I kept the two models separate and did not test for interactions.

Visual inspection of residual plots for category as well as language suggested no clear violation of normality, and Levene's test was not significant (p > .05), so homogeneity of variance can also be assumed. Every participant rated multiple ideophones, which strictly counts as a violation of independence; nevertheless, the circumstances of this project led to linear models still being the best option.

In order to compare the iconicity ratings with the experimental guessing scores collected in the Language and Collabra studies, I transformed the iconicity ratings to z-scores. The guessing scores had also been transformed to z-scores, in order to account for the difference in baseline performance between the two studies (in the Collabra study, participants were asked to choose between two opposite meanings, making it somewhat easier than the Language study, where the choice was between two related, but not opposite, meanings). For this comparison I performed a Pearson correlation analysis. I also looked at where the two methods deviated to see if a pattern could be found in the ideophones that scored high on one of the two axes and low on the other.

3. Results

3.1 Rating analysis

The mean rating across all ideophones was 2.95, with a standard deviation of 1.30. This includes the ideophones from the residual category OTHER (M = 2.85, SD = 1.22); however, for the reasons stated in §2.3, these ideophones were not included in the results reported below.

Some ideophones were consistently rated as highly iconic, like Siwu *kpa* 'tock, sound of dry impact' which was rated between 4 and 5 by every rater (M = 4.65, SD = .38). Two other ideophones referring to sounds were consistently rated between 3.7 and 5, namely the Semai *bukbukbuk* 'sound of bubbles in water', which was also the ideophone with the highest mean rating (M = 4.71, SD = .44), and the Japanese *gokugoku* 'gulping' (M = 4.57, SD = .45).
Other ideophones were very consistent in getting low ratings, like the Semai *plɔ̃s* 'sound of someone's breathing when sleeping', which was never rated higher than 2 (*M* = 1.33, *SD* = .43), or the Japanese *boo boo* 'fire burning', always scoring between 1 and 2.4 (*M* = 1.39, *SD* = .47).

Table 2. Ideophones with the highest and lowest rating per semantic category, with their language and mean rating in parentheses.

Category	Highest rating	Lowest rating
Sound	*bukbukbuk* 'sound of bubbles in water' (Korean, 4.72)	*plɔ̃s* 'sound of someone's breathing when sleeping' (Semai, 1.33)
Motion	*tugŭndugŭn* 'heartbeat' (Korean, 4.61)	*yota yota* 'walking with heavy faltering steps' (Japanese, 1.58)
Shape	*tagbaraa* 'long' (Siwu, 4.48)	*baŋʔis* 'a larger, bulging shape' (Semai, 1.66)
Texture	*grigelrigel* 'surface with ridges' (Semai, 4.49)	*gowagowa* 'stiff' (Japanese, 1.72)
Colour/Visual	*hɛ̃ɛ̃ɛ̃* 'focal red' (Ewe, 1.40)	*gelegele* 'shiny' (Siwu, 3.82)

Table 2 lists the ideophones with the highest and lowest mean ratings for each semantic category. However, these extremes only represent a small part of the dataset, as only five ideophones were stuck in the lower half of the scale (never rated 3 or above), and six in the upper half (only rated higher than 3). Most ideophones varied quite a lot in their rating scores, such as the Japanese *zuratto* 'state of things being arranged in a line', ratings of which ranged from 1 to 5 (M = 2.41, SD = 1.50).

Figure 1 shows a combined boxplot and scatterplot of the iconicity ratings as a function of semantic category. The middle lines in the boxes represent the mean rating for each category, and the dots on the right represent the mean rating for each individual ideophone. Ideophones from the categories SOUND (M = 3.47, SD = 1.33) and MOTION (M = 3.18, SD = 1.24) were rated the highest on average, both with a mean rating above 3. Especially for SOUND, the density of ideophones in the upper part of the scale is quite high. However, the spread of the ratings is very large, so there are also SOUND ideophones with much lower ratings. For the categories SHAPE (M = 2.87, SD = 1.28), TEXTURE (M = 2.68, SD = 1.21), and COLOUR/VISUAL (M = 2.45, SD = 1.21), iconicity ratings are mostly below 3 and the spread is smaller, with concentrations in the lower part of the scale.

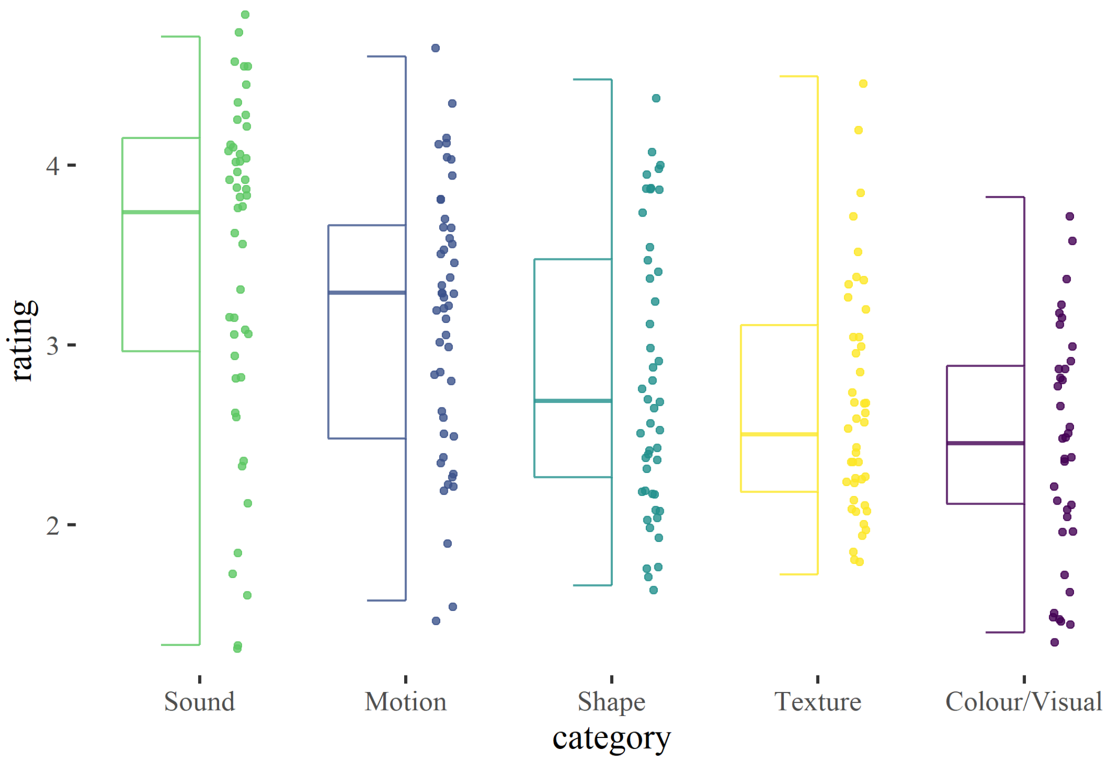

Figure 1. Combined boxplot and scatterplot showing the distribution of iconicity ratings as a function of semantic category. Each dot represents an ideophone's mean rating.

Linear regression analysis with rating as the dependent variable and semantic category as the predictor confirms that the iconicity ratings varied significantly across the different semantic categories, $F(4, 4253) = 84.01$, $p < 0.0001$, $R^2 = 0.07$. Table 3 shows the regression coefficients for the linear model used in this analysis. Pairwise comparisons with Tukey adjustments are shown in Table 4, where the values in the triangle present the *p*-values of the differences in ratings between categories, and the diagonal contains the mean ratings for each category. The *p*-values are all lower than .05 (many even < .0001) confirming that all categories differ significantly from one another.

Table 3. Regression coefficients for the linear model with semantic category as predictor (dummy-coded with Colour/Visual as reference category) and rating as dependent variable.

	Coefficient	*SE*	*t*	*p*
(Intercept)	2.45	.04	52.11	< .0001
Motion	.72	.06	11.57	< .0001
Shape	.42	.06	6.73	< .0001
Sound	1.02	.06	16.13	< .0001
Texture	.23	.06	3.58	< .0001

Table 4. P-values for the rating differences between categories, with mean ratings per category in parentheses.

	Colour/Visual	Motion	Shape	Sound	Texture
Colour/Visual	(2.45)	< .0001	< .0001	< .0001	.0005
Motion		(3.18)	< .0001	< .0001	< .0001
Shape			(2.87)	< .0001	.0409
Sound				(3.47)	< .0001
Texture					(2.68)

Figure 2 shows the distribution of iconicity ratings as a function of language. It can be noted that the distribution of ratings is largely similar across most languages. However, Semai ideophones were rated considerably lower (M = 2.57, SD = 1.34) than the ones from Ewe (M = 3.11, SD = 1.28), Siwu (M = 3.14, SD = 1.33), Korean (M = 3.03, SD = 1.24), and Japanese (M = 2.93, SD = 1.27). Also, while Korean ideophones do not have a relatively high mean rating, their spread is much smaller than that of the other languages; Korean ideophone ratings are concentrated around the middle of the scale, with none rated around 1 ('not iconic at all').

Linear regression analysis with rating as the dependent variable and language as the predictor confirms that there is indeed significant variance between languages, $F(4, 4253) = 22.06$, $p < 0.0001$, although $R^2 = .02$ indicates that language accounts for a mere 2% of the total variance in ratings. The regression coefficients for this model are presented in Table 5. Table 6 shows the p-values of the pairwise comparisons with Tukey adjustments, indicating that the source of the variance lies mostly in the Semai ideophones, as they were rated significantly lower than in all other languages ($p < .0001$).

Table 5. Regression coefficients for the linear model with language as predictor (dummy-coded with Ewe as reference language) and rating as dependent variable.

	Coefficient	*SE*	*t*	*p*
(Intercept)	3.11	.05	67.60	< .0001
Japanese	-.18	.06	-3.07	< .001
Korean	-.08	.06	-1.28	.20
Semai	-.53	.07	-7.91	< .0001
Siwu	.03	.07	.44	.66

Table 6. P-values for the rating differences between languages, with mean ratings per language in parentheses.

	Ewe	Japanese	Korean	Semai	Siwu
Ewe	(3.11)	0.0183	0.7046	< .0001	0.9922
Japanese		(2.93)	0.4069	< .0001	0.0054
Korean			(3.03)	< .0001	0.4429
Semai				(2.58)	< .0001
Siwu					(3.14)

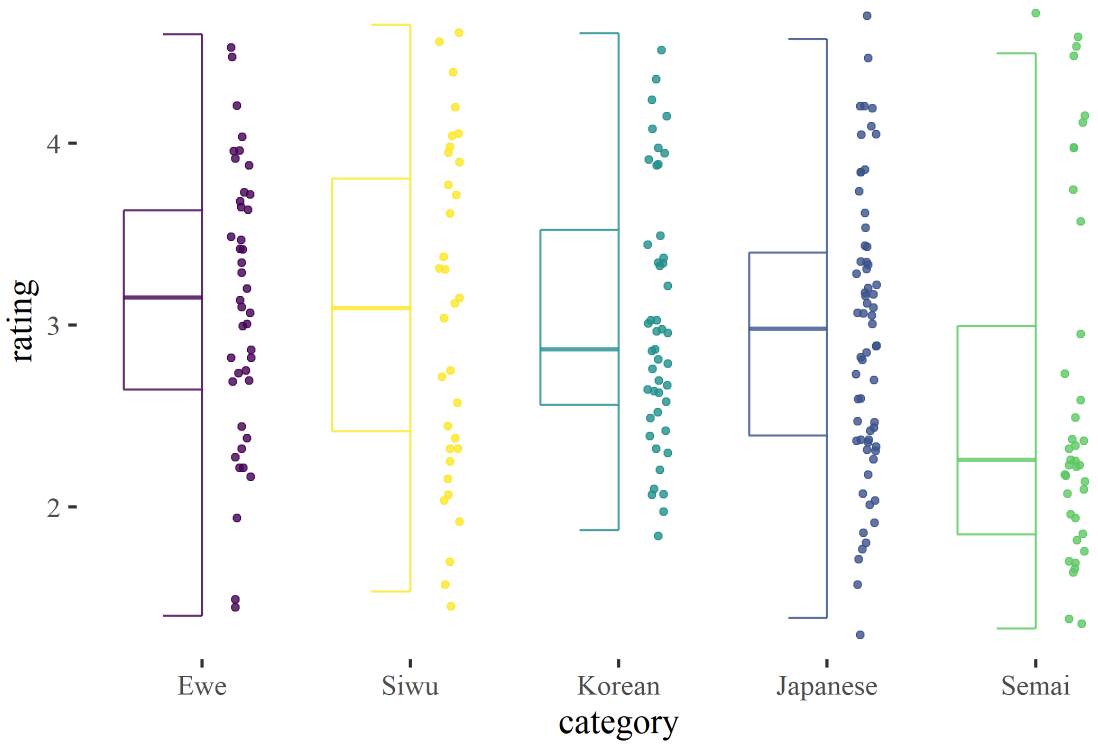

Figure 2. Combined boxplot and scatterplot showing the distribution of iconicity ratings as a function of language. Each dot represents an ideophone's mean rating.

3.2 Comparison with guessability

This section describes the relation between the iconicity ratings collected in this study and the guessing scores experimentally collected by Dingemanse et al. (2016) and Lockwood et al. (2016b). Figure 3 shows what appears to be a positive correlation between the iconicity ratings and the guessing scores, both presented as normalised z-scores. Pearson's correlation coefficient confirms that there is indeed a strong positive correlation between the rating and guessing data,

$r = .61$, $t(220) = 11.56$, $p < .0001$. This is in line with the hypothesis that ideophones rated as highly iconic in the current study were generally also highly guessable in the previous studies.

When zooming in on where the two methods deviate, it appears that some ideophones with very low ratings were in fact highly guessable. Conversely, there were no highly rated ideophones with a guessability lower than 45%. Examples from the four quadrants of iconicity ratings and guessability scores are shown in Table 7. There were only four ideophones for which the difference between the between the z-scores was larger than 2; these belong to the COLOUR/VISUAL and TEXTURE categories. For the two COLOUR/VISUAL ideophones, both from Semai, the ratings were much lower than the guessing scores, while for the two TEXTURE ideophones, from Semai and Korean, this was the other way around.

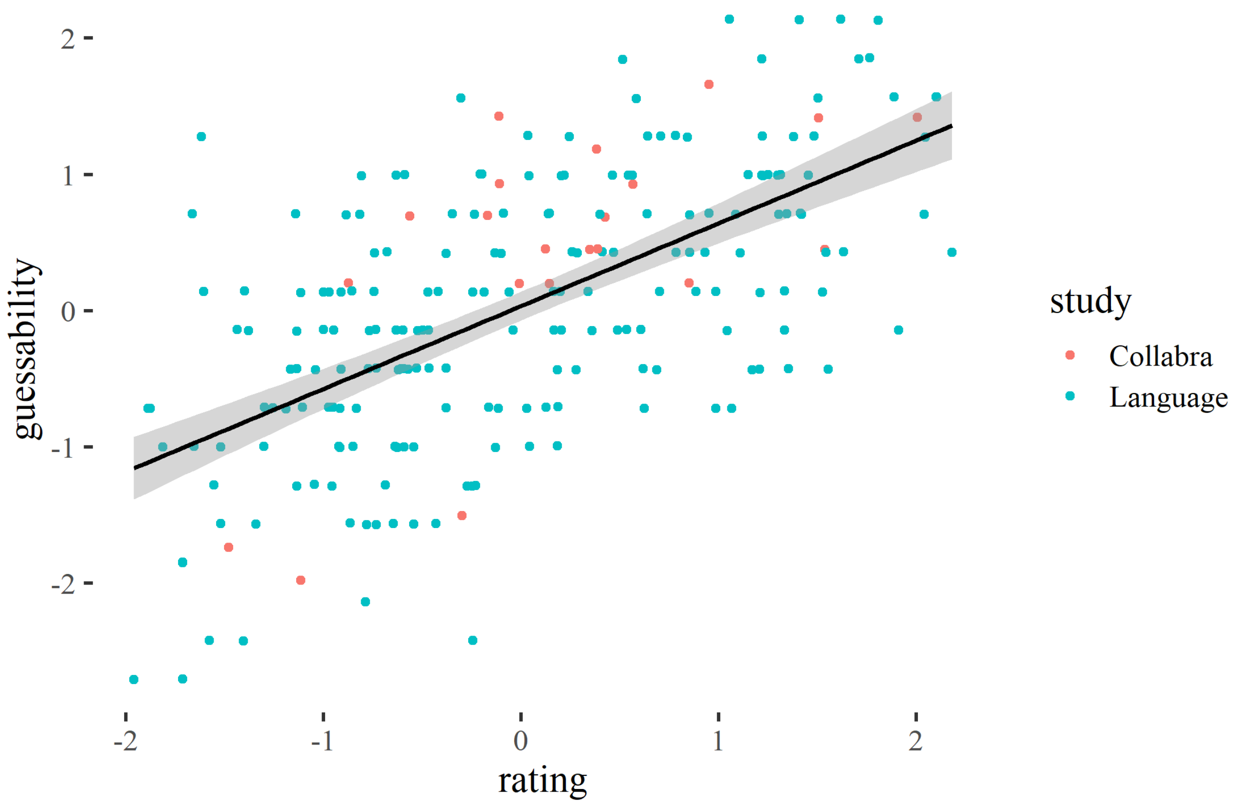

Figure 3. Relation between mean iconicity ratings and guessability scores from Dingemanse et al. (2016) and Lockwood et al. (2016b). Each dot represents an individual ideophone. The solid line indicates linear model fit, with a 95% confidence interval in the shaded area. Both axes present normalised z-scores.

Table 7. Example ideophones from the extremes of the four quadrants of iconicity ratings and guessability scores, with their language, mean iconicity rating and average guessing performance (% guessed correctly) in parentheses.

	High rating	Low rating
High guessability	- *ton ton* 'knocking on door' (Japanese, 4.41, 95%); - *kpa* 'tock, sound of dry impact' (Siwu, 4.65, 85%)	- *slʔẽẽk* 'lighter brown, pale-coloured' (Semai, 1.57, 70%); - *miɔmiɔ* 'pointy' (Siwu, 2.00, 70%)
Low guessability	- *kodzokodzo* 'walk with the body bent forward' (Ewe, 3.44, 45%); - *ɟtoonɟtoonɟtoon* 'slow up-and-down movement' (Semai, 3.74, 45%)	- *plɔ̃s* 'sound of someone's breathing when sleeping' (Semai, 1.33, 10%); - *fututu* 'purely white' (Siwu, 1.53, 10%)

4. Discussion

In this study, the perceived iconicity of 240 ideophones from five languages (Ewe, Japanese, Korean, Semai and Siwu) was rated by 75 Dutch participants with no knowledge of the languages in question. The iconicity ratings differed significantly between the semantic categories of these ideophones (SOUND, MOTION, SHAPE, TEXTURE, and COLOUR/VISUAL). On average, ideophones in the category of SOUND were rated higher than all other categories. Conversely, ideophones from the category COLOUR/VISUAL (referring to colours and visual patterns) were rated the lowest on average and could thus be regarded as the least iconic type of ideophones.

These findings are in line with the initial hypothesis that SOUND ideophones would be found the most iconic, which is in line with previous findings (e.g., Dingemanse et al., 2016; Perry et al., 2015; Winter et al., 2017). The rating differences between categories partly reflect the implicational hierarchy by Dingemanse (2012): SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES. While it should be noted that there is no one-to-one mapping of the hierarchy and the categories in this study, there is also quite some overlap. In fact, the current results almost exactly fit the slightly adjusted version of this hierarchy by McLean (2019): SOUND < MOVEMENT < SHAPE < TEXTURE < OTHER SENSORY PERCEPTIONS. The category COLOUR/VISUAL from the current study is not explicitly mentioned in this hierarchy, but it is included in the category OTHER SENSORY PERCEPTIONS.

The other categories from this study are all mirrored perfectly in this hierarchy. This suggests that the most 'basic' semantic categories of ideophones (according to the hierarchy) are also judged to be the most iconic ones, and that some sensory phenomena are apparently easier to represent with iconic mappings than others (McLean, 2019:57). This intuitively also makes sense, as it is easier to imagine a form resembling a meaning related to a sound or movement than a meaning related to a specific colour.

It would be informative to focus more on the nature of the semantic categories in future research. The category COLOUR/VISUAL includes colours like 'purple-brown' or 'purely white' as well as more dynamic visual patterns like 'glittering', and a closer look at this category might reveal a structural difference between these two subtypes. Similarly, there is still much to discover about the category OTHER, which was now excluded from the analyses as its broad, residual nature made it hard to compare it to the more interpretable semantic categories. It is very plausible that ideophones expressing sensory perceptions like smells are different from ideophones expressing emotions in terms of perceived iconicity, and yet these would both fall into the category OTHER. Breaking this category down into more seizable subfeatures could provide more insight in the iconic nature of ideophones with more complex meanings than the semantic categories defined in this research.

I had no *a priori* hypotheses about a difference between the languages in the experiment. The finding that Semai ideophones were rated significantly lower than the other languages, however, may have been due to the relatively low quality of the Semai field recordings. Another explanation could lie in the nature of the ideophones in the dataset, which contains only a part of the full ideophone inventories of the languages in question. Ideophones with meanings like 'long', which can be clearly mapped to a spoken form with acoustic features like final lengthening, only occurred in the two highest-scoring languages, Ewe and Siwu (as *lɛgbɛɛ* and *tagbaraa*, respectively, both with long final vowels and both with a high mean rating around 4). The fact that this dataset lacked Japanese, Korean and Semai ideophones meaning 'long', does not mean these languages do not have this kind of prototypically iconic ideophones.

Part of the observed differences in ratings could also be attributed to individuals' varying sensitivity to finding something 'iconic'. This was already suggested by Sapir (1929:227), who wrote that "individuals differ a good deal in the matter of sensitiveness to the symbolic suggestiveness of special sound contrasts", and has also recently been shown in neurolinguistic research by Lockwood et al. (2016b). The observation that people generally do have an idea about which words they find more iconic than others, even for languages unknown to them, supports the idea of iconicity being a universal feature of language (e.g., Dingemanse et al., 2015; Perniss et al., 2010)

A note on the nature of the rating scale is also in order. Although the current scale contained the numbers 1 to 5, it is interpreted as a more continuous scale because the ratings could take any value in between (rounded to one decimal), which forms a contrast with the more common Likert type data (Boone & Boone, 2012). However, the continuity of the scale could have been greater if the participants had only seen a slider, without the numbers 1 to 5; now, they often seemed to still conform their responses to the fixed numbers that were presented.

Despite the different sources of variation in the rating task, the strong correlation between the rating experiment and the guessing studies does suggest that these two methods measure roughly the same aspects of iconicity. In most cases, ideophones that were often guessed correctly were also given relatively high ratings, and vice versa. The few cases where the two approaches largely deviated mostly involved COLOUR/VISUAL and TEXTURE ideophones, suggesting that the concept of iconicity is slightly harder to grasp in these categories. Overall, however, the two approaches to operationalising iconicity involving guessing and rating seem to tap into the same intuitions of similarly naïve participants.

Some questions are still left open. For example, are there structural differences between ideophones that are responsible for the variation that has been found? Are there certain properties of the ideophones perceived as 'highly iconic' that are different from the ideophones perceived as low in iconicity? And could it be predicted which forms score higher than others? To address these questions, a more theoretical approach is needed where the connections between ideophones' forms and meanings are placed under scrutiny.

An ongoing study in which I am involved as part of an internship is currently looking into these questions. In this project, the same ideophones as in the present study are coded for a number of characteristics related to form and meaning by two linguistically informed researchers. The goal is to capture specific resemblance-based form-meaning associations previously attested in the literature, such as the link of vowel quality and consonant voicing to 'light' versus 'heavy' or 'dark' meanings (Westermann, 1927; 1937). The more aspects of an ideophone's form are congruent with certain aspects of its meaning, e.g. when the form is monosyllabic and the meaning involves an abrupt ending, the higher its cumulative iconicity score.

While this coding study is still in development, the first findings already show a moderate positive correlation with the ratings collected in the current study, $r(220) = .42$, $p < .001$, as well as with the experimental guessing scores, $r(220) = .36$, $p < .001$. This implies that the ideophones that were given high iconicity ratings and often guessed correctly, indeed have some structural properties that have been said to represent iconic relations. Further investigation of the mismatches between the different methods will have to shed light on the associations we might still be missing. Combining approaches of ratings, guessability and structural iconic mappings could help us predict which words will be perceived as more iconic than others, and lead to a better understanding of the concept of iconicity and how sensory information is incorporated in human language. If iconic words are indeed easily learnable, as has been suggested in various studies (e.g., Imai et al., 2008; Lockwood et al., 2016a), these insights could also have implications for studies of language learning.

5. Conclusion

This study of iconicity in ideophones can be considered unique in two ways: (1) iconicity ratings have never been collected for ideophones from different languages that were unknown to the raters, and (2) it is the first time a rating study is directly linked to a forced-choice method of guessing meanings using the same set of words. I aimed to discover more about how iconicity is perceived by means of the following research question: *How do Dutch speakers rate the iconicity of ideophones from five foreign languages with regard to semantic category, and to what extent do these ratings correlate with measures of guessability collected for the same ideophones?* Regarding the first part of this question, iconicity ratings differed between the semantic categories of the ideophones: sound-related ideophones were found the most iconic, while ideophones referring to colours and/or visual appearances received the lowest ratings. The answer to the second part is that there is a strong positive correlation between the iconicity ratings and the previously collected guessing scores, suggesting that similar intuitions are used in guessing and rating studies and making these approaches equally applicable in the study of iconicity; it is not so much a question of which method is better than the other, but rather which method best fits one's research questions and resources. Both methods can contribute to a fuller understanding of the concept of iconicity and how humans perceive and incorporate sensory information in language. In addition, they form a sound basis for the study of predicting iconicity, which could in turn have implications for studies of language learning and processing. Finally, humans are apparently quite adept at distinguishing more iconic from less iconic words in different ways, even for languages they do not speak, which supports the idea that iconicity is a universal feature of language.

References

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

Akita, K., & Dingemanse, M. (2019). Ideophones (Mimetics, Expressives). In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. doi: 10.1093/acrefore/9780199384655.013.477

Ameka, F. K. (2001). Ideophones and the nature of the adjective word class in Ewe. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* (Vol. 44, pp. 25-48). Amsterdam: John Benjamins Publishing Company.

Arnold, J. B. (2019). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 4.2.0. https://CRAN.R-project.org/package=ggthemes

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818-10823. doi:10.1073/pnas.1605782113

Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. Journal of extension, 50(2), 1-5.

Caselli, N. K., Sevcikova Sehyr, Z., Cohen-Goldberg, A. M., & Emmorey, K. (2017). ASL-LEX: A lexical database of American Sign Language. *Behavior research methods*, *49*(2), 784-801. doi:10.3758/s13428-016-0742-0

Childs, T. (1994). African ideophones. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 178-204). Cambridge: Cambridge University Press.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4-19. doi:10.1016/j.jesp.2015.07.006

Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, *5*(1), 39-62.

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, *19*(3), 465-490.

Diffloth, G. (1972). *Notes on expressive meaning.* Paper presented at the Chicago Linguistic Society.

Dingemanse, M. (2011). *The meaning and use of ideophones in Siwu* (Doctoral dissertation, Radboud University).

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass, 6*(10), 654-672. doi:10.1002/lnc3.361

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics, 3*(1). doi:10.5334/gjgl.444

Dingemanse, M. (2019). “Ideophone” as a comparative concept. In K. Akita, & P. Pardeshi (Eds.), *Ideophones, mimetics, and expressives* (pp. 13-33). Amsterdam: John Benjamins Publishing Company.

Dingemanse, M., & Akita, K. (2016). An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese 1. *Journal of Linguistics, 53*(3), 501-532. doi:10.1017/S002222671600030X

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences, 19*(10), 603-615. doi:10.1016/j.tics.2015.07.013

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language, 92*(2), e117-e133. doi:10.1353/lan.2016.0034

Doke, C. M. (1935). *Bantu linguistic terminology*: Longmans, Green and Company.

Donlon, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group-determined item difficulties. *Educational and Psychological Measurement, 28*(1), 105-113.

Dryer, M. S., & Haspelmath, M. (2013). The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130301. doi:10.1098/rstb.2013.0301

Fay, N., Ellison, M., & Garrod, S. (2014). Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition, 22*(2), 244-263. doi:10.1075/pc.22.2.05fay

Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in psychology, 9,* 716. doi:10.3389/fpsyg.2018.00716

Fischer-Jørgensen, E. (1978). On the universal character of phonetic symbolism with special reference to vowels. *Studia linguistica, 32*(1-2), 80-90.

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression (3rd edition).* Thousand Oaks, CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Garnier, S. (2018). *viridis: Default Colour Maps from 'matplotlib'*. R Package version 0.5.1. https://CRAN.R-project.org/package=viridis

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science, 7*(2), 155-170.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. A*merican psychologist, 52*(1), 45.

Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. *Iconicity in syntax*, 187-219.

Haiman, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language*, 515-540. doi:10.1515/COG.2008.002

Haiman, J. (2008). In defence of iconicity. *Cognitive Linguistics, 19*(1), 35-48.

Hamano, S. (1986). *The Sound-Symbolic System of Japanese* (Doctoral Dissertation, University of Florida).

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics, 19*(1), 1-33. doi:10.1515/COG.2008.001

Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound symbolism*: Cambridge University Press.

Ibarretxe-Antuñano, I. (2017). Basque ideophones from a typological perspective. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, *62*(2), 196-220.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54-65. doi:10.1016/j.cognition.2008.07.015

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007). What do English Speakers Know about gera-gera and yota-yota?: A Cross-linguistic Investigation of Mimetic Words of Laughing and Walking. *Japanese language education around the globe, 17*, 53-78. 10.20649/00000380

Jespersen, O. (2010). *Selected Writings of Otto Jespersen (Routledge Revivals)*. New York: Routledge.

Johansson, N. E., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology, 1*(ahead-of-print). doi:10.1515/lingty-2020-2034

Köhler, W. (1947). *Gestalt Psychology; an Introduction to New Concepts in Modern Psychology* (Rev. ed.). Oxford: Liveright.

Kwon, N. (2018). Iconicity correlated with vowel harmony in Korean ideophones. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*(1). doi:10.5334/labphon.53

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* R package version 1.4.7. https://CRAN.R-project.org/package=emmeans

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in psychology, 6*, 1246. doi:10.3389/fpsyg.2015.01246

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016a). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(8), 1274. doi:10.1037/xlm0000235

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016b). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra, 2*(1). doi:10.1525/collabra.42

Lodhi, A. (2002). Verbal Extensions in Bantu (the Case of Swahili and Nyamwezi). *Africa & Asia, 2*, 4-26.

Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: implications for vocabulary learning. *Frontiers in Communication, 2*, 4. doi:10.3389/fcomm.2017.00004

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental science*, *9*(3), 316-322. doi:10.1111/j.1467-7687.2006.00495.x

McLean, B. M. (2019). *One form, many meanings: iconicity in phonological and semantic development* (BA Honours Thesis, Australian National University).

Meir, I., Padden, C., Aronoff, M., & Sandler, W. (2013). Competing iconicities in the structure of languages. *Cognitive Linguistics, 24*(2), 309-343. doi:10.1515/cog-2013-0010

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130299. doi:10.1098/rstb.2013.0299

Motamedi, Y., Little, H., Nielsen, A., & Sulik, J. (2019). The iconicity toolbox: empirical approaches to measuring iconicity. *Language and Cognition, 11*(2), 188-207. doi:10.1017/langcog.2019.14

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology, 45*(1), 53-75.

Nielsen, A. K., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech* (ahead-of-print). doi:10.1177/0023830920914339

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual review of anthropology, 28*(1), 225-252.

Nuckolls, J. B. (2001). Ideophones in Pastaza Quechua. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* (Vol. 44, pp. 271-286). Amsterdam: John Benjamins Publishing Company.

Oda, H. (2001). *An embodied semantic mechanism for mimetic words in Japanese* (Doctoral dissertation, Indiana University).

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge: Cambridge University Press.

Peirce, C. S. (1974). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science, 2*(8), 150152. doi:10.1098/rsos.150152

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: a comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in psychology, 9*, 1433. doi:10.3389/fpsyg.2018.01433

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology, 1*, 227. doi:10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130300. doi:10.1098/rstb.2013.0300

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one, 10*(9). doi:10.1371/journal.pone.0137147

R Core Team (2020). R: A language and environment for statistical computing. Version 4.0.0. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of consciousness studies, 8*(12), 3-34.

Samarin, W. J. (1970). Inventory and choice in expressive language. *Word, 26*(2), 153-169.

Samarin, W. J. (1971). Survey of Bantu ideophones. *African Language Studies, 12*, 130-168.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of experimental psychology, 12*(3), 225.

Schaefer, R. P. (2001). Ideophonic adverbs and manner gaps in Emai. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.), *Ideophones* (Vol. 44, pp. 339-354). Amsterdam: John Benjamins Publishing Company.

Sevcikova Sehyr, Z., & Emmorey, K. (2019). The perceived mapping between form and meaning in American Sign Language depends on linguistic knowledge and task: evidence from iconicity and transparency judgments. *Language and Cognition, 11*(2), 208-234. doi:10.1017/langcog.2019.18

Smith, E. W. (1920). *The Ila-speaking tribes of Northern Rhodesia*. London: MacMillan.

Thompson, A. L., & Do, Y. (2019). Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: a journal of general linguistics, 4*(1). doi:10.5334/gjgl.872/

Tiedemann, F. (2020). gghalves: Compose Half-Half Plots Using Your Favourite Geoms. R package version 0.1.0. https://CRAN.R-project.org/package=gghalves

Vigliocco, G., & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes, 21*(7-8), 790-816. doi:10.1080/016909600824070

Vigliocco G, Perniss P, & Vinson D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Phil. Trans. R. Soc. B 369*: 20130292. doi:10.1098/rstb.2013.0292

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior research methods, 40*(4), 1079-1087. doi:10.3758/BRM.40.4.1079

Voeltz, F. K. E., & Kilian-Hatz, C. (2001). *Ideophones* (Vol. 44). Amsterdam: John Benjamins Publishing Company.

Westermann, D. H. (1927). Laut, Ton und Sinn in westafrikanischen Sudansprachen. *Festschrift Meinhof*, 315-328.

Westermann, D. H. (1937). Laut und Sinn in einigen westafrikanischen Sprachen. *Archiv für vergleichende Phonetik, 1*, 154-172.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software, 40*(1), 1-29. http://www.jstatsoft.org/v40/i01/.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files*. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.0. https://CRAN.R-project.org/package=dplyr

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?: Iconicity in English sensory words. *Interaction Studies, 18*(3), 443-464. doi:10.1075/is.18.3.07win

Wundt, W. (1900). *Völkerpsychologie. Eine Untersuching der Entwicklungsgesetze von Sprache, Mythus und Sitte* 1. Die Sprache, Leipzig: Wilhelm Engelmann.

Appendix A. Distribution of ideophones in rating experiment

This table shows how the ideophones were distributed over the online questionnaire, which was divided into four lists. The category 'Other' was only present in the ideophones from Japanese (Collabra), which is why it is marked as '-' everywhere else.

Table 1. Distribution of ideophones over the four lists in the survey

List	Category	Language						Total
		Ewe	Japanese (Language)	Japanese (Collabra)	Korean	Semai	Siwu	
A	Sound	2	3	1	3	2	1	12
	Texture	2	2	1	3	2	1	11
	Motion	3	2	2	2	2	1	12
	Shape	2	2	1	2	2	3	12
	Colour/Visual	2	2	0	2	1	2	9
	Other	-	-	4	-	-	-	4
	Total	11	11	9	12	9	8	60
B	Sound	2	3	1	3	2	1	12
	Texture	2	3	1	2	1	2	11
	Motion	3	2	2	2	2	1	12
	Shape	2	2	1	2	2	2	11
	Colour/Visual	2	1	1	2	2	2	10
	Other	-	-	4	-	-	-	4
	Total	11	11	10	11	9	8	60
C	Sound	2	3	1	2	2	2	12
	Texture	2	2	1	2	2	2	11
	Motion	3	2	1	3	2	1	12
	Shape	2	2	1	2	2	2	11
	Colour/Visual	1	1	1	2	2	2	9
	Other	-	-	5	-	-	-	5
	Total	10	10	10	11	10	9	60
D	Sound	2	2	1	2	3	1	11
	Texture	2	3	1	2	2	1	11
	Motion	2	2	1	3	2	2	12
	Shape	2	3	1	2	2	3	13
	Colour/Visual	2	1	1	3	1	1	9
	Other	-	-	4	-	-	-	4
	Total	10	11	9	12	10	8	60

Appendix B. Full survey instructions (Dutch)

The original Dutch instructions for participants in the rating experiment are presented below.

Zometeen krijgt u 60 woorden te horen die niet uit het Nederlands komen. De Nederlandse vertalingen krijgt u erbij te zien. Zorg ervoor dat u zich in een rustige omgeving bevindt waarin u het geluid goed kunt horen. Zorg eventueel voor een koptelefoon of oordopjes.

De woorden beelden telkens iets uit, bijvoorbeeld een geluid, beweging of vorm (of een combinatie van meerdere dingen). Ik zou graag willen weten hoe iconisch u deze woorden vindt. Eerst zal ik uitleggen wat ik precies bedoel met 'iconisch'.

Wat is 'iconisch'?

Van sommige woorden passen de klank en de betekenis erg goed bij elkaar. Van andere woorden juist helemaal niet. Neem als voorbeeld de woorden hobbelen en lopen.

*Het woord hobbelen past qua klank bij de manier van voortbewegen die daarmee wordt uitgebeeld. Hobbelen klinkt dus als wat het betekent. Ook als u geen Nederlands sprak, zou u misschien de betekenis van dit woord kunnen raden. We zeggen ook wel dat dit woord **iconisch** is.*

*De klank van het woord lopen past daarentegen op geen enkele manier bij wat het betekent. De vorm van dit woord is totaal willekeurig. Als u geen Nederlands sprak, zou u de betekenis van dit woord niet kunnen raden. Dit woord noemen we **niet iconisch**.*

De vraag is dus telkens hoe goed u de klank en de betekenis van een woord bij elkaar vindt passen. Als de klank van een woord totaal niet past bij wat het betekent, is dat woord helemaal niet iconisch. Als de klank en de betekenis van het woord juist heel erg bij elkaar passen, is dat woord zeer iconisch.

U beoordeelt de woorden die u hoort op een schaal van 1 tot 5, waarbij 1 = helemaal niet iconisch en 5 = zeer iconisch.

Probeer niet al te lang na te denken over uw antwoord. Onthoud ook: er zijn geen goede of foute antwoorden! Er volgt nu eerst een oefenvraag.

Appendix C. R Markdown used for data restructuring and analysis

Code notebook for a study of the relation between iconicity ratings and experimentally collected guessability scores.

Setup

```{r preliminaries, results="hide"}
# Packages
list.of.packages <-
c("tidyverse","readxl","ggthemes","viridis","gghalves","plyr","emmeans","car")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
lapply(list.of.packages, require, character.only=T)

# useful functions
`%notin%` <- function(x,y) !(x %in% y)
mean.na <- function(x) mean(x, na.rm = T)
sd.na <- function(x) sd(x, na.rm = T)

```

Rating: descriptives

```{r}
# load dataframe with all ratings
read.csv("data\\ideophones_rated_uncorr.csv",header=TRUE,na.strings=c("",
"NA"),sep=";",check.names=FALSE) -> d.uncorr

# to long format
d.uncorr <- gather(d.uncorr, item, rating, 6:245)
d.uncorr <- d.uncorr[,c(1,6,7)]
colnames(d.uncorr) <- c("rater","item","rating")
# remove empty cells
d.uncorr <- subset(d.uncorr, !is.na(d.uncorr[,"rating"]))
# change ratings to numeric values with decimals
as.numeric(sub(",", ".", d.uncorr$rating, fixed = TRUE)) -> d.uncorr$rating

# check for inconsistent raters with person-total correlation based on Motamedi et al. (2019)
ptc <- d.uncorr[,c("item","rater","rating")]
personTotalCorrelationCorrected <- function(ptc)
  {
  ptc %>%
    dplyr::rename(raterFocal = rater,
          ratingFocal = rating) %>%
    full_join(ptc, by='item') %>%
    filter(rater!=raterFocal) %>%
    group_by(item, raterFocal, ratingFocal) %>%
    summarise(ratingsOthers = mean(rating, na.rm=TRUE)) %>%
    base::split(.$raterFocal) %>%
```

```
    map(~cor(.$ratingFocal, .$ratingsOthers,
            use="pairwise.complete.obs")) %>%
    as.data.frame %>%
    gather %>%
    dplyr::rename(rater = key, perTotCor = value)
}
View(personTotalCorrelationCorrected(ptc))

# plot person-total correlations based on Motamedi et al. (2019)
ptc %>%
select(item, rater, rating) %>%
personTotalCorrelationCorrected %>%
ggplot(aes(x=perTotCor)) +
  geom_density() +
  geom_histogram(stat="bin",
               alpha=0.4,
               aes(y=..density..),
               binwidth=0.1) +
  labs(x="Person-total correlation")
ggsave("figures\\rating_PerTotCorr.png",height=5,width=7.5)
```

One participant (N045) showed a negative person-total correlation, which according to Curran (2016) indicates a careless responder rendering invalid data. Inspection of this participant’s data showed they had indeed given the same answer to every question. I therefore decided to remove this participant from the dataset.

Two other participatns (N022 and N034) also had to be removed because of their knowledge of Japanese and Korean, resulting in a total of 75 participants to be included in the analysis.

```{r}
# load corrected data
d <- read_xlsx("data\\ideophones_rated.xlsx")
d.means <- read_xlsx("data\\ideophones_rated_means.xlsx")

# get means and standard deviations across all ideophones and for each language and category
mean(d$rating)
  # 2.948339
sd(d$rating)
  # 1.297106
sumcat <- ddply(d,~category,summarise,mean=mean(rating),sd=sd(rating))
sumlang <- ddply(d,~language,summarise,mean=mean(rating),sd=sd(rating))

```

Because the category "Other" is quite a vague residual category with only 17 items (as opposed to 30-40 items for the other categories), I decided to remove this category for now. I will keep a copy that includes the ratings for "Other".

```{r}
# create copy of the data including the category "Other"
```

```
d.other <- read_xlsx("data\\ideophones_rated.xlsx")
d.m.other <- read_xlsx("data\\ideophones_rated_means.xlsx")

# remove category "Other" from d and d.means
d <- d[!grepl("Other",d$category),]
d.means <- d.means[!grepl("Other",d.means$category),]

# make new summary for language (without the Japanese "other" ideophones)
sumlang2 <- ddply(d,~language,summarise,mean=mean(rating),sd=sd(rating))
```

Here we plot the distribution of the rating data.

```{r}
# convert some columns to factors
cols <- c("category","language","study")
d[cols] <- lapply(d[cols], factor)
d.means[cols] <- lapply(d.means[cols], factor)

# combined boxplot and scatterplot by category
ggplot(d.means, aes(x=reorder(category,-rating),y=rating,color=category)) +
  theme_tufte(base_size=16) +
  geom_half_boxplot(aes(middle=mean(rating)),show.legend=F) +
  geom_half_point(show.legend=F) +
  scale_colour_viridis_d(option="D",alpha=0.8) +
  xlab("category") + scale_x_discrete(labels = c("Sound","Motion","Shape",
  "Texture","Colour/Visual"))
ggsave("figures\\rating_category.png",height=5,width=7.5)

# same, but by language
ggplot(d.means, aes(x=reorder(language,-rating),y=rating,color=language)) +
  theme_tufte(base_size=16) +
  geom_half_boxplot(aes(middle=mean(rating)),show.legend=F) +
  geom_half_point(show.legend=F) +
  scale_colour_viridis_d(option="D",alpha=0.8) +
  xlab("category")
ggsave("figures\\rating_language.png",height=5,width=7.5)

```

Statistics

Now we want to see which categories and languages differ significantly from the others in their iconicity ratings.

First, we check the assumptions.

```{r}
# check normality by visual inspection
ggqqplot(d, "rating", facet.by = "category")
```

```
  # all points fall approximately along the reference line, so normality can be assumed

# check homogeneity of variance with Levene's test
leveneTest(rating~category,data=d)
leveneTest(rating~language,data=d)
  # both are not significant (> .05), so homogeneity of variance can be assumed
```

Now we construct a linear model with category as a predictor of the iconicity ratings.

```{r}
summary(lm.cat <- lm(rating ~ category, d))

# pairwise comparisons with Tukey adjustments in the emmeans package
emm.cat <- emmeans(lm.cat, "category")
pairs(emm.cat)
pwpm(emm.cat)
```

And here we do the same with language as predictor.

```{r}
summary(lm.lang <- lm(rating ~ language, d))

# pairwise comparisons in emmeans package
emm.lang <- emmeans(lm.lang, "language")
pairs(emm.lang)
pwpm(emm.lang)
```

Combined data

Here we load the combined coding and guessability data and then add the rating data.

```{r}
# get coded and guessed data from previous step (TI_02_congruence)
d = read.csv("data\\ideophones_coded_guessed.csv")

coding_categories <- d %>% dplyr::select(matches('F_|M_'),-matches('notes|meaning')) %>%
names()

# add mean ratings
d.ratings = read_excel("data\\ideophones_rated_means.xlsx") %>%
  dplyr::select(-c(item,category,list))
d <- left_join(d,d.ratings,by=c("filename","language","study"))

# convert ratings to z-scores for comparison with guessability (logodd) scores
d$rating_z <- scale(d$rating,center=T,scale=T)

#remove category "Other"
```

```
d <- d[!grepl("Other",d$category),]
```

Visualisations

Let's visualise the relation between the rating and guessing data.

```{r}
ggplot(d,aes(x=rating_z,y=score_z)) +
  geom_point(aes(colour=study),position="jitter") +
  geom_smooth(method=lm,colour="black") +
  theme_tufte(base_size = 16) +
  xlab("rating") + ylab("guessability")
ggsave("figures\\rating_guessing_corr.png",height=5,width=7.5)

ggplot(d,aes(x=rating_z,y=score_z)) +
  geom_point(aes(colour=category,shape=study),position="jitter") +
  geom_smooth(method=loess,colour="black") +
  theme_tufte(base_size = 16) +
  xlab("rating") + ylab("guessability")
```

There appears to be a positive correlation between the rating and guessing data.

Correlation

```{r}
# Pearson correlation
cor.test(d$score_z,d$rating_z,method="pearson")
cor.rank <- cor.test(d$score_z,d$C_cumulative,method="spearman")
```

Pearson's correlation coefficient indicates that there is indeed a strong positive correlation between the guessing and rating data, $r(220) = .61$, $p < .001$.

```{r}
#check where the guessing and rating data deviate
d %>%
  filter(rating_z-score_z > 2 | rating_z-score_z < -2)%>%
  dplyr::select(filename,ideophone,rating,rating_z,language,logodds,score,score_z) %>%
  arrange(score) %>% ungroup()
```