Mappings between linguistic sound and motion

Christine Cuskley[1]

Università La Sapienza, Institute for Complex Systems (ISC-CNR), Rome, Italy
University of Edinburgh, Language Evolution and Computation Unit, Edinburgh, UK

This paper provides an overview of the possible function of non- arbitrary mappings between linguistic form and meaning, and presents new empirical evidence showing that shared cross-modal associations may underlie motion sound-symbolism in particular. In terms of function, several lines of empirical and theoretical evidence suggest that non-arbitrary form-meaning connections could have played a crucial role in lexical emergence during language evolution. Furthermore, the persistence of such non-arbitrariness in some areas of modern language may also be highly functional, as recent data has shown that non-arbitrary forms may help to bootstrap learning in children (Imai, Kita, Nagumo, and Okada, 2008) and adults (Nielsen and Rendall, 2012). Given the functional role of these non-arbitrary mappings between linguistic form and meaning, this paper describes new experimental data demonstrating shared mappings between non-sense words and visual motion using a direct matching task. Participants were given nonsense words that varied in terms of their voicing, reduplication, and vowel quality, and asked to change the movement of a ball to match a given word. Results show that back vowels are mapped onto slower speeds, and consonant reduplication with vowel alternation is mapped onto faster speeds. These results show a shared cross-modal association between linguistic sound and motion, which is likely leveraged in sound-symbolic systems found in natural language.

1. Introduction

Research into sound symbolism has seen a definitive resurgence in recent years, from both linguistic and psychological perspectives. From the linguistic perspective, a growing body of work has emerged describing sound symbolic phenomena in natural language (e.g., Hinton, Nichols, and Ohala, 1994; Voeltz and Killian-Hatz, 2001; Perniss, Thompson, and Vigliocco, 2010). In psychology, a classic cross-modal paradigm wherein participants match non-words with shapes has expanded particularly in the last ten years (see Figure 1). Given a choice of two nonsense words and two shapes, people share preferences to pair, for example, angular shapes, with words like *takete* and *kiki* (see also Ahlner and Zlatev, 2010; Nielsen and Rendall, 2011; Monaghan, Mattock, and Walker, 2012). Such preferences have been shown not only for shape angularity, but also for size (Peña, Mehler, and Nespor, 2011; Thompson and Estes, 2011) and taste (Simner, Cuskley, and Kirby, 2010; Gallace, Boschin, and Spence, 2011; Crisinel, Jones, and Spence, 2012).

Broadly, the preference to pair certain sounds with certain sensory experiences shows a potential for goodness-of-fit between linguistic sound and sensory meaning. This evidence supports a more nuanced revision of the traditional view that the relationship between linguistic form and meaning is entirely arbitrary (e.g., Saussure, 1959; Hockett, 1960). Arbitrariness may remain the dominant relationship between form and meaning in modern language, but a wealth of evidence from both the experimental and natural language perspectives supports a broader picture wherein non-arbitrariness is also a pervasive, if weaker, force in language (Perniss et al., 2010).

---

[1] Email: ccuskley@gmail.com
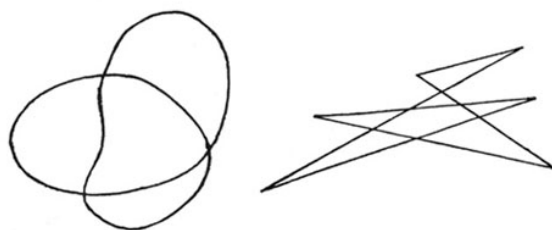
Figure 1. Wolfgang Köhler's (1929, 1947) original *takete/maluma* experiment. Köhler asked participants which of the two shapes above they would call the *takete* and which the *maluma*. Participants reportedly "answered without hesitation" (1947, p. 224), identifying the angular shape (right) as the *takete* and the rounded shape (left) as the *maluma*. This has since been interpreted as evidence of a naming bias to pair certain linguistic sounds with certain visual qualities of shape (Ramachandran & Hubbard 2001), and variations on this paradigm have been used in several more recent studies (Westbury, 2005; Maurer, Pathman, and Mondloch, 2006; Ahlner and Zlatev ,2010; Nielsen and Rendall, 2011; Monaghan, Mattock, and Walker ,2012).

This paper aims to consider non-arbitrariness from two distinct perspectives. First, section 2 provides a functional explanation for the pervasiveness of non-arbitrariness. I argue that the presence of non-arbitrariness is not merely peripheral as some traditional accounts argue (e.g. Chomsky, 1986), but serves a set of specific and essential functions within a language. Furthermore, non-arbitrariness may emerge and persist in language as a result of these functions. The latter half of the paper addresses non-arbitrariness in language with respect to motion specifically. Section 3 examines the natural relationship between sound and motion, and what we know about how natural language uses this relationship.

Lastly, the final section provides new experimental evidence that we share biases to pair certain types of linguistic sound with speed. Previous experiments have focused predominantly on associations between sound and visual angularity or size (e.g., Ramachandran and Hubbard, 2001; Sapir, 1929), with some more recent explorations into taste (Simner et al., 2010; Gallace et al., 2011; Crisinel et al., 2012). Motion is, as yet, a relatively unexplored modality for linguistic cross-modal associations, though recent explorations do show some trends between English and Japanese speakers (see Saji et al., 2012). This experiment provides a particular connection to motion sound symbolism in natural language, tying experimental and natural language evidence together. Lastly, I conclude by suggesting further directions for this line of research.

2. The function of non-arbitrariness

This section seeks to explain why non-arbitrariness persists, and in some cases, proliferates (e.g., in child directed speech, Yoshida, 2012), instead of being eclipsed by arbitrary forms. Before turning to more detail regarding the function of non-arbitrariness, I will first clarify the concepts of arbitrariness and non- arbitrariness.

In a completely arbitrary vocabulary, such as those which dominate modern language, two fundamental things are true: a) forms related in meaning are no more likely to share features than forms which are unrelated in meaning[2], and b) the relationships between word forms and their meanings are rooted in convention among speakers, rather than form and meaning being connected by some goodness-of-fit. Where a language is non-arbitrary, one or

---

[2] Forms, in this case, refers to monomorphemic forms. At the multimorphemic level, the presence of type (a) non-arbitrariness is uncontroversial, and in fact, essential. For example, the meaning of the monomorphemic word *cat* cannot be derived by considering its (phonemic) subparts, but the meaning of a multimorphemic word like *misconstruing* is derived from its (morphemic) subparts

both of a) and b) is no longer true. Put differently, non-arbitrariness can occur on two separate levels: the level of (a) *systematicity* and the level of (b) *iconicity*. Systematicity is the property of shared forms also sharing features of meaning across the lexicon, whereas iconicity is the property of meaning dictating the use of a specific form by virtue of cross-modal goodness-of-fit. In other words, a language in which all words for sharp items have front vowels (as the /i/ in *kiki*) is systematic by virtue of the fact that vowel backness relates to sharpness across the system, and iconic by virtue of the fact that front vowels are particularly well-suited to sharp objects (Ahlner and Zlatev, 2010; Monaghan et al., 2012).

It is important to note that though iconicity exhibits goodness-of-fit where systematicity does not, the percieved goodness-of-fit can be at least partially dependent on convention. Ahlner and Zlatev (2010) describe the distinction between primary iconicity and secondary iconicity (after Sonesson 1997). In primary iconic signs, the goodness-of-fit relationship between the sign and signifier is evident at the outset. However, in signs with secondary iconicity, the goodness-of-fit relationship is evident only in a post-hoc manner (e.g., a circle with two dots could iconically depict pig or power outlet depending on context). This highlights the fact that convention and arbitrariness are often conflated. The interpretation of a circle with two dots as a sign for pig is non-arbitrary in the iconic sense, but fully grounding the sign still requires convention. Below, I will argue that the combination of systematicity and iconicity at the phonemic level served to bootstrap the evolution of an arbitrary vocabulary, and that iconicity is still used to generate new forms, maintaining a subset of language which is highly learnable.

The first function, for bootstrapping an emerging vocabulary, stems from a much older set of theories about non-arbitrary language origins. So-called 'bow-wow' theories of language origins suggested that language was built on direct imitations of sounds (see Aitchison, 2000 for a review). These early 'bow-wow' theories lacked specific detail regarding mechanisms which would have led the transition into more arbitrary systems given a non-arbitrary origin. As a result, the idea of a non-arbitrary origin to language was not seriously considered until much more recently.

However, Gestalt psychologists, unconstrained by the notion of arbitrariness in theoretical linguistics, began to explore the notion of non-arbitrary naming in more detail. Wolfgang Köhler began experiments in motivated naming in the early 20th century (described in Figure 1). Similar experiments have since been performed with children (Irwin and Newland 1940; Maurer et al., 2006), as well as cross-culturally (Davis, 1961) and with varying visual and linguistic stimuli (Ramachandran and Hubbard, 2001; Nielsen and Rendall, 2012; Nielsen and Rendall, 2011). In the vast majority of cases[3], people agree significantly in their non-word/object pairings, demonstrating a shared 'naming bias' to map certain names to certain visual forms. Several early authors hinted that this naming bias might have played an important role in language evolution (Köhler,1929; Köhler, 1947; Werner, 1957; Werner and Wapner, 1952). However, Ramachandran and Hubbard (2001) were the first to outline in detail a "synaesthetic bootstrapping theory of language origins" (2001, p. 15), suggesting shared associations linked linguistic sound to other sensory experiences.

An iconic theory for language origins fills a particular gap in the puzzle of language origins, providing a way to ground a fledgling vocabulary in our shared sensory experiences. An arbitrary vocabulary has no obvious starting point or connection to an existing conceptual system - in other words, there is a question as to how the symbol system is grounded (Harnad, 1990). With an iconic origin to language, linguistic symbols (both in the speech and gestural

---

[3] One notable exception is Rogers and Ross (1975), where the effect was not found among the Songe in Papua New Guinea.

channels, e.g., Brown, 2012; McNeill, 2005) - could have been grounded in our perceptual system. In the case of speech signals, this grounding was primarily through shared cross-modal associations: associations between different sensory modalities (e.g., sound and vision) that are shared across a population of speakers. These associations scaffolded the emergence of a small, predominantly non-arbitrary communication system, or protolanguage (e.g., Bickerton, 1990; Wray, 1998), from which modern languages themselves emerged. This protolanguage was initially small, changing and expanding in response to various pressures over time, and eventually maturing into the predominantly arbitrary modern language systems we observe today (for further detail on mechanisms active in this theoretical transition, see Cuskley and Kirby, 2013; Cuskley, 2013; Deutscher, 2005).

This leads to the modern functions of arbitrariness and non-arbitrariness. Pressures on a language system to expand - for example, a small protolanguage shifting to a full-fledged modern language - would have resulted in a shift from iconicity to arbitrariness. However, there remain advantages to maintaining some non-arbitrariness within a system. The contrast between advantages inherent in non-arbitrariness with the advantages inherent in an arbitrary system shed light on why we continue to see both types of form-meaning relationships in language today. Non-arbitrariness provides increased learnability and decreased processing demands, while arbitrariness increases the expressive power of a language.

Theoretically, learnability of a system greatly increases when it is non-arbitrary. If the connection between form and meaning is intuitive, storage and fidelity of a form-meaning pair will be quicker at the outset and more robust against loss over time (Gasser 2004). Learnability is not trivial; in fact, it is considered a core property of human language (Hockett 1960) - a language must be learnable by its users in order to be useful. For example, imagine a 'language' of road signs, each with a different meaning. A learning driver who must memorise pictorial road signs will have an easier time of doing so if the signs are non-arbitrary, and even more so if they are specifically iconic. If the sign for 'no parking' is a car with a line through it, the connection between the sign and its meaning is obvious, making it easier to remember. Moreover, such obvious connections throughout a system decrease the need for rote memorisation altogether - interpretation of meaning is so clear and automatic that permanent storage of the relation might not be necessary at all. Even in a system of road signs that was not necessarily iconic but remained systematic - for example, where any sign indicating a need to decrease speed had a grey border - learnability is greatly increased. Rather than remembering disparate signs for a set of related actions (e.g., stop, merge, yield and slow), the learner can associate similar actions (those requiring a decrease in speed) with a single sign feature (a grey border). Of course, this doesn't eliminate the need to have other features distinguishing the separate but related actions, but the learning burden is decreased by the relatedness of signs. Using this example, we can even turn the question on its head: why would signs for related actions not share related visual features?

This leads onto the second advantage: a non-arbitrary system requires less effort during retrieval (either of a form during production or a meaning during comprehension) by providing a constrained search. Iconicity is, by definition, a natural strength in the bond between form and meaning, making the retrieval of a meaning given only the form - or the retrieval of a form given only the desired meaning - more automatic. Conceptualise the task of retrieval as one of search: where we see a form, our task is to search in our memory for its meaning; where we see a meaning and want to express it, we are searching in memory for the correct form. If all forms denoting angular objects contain the front vowel /i/, the task of searching for the meaning of a form such as *kiki* is immediately constrained to meanings with the feature [+angular]. In an arbitrary system, however - where the form *kiki* might denote any type

of object - the task of searching for the relevant meaning remains largely unconstrained, and the search must be exhaustive.

Given these advantages, why would language ever become an arbitrary system? Arbitrariness presents a different set of advantages. Most notably, it allows for the expansion of a system while minimising potential ambiguity. Gasser (2004) set out to examine the advantages of arbitrariness and non- arbitrariness relative to vocabulary size in a series of simulations using a simple feed forward network[4]. The network learns systematic languages with much lower error, in accordance with the notion that non-arbitrary languages have superior learnability. However, this advantage only persists while the systematic language is small. When the size of the language was expanded, the trend reversed: arbitrary languages were learned with lower error than non-arbitrary languages.

Why are large languages more learnable if they are arbitrary? Gasser suggests that the manner in which the form-meaning space is occupied in non-arbitrary versus arbitrary systems is essential to understanding the large arbitrary advantage. If there is a non-arbitrary relationship between form and meaning, a given meaning cannot have just any form, and so the space is occupied in a regular and organised fashion, as in Figure 2a. To occupy the space in such an organised fashion is sensible if we only have five meanings. However, if we wish to expand the size of the language such that there are more items and maintain such organisation, similar forms and similar meanings are now effectively on top of one another (Figure 2b) - so close that confusion is likely. This increases the potential for ambiguity: a given form may be taken for the wrong meaning. In contrast, if we allow these ten meanings to occupy the space in a more distributed - and necessarily arbitrary - manner, then forms have the distance necessary to maintain accurate discriminability (Figure 2c).
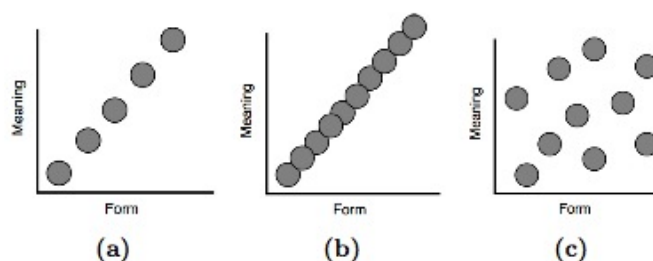
Figure 2. Occupation of the form meaning space in a small non-arbitrary lan- guage (a), a large non-arbitrary language (b), and a large arbitrary language (c). Adapted from Gasser (2004).

Monaghan, Christiansen, and Fitneva (2011) have demonstrated experimentally that non-arbitrariness can create an ambiguity problem in a vocabulary system where the meaning space outgrows the form space. In their experiment, participants were tasked with learning either an arbitrary or a systematic language, mapping non-words to a series of 8 actions and 8 objects. In the arbitrary language, there was no systematic relationship between form and meaning; in the non-arbitrary language a particular type of sound (e.g., fricatives), was al

---

[4] A feed forward network is a simple simulated neural network (a common connectionist method, e.g., see Hopfield, 1982) wherein information moves forward only, with no cycles or loops between input and output nodes.

ways present with a particular meaning type, for example, actions[5]. Following training, participants' performance was measured in terms of two metrics: individuation (whether or not they selected the correct specific meaning for a given form) and categorisation (whether they selected correct category - object or action - for a given form). Monaghan and colleagues found that participants learning the systematic language were significantly more accurate at categorisation, whereas at final test there was no difference in individuation accuracy between the systematic and arbitrary languages.

Monaghan and colleagues (2011) interpret these results specifically as evidence that non-arbitrariness is useful for indicating broad category membership, but not necessarily for individuating particular referents. However, in light of Gasser's (2004) language size perspective, ambiguity problems found in the individuation task could be interpreted in terms of size of the meaning and form spaces rather than task type (categorisation vs. individuation). Indeed, performance in both indviduation and categorisation was measured based on performance in the same form-meaning pairing task. In the case of individuation, participants were measured on discrimination between 16 meanings, but measuring the same performance only in terms of categorisation shrinks the meaning space from 16 to 2. The form space in these experiments was small relative to natural language form spaces, consisting of only four consonants and four vowels whose combination was constrained to particular CVC words, providing 16 possible forms. Since the meaning space contained 16 meanings, measures of individuation would be measuring performance where the size of the meaning space is equal in size to the form space, resulting in ambiguity equivalent to many meanings trying to occupy a small form space systematically (as in Figure 2b). However, measuring categorisation rather than individuation leads to only two relevant meanings (object and action), such that the form space dwarfs the meaning space. Thus, arbitrariness becomes a better strategy for a language which must cover a large meaning space with a finite form space, but non-arbitrariness can still be effective in particular pockets of meaning within a largely arbitrary system.

Further experimental evidence has bolstered particularly the learnability advantage of non-arbitrariness. Learnability has been used in several studies to implicitly show a cross-modal association, meaning that not only systematicity (as in the study discussed above), but also iconicity, play an important role in learnability. For example, participants learn to pair words like *takete* with angular shapes more readily and accurately than they do with rounded shapes (Nielsen and Rendall, 2012; Monaghan et al., 2012), and process such pairings more quickly (Westbury, 2005; Kovic, Plunkett, and Westerman 2010). Crucially, evidence for increased learnability of iconic forms extends to iconicity in natural language.

A simple way to test for non-arbitrariness in natural language is to ask whether the forms of foreign words contain any information regarding their meaning. In other words, if there is an entirely arbitrary relationship between form and meaning, monolingual English speakers who are given the Finnish forms kylma and kuuma, and asked to guess which means cold and which means hot, will guess the meanings of these words at levels no different to chance. The fact that people are able to guess the meanings of foreign words with no knowledge of the foreign language in question provides evidence of iconicity. This sort of experiment, where participants who do not speak a language are asked to guess the meanings of words, is particularly suited to *ideophones*. Ideophones form a class of words with acknowledged iconicity, reported by speakers as being evocative of sensory experience (Dingemanse, 2011). Iwasaki, Vinson, and Vigliocco (2008) asked both Japanese and English speakers for their intuitions regarding the meanings of Japanese ideophones for motion (e.g., *bura-bura*:

---

[5] Note that this non-arbitrary relationship was only systematic in nature (in that it was implemented uniformly throughout the vocabulary), but that there is no evidence of *iconic* goodness-of-fit between fricatives and actions.

strolling). Despite no knowledge of Japanese, the English speakers rated the meanings of Japanese ideophones similarly to the Japanese speakers. For example, both English and Japanese speakers chose the ideophone *toko-toko* to refer to the gait of a small person walking quickly and energetically in an informal manner. Similar judgements between Japanese and English speakers were also found for Japanese pain ideophones (Iwasaki, Vinson, and Vigliocco 2007a).

As the biases underlying cross-modal matching experiments can also be demonstrated using learnability, the iconicity of ideophones for naïve participants can likewise be demonstrated with a learnability-based approach. If ideophones are forms motivated by their meanings, the natural connection between form and meaning should make them more easy to learn than forms with no natural relationship to their meaning. Imai et al. (2008) found that three year old children were able to more effectively learn the meanings of novel verbs when they were designed to be ideophonic in nature (e.g., voicing was contrasted to indicate lightness or heaviness, as it is used among many Japanese ideophones). To eliminate the possibility that this result might simply demonstrate Japanese children's familiarity with Japanese ideophones, Kantartzis, Imai, and Kita (2011) showed that English speaking children with no exposure to Japanese demonstrated the same ideophonic learning advantage. Yoshida (2012) has also shown that sound-symbolic forms are used more in child directed speech than normal speech. Even among English speaking adults, who have no formal ideophone system to leverage in child-directed speech, there is significant use of idiosyncratic sound-symbolic forms when speaking to chidren. Nagumo, Imai, Kita, Haryu, and Kajikawa (2006) have shown that Japanese speakers are almost twice as likely to use ideophones when speaking to children than to adults, suggesting a special role for ideophones in language acquisition. Together, the results of Imai and colleagues (2006) and Nagumo and colleagues (2006) suggest that the learnability conferred by iconicity is especially important for verb acquisition in child learners, when language learnability is arguably most important.

Experiments with Japanese ideophones have also shown that their function may be of particular importance for verb acquisition in children. There is ample evidence that verbs are generally more difficult to acquire, and thus are later acquired, than nouns (Childers and Tomasello, 2006; Werker, Cohen, Lloyd, Casasola, and Stager, 1998). Imai and colleagues (2008) suggest this may be due to the more "ephemeral" nature of verbs (p. 56): specifically, they are more difficult to individuate in a complex scene (Gentner 1982). In other words, the learner must separate the action from the actor and learn to identify the action in disparate contexts. An experiment confirmed this difficulty children often have in learning verbs; Japanese and Chinese children failed to generalise the mean- ing of novel (non-sound symbolic) verbs when the object acted upon was altered (Imai et al., 2008). In other words, they could learn to associate a video of a person moving a ball in a specific way with a novel word, but would not generalise the label to a video of a person performing the same action with a cube.

A follow-up experiment has demonstrated that the difficulty in learning form-meaning pairings for actions may be mitigated by iconicity. Considering the difficulties inherent in verb learning, this may mean that connections between linguistic sound and action are highly functional in this particular domain. Imai, Kita and colleagues (2008) showed that generalisation (e.g., of an action to various objects) was robust if the novel words were designed to be sound symbolic. In other words, if the non-words used the features present in Japanese ideophone systems, children were considerably more adept at learning the form-meaning pair accurately and applying it across diverse contexts.

This section has shown that non-arbitrariness in language serves a specific set of functions, particularly related to grounding emerging vocabulary, and providing advantages in learnability and processing over arbitrary forms when used for a confined set of the meaning

space. The next section will focus in on a particular type of sound symbolic language: motion sound symbolism. The work outlined above shows cross-modal biases in naming using non-words and shapes, and connected these experiments to functional aspects of iconicity in language. However, only a handful of studies (e.g., Imai et al., 2008, Kantartzis, Imai, and Kita, 2011) have used iconicity in natural language to motivate the use of particular linguistic sounds and make specific predictions in Köhler-style linguistic cross-modality studies.

Below, I present experiments showing that people make regular mappings between features of linguistic sound and motion. Specifically, participants hear non-words which vary systematically in terms of their linguistic features (voicing, vowel quality, and reduplication) and change the motion of an animated ball to 'match' the word they hear. Results show that some features of linguistic sound are matched to motion systematically. First, I provide a brief overview of cross-modality between sound and motion more generally, and address the motivation for the experiment from a natural language perspective.

3. Connecting sound and motion

Like many sensory experiences, sound and motion perception are fundamentally connected via low-level multi-modal representations: things that move tend to make sounds when they do so, and so we learn to have expectations about how sound and motion should go together. Indeed, some have argued that motion perception is fundamentally multi-modal, and research in motion perception has shown that auditory information in particular is vital to the perception of motion (e.g., Poirier et al., 2005; Poirier et al., 2006; Shams, Kamitani, and Shimojo, 2002; Soto-Faraco, Lyons, Gazzaniga, Spence, and Kingstone, 2002; Colonius and Arndt, 2001).

The integration of visual and auditory information is of particular importance in speech perception: accurate processing of speech integrates information from both the acoustic signal and from articulatory movements during speech (e.g. in the well-known McGurk Effect; Macdonald and McGurk 1978, but see also Green, 1997). Furthermore, there is evidence that visual articulatory information improves speech perception where the auditory signal is degraded or noisy (Massaro, 2004), and such visual information is integral for second language learners, who rely on articulatory cues to improve perception of foreign phonemes (Navarra and Soto-Faraco, 2007).

Automatic audiovisual integration, in both the perception of motion generally as well as during speech perception, demonstrates that there is a strong natural connection between motion and sound, and that it plays a role in language. However, the sound-motion associations described above arise from natural co-occurrence; in other words, they are fundamentally indexical. For example, we associate particular oral gestures with certain linguistic sounds because they reliably co-occur. However, this type of indexical natural co- occurrence cannot account for less obvious mappings between, for example, bitter tastes and low vowels (Simner et al., 2010), or round objects and sonorant consonants (Ahlner and Zlatev, 2010; Nielsen and Rendall, 2011). The next section explores evidence for cross-modal associations between linguistic sound and motion which are not obviously indexical.

3.1. Motion Sound Symbolism

Experiments using abstract shapes and non-words, such as Köhler's classic takete/maluma experiment, show shared cross-modal associations with no apparent indexical underpinnings. There is little experimental evidence for these sorts of associations between motion and linguistic sound - but the prevalence of sound symbolic language which depicts motion speaks to the potential for shared associations between motion and linguistic sound. In other words,

we can posit that the ability of sound symbolic forms to evoke motion events is derived from some cross-modal goodness-of-fit be- tween linguistic sound and motion. Motion sound symbolism has been well documented across a diverse group of languages, including Japanese, Thai, Indonesian, and Korean (Wienold, 1995), Basque (Ibarrexte-Antunaño, 2006), Quechua (Nuckolls ,2001), and Balto-Finnic languages (Mikone, 2001).

Documentation of ideophones cross-linguistically is still in its relative infancy, and therefore, specific data is fairly sparse. Thus far, most efforts have considered ideophone systems in distant languages in parallel rather than in concert (though see Dingemanse, 2013 for a recent synthesis). There is a broad literature documenting the existence of ideophones in languages as disparate as Quecha and Japanese (e.g., see Voeltz and Killian-Hatz, 2001). Such examinations focus primarily on arguing for the status of ideophones in language writ large, as well as their particular class and function within specific languages. There is also a body of literature focused on the sensory/perceptual nature of ideophones and their overall aesthetics (see, e.g., Majid and Levinson, 2011; de Sousa, 2011; Tufvesson, 2011). However, there exist few examinations of form and meaning correlation in ideophones as they occur cross-linguistically[6].

Detailed knowledge of the cross-linguistic inventory of motion ideophones is not yet mature enough to have produced a large cross-linguistic corpus such that we might, for example, examine if certain phonemes or sound patterns are likely to occur across motion ideophones (rather than being unique to a particular language's ideophone system). Until such a well-rounded cross-linguistic corpus of motion ideophones exists, we are left instead with a few instances of similarities among ideophones across lan- guages. Nonetheless, Dingemanse (2013) points out that these similarities do display some patterns, including features of reduplication, voicing, and vowel quality which may be particularly important in depicting motion. This section will discuss similarities among some salient features of Basque and Japanese motion ideophones, as these are currently well-documented and motivate the experimental examination to follow.

Ideophones form an integral part of the Basque vocabulary. Trask (1997) identifies several ways in which Basque ideophones are used, including, but not limited to, "interjections, onomatopoeic words, and adjectives or nouns denoting physical or moral defect" (p. 257). Most relevant to the current discussion, Trask identifies "adverbs which denote particular sorts of actions or ways of doing things, e.g., *mara-mar*a: smoothly, steadily, continuously." (as described in Antunaño 2006, p. 502). Crucially, Antunaño points out that Basque ideophones accomplish the depiction of motion in particular ways, one of which is the use of total or partial reduplication. In a corpus of Basque motion ideophones, Antunaño found that 44% made use of reduplication, for example: *tapa-tapa* (tip-toeing) and *zapa-zapa* (walk without stopping; p. 5). Indeed, (Hinton, Nichols, and Ohala) suggest that reduplication is widely used in ideophones cross-linguistically, particularly to depict visual motion events. Basque also makes use of partial reduplication to depict motion, wherein the reduplicated element is confined to consonants, with the vowel alternating between a high front and low or back vowel. For example, *bilin-balan* depicts a tumbling action, while *kili-kolo* depicts wobbling (Antunaño 2006).

---

[6] There is a rather mature typological literature examining how motion is expressed cross-linguistically (e.g., Talmy 2000; Slobin 2004); this body of work makes typological distinctions between languages based on the relationship between the figure, ground, manner, and path in a motion event. However, these distinctions do not capture whether the expression of motion in a language is accomplished using ideophones (Beavers, Levin, and Wei Tham 2010). Since ideophones are of specific interest for their potential cross-modal goodness-of-fit, isolating them is of great importance.

The process of reduplication is, of course, not strictly confined to ideophones, and is often used to express plurals, superlatives, and iteration. For example, in Pagnisan, an Austronesian language spoken in the Phillipines, *amigo* means "friend" while *amimigo* means "friends" (Rubino 2005) . In Ilocano (also an Austronesian language spoken in the Phillipines), reduplication forms both superlatives (*dakkel* 'big' → *dakdakkel* 'bigger') and iteration (*ag-tilmon* 'swallow' → *ag-tilmotilmon* 'swallow repeatedly'; Rubino, 2005). Thus, reduplication is often used to express dimensions which lend themselves to being expressed in terms of more or less (i.e., prothetic dimensions; Walker, Bremner, Mason, Spring, Mattock, Slater, and Johnson 2010)[7]. Japanese ideophones also make use of reduplication. For example, *gorogoro* depicts a heavy object rolling (while *korokoro* depicts a light object rolling; Imai, Kita, Nagumo, and Okada 2008; Hamano 1998). However, reduplication appears to be widely used throughout Japanese ideophones to indicate repeated action, and is not confined to depicting motion. For example, the bound root *pota* depicts the sound of "thin liquid hitting a solid surface," while the geminate form (*potapota*) means "thin liquid hitting a solid surface repeatedly." (Imai, Kita, Nagumo, and Okada 2008, p. 55). Another important feature of Japanese ideophones is the use of voicing to depict weight, largeness, or coarseness, as in the contrast between *gorogoro* and *korokoro*.

The evidence from Basque and Japanese provides a starting point for investigating associations between linguistic sound and motion. First, this evidence highlights the potential for linguistic sound to depict motion events, and thus provides a reasonable expectation that there are, in fact, shared associations between features of linguistic sound and motion. Second, the specific sound patterns found in these ideophones can provide the basis for several hypotheses regarding the nature of these associations. For example, it is likely that people will associate features of motion with reduplication (as found in both Basque and Japanese). The use of reduplication to express concepts such as plurality, superlatives, and iteration also lends it to possible prothetic matching, wherein increased motion is related to reduplication. On these grounds, we can predict that reduplicated items may be matched with increased speed of motion.

Additionally, voicing and vowel quality may also map onto speed, as these contrasts have been found to play a role in other linguistic cross-modal map- pings involving shape (Cuskley, 2013; Nielsen and Rendall, 2011), taste (Simner et al., 2010), and perhaps most relevantly, size. Thompson and Estes (2011) show that voiceless consonants and front vowels are rated as being smaller, while voiced consonants and back vowels are considered large. Size sound symbolism has been tied to the frequency code hypothesis (Ohala, 1994), which holds that certain phonemes exhibit features of low frequency (such as voiced segments and back vowels), and low frequency is fundamentally physically connected to larger size (i.e., larger animals make lower pitched sounds, Fitch, 1997). Speed of motion also connects fundamentally to size: larger entities tend to move slower than small entities thus, front vowels may be considered faster and back vowels slower[8] (see also Berlin, 2006). The next section describes an experiment which examines how the features of voicing, vowel backness, and reduplication are associated with speed in a direct perceptual matching task.

---

[7] Prothetic dimensions stand in contrast with *metathetic* dimensions, which cannot be straightforwardly described in terms of more or less. For example, different taste qualities represent a metathetic dimension.

[8] The author thanks an anonymous reviewer for this suggestion.

4. Experiment: Mappings between linguistic sound and speed

The following experiment examines associations between the motion of a bouncing ball and non-word items designed to vary in terms of reduplication, voicing, and vowel quality. Using these features, a corpus of 105 non-words (described in detail below), was created. Participants heard a subset of these words and altered the speed of a simple animated bouncing ball to 'match' the words. Unlike many previous sound-symbolism experiments which use a forced choice paradigm (e.g., the classic takete/maluma style task used to examine angularity and adapted to motion in more recent studies, see Imai et al., 2008), this method allows participants to provide gradable associations, and has also been used in several other recent studies (Simner et al., 2010; Gallace et al., 2011). This method has the advantage of reavealing only actual associations, rather than default associations. For example, in a forced choice paradigm, a strong association between vowel frontness and angularity could also (erroneously) be interpreted as an association between vowel backness and shape roundness.

4.1. Participants

Participants for this experiment were recruited online using Amazon's Mechanical Turk (MTurk), reviewed in detail in Buhrmester, Kwang, and Gosling (2011):

> MTurk...provides people and tools that enable task creation, labor recruitment, compensation, and data collection. The site boasts a large, diverse workforce consisting of over 100,000 users from over 100 countries who complete tens of thousands of tasks daily. Individuals register as "requesters" (task creators) or "workers" (paid task completers). Requesters can create and post vir- tually any task that can be done at a computer...Workers can browse available tasks and are paid upon successful completion of each task. Requesters can refuse payment for subpar work. (Burhmester et al., 2011: 3)

Burhmester et al. (2011) also detail that MTurk's workforce is more diverse than traditional samples for psychological experiments (i.e., undergraduate psychology students; see Sears, 1986; Henrich, Heine, and Norenzayan, 2010), and the quality of the data obtained is comparable to in-lab data collection (see also Gosling, Vazire, Srivastava, and John, 2004).

Participants in the current study were compensated 50 US cents for a minimum of five minutes' participation[9]. Data was originally collected from 60 participants, but some participants were automatically excluded for one of two reasons: inattention to the task or insufficient evidence of task completion (explained in further detail in the procedure section). In all, 11 participants were not considered, leaving a subject pool of 49 participants.

---

[9]The rationale behind a time limit rather than a limit on the number of items was to prevent quick responses which were not carefully considered, a particular danger considering the unsupervised online nature of completing the task. In other words, if participants were encouraged to complete, for example, 10 items regardless of time, they may have rushed through these in a very short time without attending to the main goal of making an intuitive match between the sounds of the non-words and the speed of the ball. Thus, participants were instructed that they could take as long as they wanted with each item, but had to 'play' the game in total for at least five minutes, and could continue for longer if they wished. The time was measured through the applet by taking a timestamp from the participant's machine for each response. If the time between the first response and the last response was less than four minutes (allowing a possible minute for the first response), the participant was not compensated. The participant who responded to the fewest items responded to only six in just under six minutes, while the person who responded to the most items responded to 88 items in twelve minutes. This meant that each participant responded to a varying number of items, and took a varying amount of time to do so.

4.2. Materials

A corpus of distinct two-syllable CVCV words were designed to vary systematically in terms of voicing, vowel backness, and reduplication. The set of words was designed by taking the seven voiced/voiceless consonant pairs in Figure 3 combined with either the front vowel /i/ or the back vowel /u/.

Figure 3. Consonants used in creating CVCV non-words for motion experiments. In each pair, the voiceless consonant is on the left and the voiced on the right.

Within a word, only the voicing of the consonant sounds could vary, but not the place or manner of articulation. In other words, kigu was a valid word while kivu was not, since /k/ and /v/ vary in terms of both their place (k: velar and v: labial) and manner (k: plosive and v: fricative) of articulation. With 16 possible words for each of the seven voiced/voiceless pairs, 112 CVCV words resulted. Figure 4 shows how a set of sixteen words was made for each consonant pair using the example of the velar plosives /k/ and /g/. The resulting words could be classified in several ways (see Figure 4); in terms of voicing (voiced, voiceless, mixed), vowel backness (all front, all back, mixed), and reduplication (reduplicated, only consonant reduplicated, only vowel reduplicated, unredpulicated).

k, g + i, u ➡

kiki	gigi	kigi	giki
kuku	gugu	kugu	guku
kiku	gigu	kigu	giku
kuki	gugi	kugi	guki

	Voiced	Voiceless	Mixed
Voicing	gigi, gugu, gigu, gugi	kiki, kuku, kiku, kuki	kigi, kugu, kigu, kugi, giki, guku, giku, guki

	Front	Back	Mixed
Vowel Quality	kiki, kigi, gigi, giki	kuku, kugu, gugu, guku	kiku, kuki, gigu, gugi, kigu, kugi, giku, guki

|	Total	Consonant Only	Vowel Only	None
Reduplication	kiki, kuku, gigi, gugu	kiku, kuki, gigu, gugi	kigi, kugu, giki, guku	kigu, kugi, giku, guki

Figure 4. The process used to create 16 words for each of the seven voiced/voiceless consonant pairs, using the exemplar pair of /k/ and /g/ and the two vowels /i/ and /u/. The word /kuki/ was not used due to its proximity to the word kooky. The resulting words were classified according to voicing, vowel quality, and reduplication.

Seven of these words (*duti, kuki, pupi, pupu, pipi, tSutSu* and *tutu*) were removed from the set due to their pronunciation being identical or near identical to familiar lexical items (*duty, kooky, poopy, poopoo, peepee, choochoo* and *tutu*). Recordings of each word were synthesised using the Festival Speech synthesis software and Praat[10].

The animated experimental interface was programmed using Processing and then exported to a Java applet embedded in a webpage. The animation consisted of a black ball on a white background which bounced upon hitting the 'ground' (i.e., the bottom of the frame for the applet). The user interface was located on the left side of the applet (see Figure 5), and

---

[10] To ensure roughly equal durations of all items (all between 450-550ms), as well as even stress, relevant phonetic segments were taken from Festival and re-concatenated using a Praat script (e.g., because typing *kiki* directly into Festival does not result in an evenly stressed reduplicated CVCV word). Appropriate transitions between segments within words were ensured by having separate recorded segments for the different environments in which a given segment occurred (e.g., the /i/ was slightly different in its word final and inter-consonantal forms, and also differed depending on the character of preceding/following consonant, i.e., voicing, place, and manner).

had controls which allowed the participant to change the speed of the bouncing ball using a dial, and buttons to play the non-word, submit their speed choice and move onto the next non-word, and exit the game altogether.
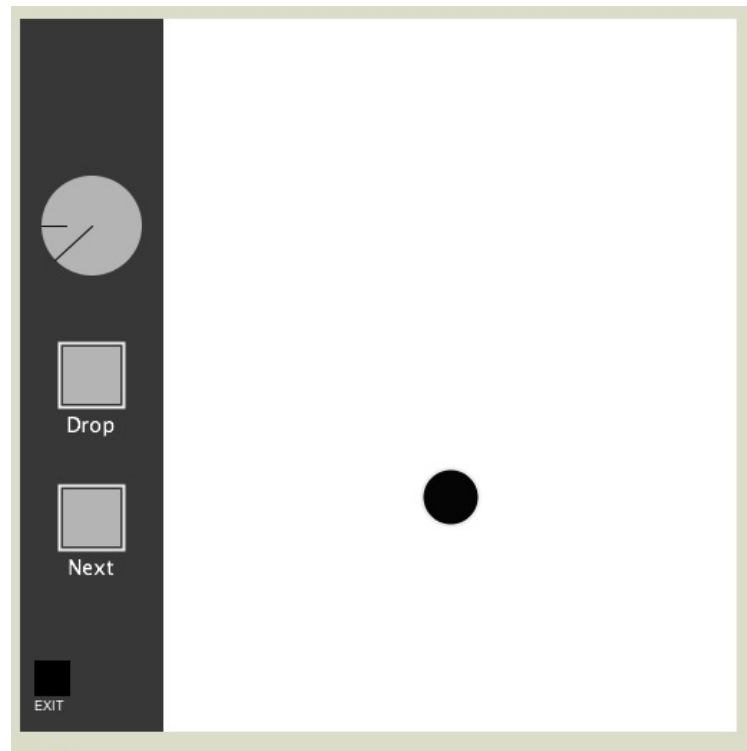
Figure 5. A screenshot of the interface for the experiment, with controls on a panel to the left. The dial at the top controlled the speed of the ball (as determined by the Mass parameter). The 'Drop' button served two functions: to re-release the ball into the space so changes in speed could take effect, and to play audio of the non-word for that trial. Dial changes would only take effect once the 'Drop' button was pressed to ensure participants actually observed changes in the speed dial before submitting them. The word was played at the beginning of the trial as well as each time the ball was dropped. The 'Next' button served to record a participant's choice of match and advance them to the next trial by simultaneously resetting the ball speed to the default midpoint of the speed dial (250), re-releasing it into the space, and playing the non-word for the next trial. The 'Exit' button on the lower left allowed participants to 'end play' and display a HIT code to obtain compensation.

The speed of the bouncing ball was changed by moving a dial. Within the program, the value of this dial was tied to a mass parameter which varied between 0-500. This parameter does not translate directly to speed, but does affect perceived speed of the ball[11]. In this parameter, low values indicated fast speed and high values indicate slow speed.

---

[11] The ball was a particle which was re-released into the space each time the participant dropped the ball. The mass parameter changed the perceived speed of the ball such that when it had high mass it appeared to move slowly, and with a low mass it appeared to move fast. Although mass normally does not affect speed, within the Traer physics package used to code the animation, heavier particles have more inertia and accelerate slower after their release into the space. This resulted in overall slower motion for the ball on a heavier setting, and faster motion on a lighter setting. The updated release of Traer Physics library (3.0) now provides a velocity parameter for particles, but this was unavailable at the time of testing (fall 2010). As such, the Mass setting was used as a proxy.

When a participant chose to view the task associated with the experiment, they were given information and brief directions regarding the experiment, as well as a link to click through to the applet itself. At this stage, participants were also asked to list any languages they spoke aside from English - thus, monolingual and multilingual status was self-reported. Rather than confining the participant pool to English speaking monolinguals, multilingualism was a factor in analysis. Of the 49 participants, 29 were native English speakers, 10 of whom also reported speaking another language, making for 18 monolingual English speakers and 31 multilinguals. Once at the applet, participants began an instructional process to familiarise themselves with the interface. The instructions included an audio captcha to ensure participants were listening to the relevant audio[12]. After the instructional period, participants began with the test words.

Each trial began with audio of a non-word playing, and the ball dropping at the default speed of 250 (the midpoint of the dial). The ball speed did not change automatically with movement of the dial; rather, for the change indial setting to take effect, the participant had to press the 'Drop' button. Whenever a participant reset the ball speed by pressing the Drop button, the non-word of the current trial also played. When the participant submitted their speed choice for the current non-word, the ball speed was reset to the default value of 250, the ball was re-released into the space, and a new non- word played. The non-words were played in a completely random order for each participant.

4.3 Analysis and Results

Analysis of this data presented a slightly unusual challenge since items were grouped in different ways according to different phonetic features of interest (vowel quality, voicing, reduplication), and some groupings contained more items than others (e.g., there were more mixed voicing items than voiced items). For this reason, several linear mixed effects models were run using the R statistical package; specifically, the lmer library. The overall goal of the linear mixed effects models was to assess if the relevant phonetic features could significantly explain variations in speed ratings. For each of the phonetic features of interest, a model was run with the relevant phonetic feature as a fixed effect and individual participants as a random effect. Using the participant as a random effect was also an advantage because of our unusual sampling method, where not all participants responded to the same number of items.

In total, three mixed effects models were run. As this is analogous to running multiple comparisons, the α for these analyses was set to $\alpha = 0.017$ (the standard α of 0.05 divided by 3, after Curtin & Schulz, 1998). The first model aimed to examine if bilingual participants were responding to stimuli differently than monolingual participants. This left two mixed effects models designed to examine sound-motion associations more specifically: one which considered voicing and vowel quality, and one which considered reduplication. Voicing and vowel quality were eligible to be considered together because they are independent; in other words, where there is a voiced consonant, this fact gives no indication of what vowels may be in the word. On the other hand, reduplication is not independent of either voicing or vowel

---

[12] While the audio captcha could ensure that participants had working audio with a minimum quality to comprehend the audio captcha, this was the only quality control on participants' audio. This differs markedly from more traditional psychophysical or even cross-modal experiments where delivery of audio stimuli is usually highly controlled across participants. However, removing this high level of control is unlikely to have resulted in a systematic difference in the delivery of audio which led to spurious significant results. In other words, given the stimuli randomisation, it is very unlikely that, e.g., voiced items were systematically heard at a higher volume across or within participants. Rather, the absence of such controls is likely to have added a great deal of random noise to the data, such that it is possible we would find even stronger associations given a more strictly controlled auditory environment.

quality. For example, where we have total reduplication, this tells us that the voicing and vowel in the word are not mixed. Thus, effects of reduplication were considered in a separate model. In terms of vowel effects, the model compared mixed vowel quality (e.g., a word which contained both front and back vowels such as /kiku/ or /kugi/) with front and back vowels, as well as front and back vowels with one another. In terms of voicing, the model compared mixed voicing with voiced and voicelessness, as well as voiced and voicelessness with one another. For the reduplication model, total reduplication and no reduplication were compared with each other as well as consonant and vowel only reduplicated words.

Examination of the data before analysis revealed that responses in our dependent measure were not normally distributed; rather, they were skewed with responses heavily favoured on the low end of the scale. In other words, people preferred to make the ball faster generally (recall that a low response means a low mass and thus faster speed). This may have been because differences in speed were not noticeable where the ball was moving slowly, such that the difference between 500 and 400 was not as salient as the difference between 300 and 200. In order to make the data suitable for the linear mixed effects model, it was transformed using a Yeo-Johnson power transformation (Yeo and Johnson, 2000). This transformation was only used to normalise the distribution for analysis; unadjusted means will be reported when discussing the results. Significance was calculated using the pvals.fnc() function in the languageR library[13].

First, a linear mixed effects model which examined possible effects of native language found that the responses of English monolinguals and non-native English speakers did not differ significantly overall or in terms of any of the sound features (all p's >0.05), and henceforth all participants are collapsed into a single group. In terms of voicing, we found that mixed voicing was rated as near significantly slower than items which were entirely voiced (p = 0.06) and items which were entirely voiceless (p = 0.03). As there was no significant difference between voicing and voicelessness (p = 0.8480), this effect is likely related to the significantly faster ratings for reduplicated consonants, explained in detail with the Reduplication results. Mean voicing ratings are shown in Figure 7.

---

[13] This function runs a Markov Chain Monte Carlo (MCMC) for the model in question, and returns the likelihood of the actual result as a p-value.The MCMC function performs ten thousand runs to create a distrubution of possible means, returning a p-value reflecting where the actual results fall within this real distribution. Thus, this p-value (reported here simply as p) is more conservative than one based on traditional t-tests, which would require more extensive correction.
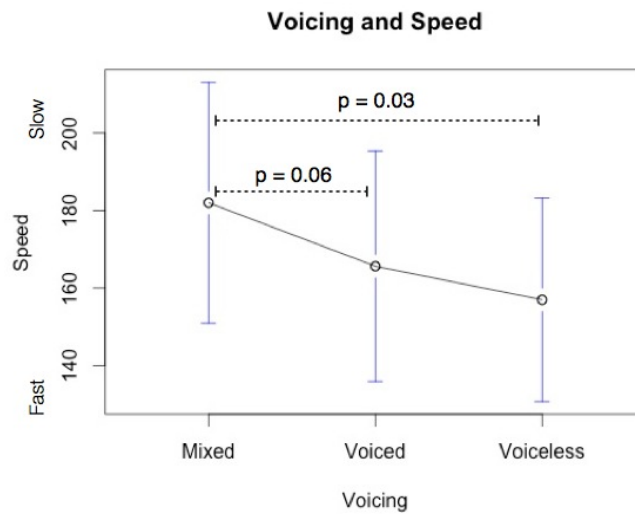
**Figure 6.** Speed ratings in terms of consonant voicing

In terms of vowel quality, participants rated back vowels marginally significantly slower than mixed vowels ($p = 0.05$) and significantly slower than front vowels ($p = 0.005$; see Figure 8). There was no significant difference between high vowels and mixed vowels ($p = 0.2176$), indicating high vowels were not rated as fast per se, but back vowels were considered particularly slow.
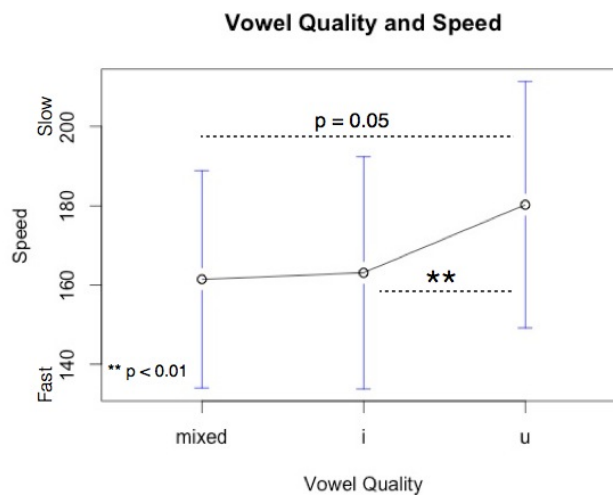
Figure 7. Speed ratings in terms of vowel quality.

In terms of reduplication, total reduplication was not significantly different to no reduplication ($p = 0.1128$), though reduplication in terms of vowels and consonants independently showed significant differences in speed ratings. The difference between vowel only reduplication and total reduplication trended towards significance ($p = 0.09$), with

total reduplication being slightly faster than vowel-only reduplication. There was a marginally significant difference between consonant-only reduplication and no reduplication (p = 0.05), with consonant reduplication being faster than no reduplication at all (see Figure 9). This means that reduplicated consonants were rated the fastest, followed by total reduplication and no reduplication, with vowel-only reduplication resulting in the slowest ratings[14].

Reduplication and Speed
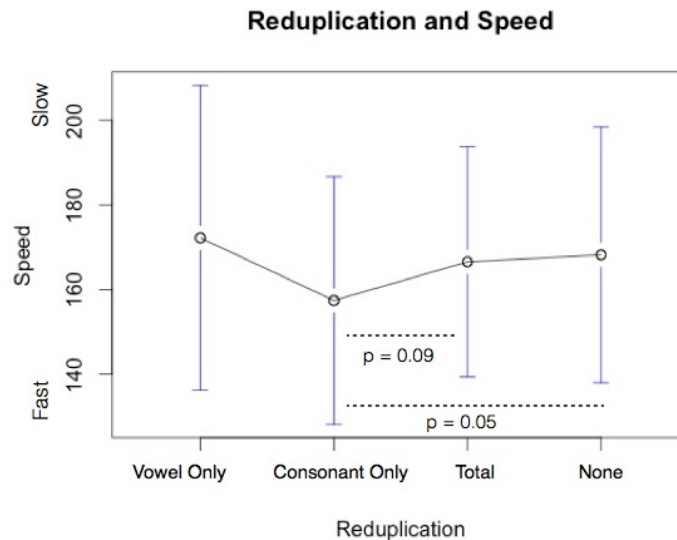
Figure 8. Speed ratings in terms of reduplication.

4.4 Discussion

This study examined cross-modal associations between linguistic sound and motion using a graded direct perceptual matching task. Participants were presented with non-words specifically designed to vary in terms of voicing, vowel quality, and reduplication and asked to change the speed of a bouncing ball to intuitively match the words. The results showed that the sound properties of the non-words had significant influence on participants' speed ratings. Specifically, words with mixed voicing (e.g., containing one voiced and one voiceless consonant) were rated as being slower than words with either all voiced or all voiceless consonants. These sorts of words could also be classified as having consonant only reduplication, and indeed, the reduplication model revealed that words which had reduplicated consonants were rated slightly faster than words with reduplicated vowels, total reduplication, and no reduplication at all. Lastly, results showed that words containing back vowels were rated as significantly slower than those containing front vowels or mixed vowels (e.g., containing one front and one back vowel).

---

[14] This means that vowel only reduplication was likely significantly slower than consonant only reduplication, but this was not examined as the primary aim of our models (based on our hypotheses) was to compare total reduplication and no reduplication. The comparison between vowel-only reduplication and consonant reduplication was omitted because there was no well-motivated hypothesis regarding their difference, and we aimed to keep the total number of comparisons low after Curtin and Schulz (1998).

In terms of voicing, the tentative hypotheses were not confirmed. A hypothesis was proposed wherein the use of voicing to express weight in Japanese ideophones might translate to motion (e.g., perhaps heavier objects move more slowly). In fact, this study found no specific effects of the presence or absence of voicing. Rather, mixed voiced items were significantly slower than both voiced and voiceless items. This is likely due to effects of reduplication, and was echoed by findings from the reduplication model, which suggested that the reduplication of consonants in particular was influential. This indicates that voicing and reduplication were inherently confounded in the design of the items. These two factors were indivisible, as indicated by the need to create separate models to examine the effects. In other words, where a word was completely voiced or voiceless, it also exhibited consonant reduplication. In part, these two factors were confounded because consonant place and manner did not vary within a given word. Future studies might consider voicing and reduplication separately by varying place and manner within words, and examining how these features of articulation might affect speed ratings.

Looking more specifically at reduplication, we expected to find that completely reduplicated items would be rated as faster than items with no redu- plication at all. In fact, there was no significant difference between the presence of absolute reduplication and its absence. However, the results from both the voicing model and the reduplication model seemed to point to a specific role for the reduplication of consonants in particular. This is not enttirely surprising particularly in light of partial reduplication patterns found in Basque, wherein the consonant sound is entirely preserved while the vowel sound changes or alternates (Ibarratxte-Antunaño, 2006). While total reduplication is common among Basque motion ideophones, partial reduplication also ac- counts for a large proportion of motion ideophones. Moreover, the majority of partial reduplication involves consonant reduplication and vowel alternation (Ibarratxte-Antunaño, 2006). Partial reduplication is also widely used across diverse languages (e.g., see Hurch, 2005), including English, where it is used in terms such as *pitter-patter* and *mish-mash*.

The *pitter-patter* variety of partial reduplication makes simultaneous use of both consonant reduplication and vowel alternation, and these properties in particular were found to be significantly related to speed in our data. Results indicate that it is not only the use of reduplicated consonants, but the combination of reduplicating consonants and alternating vowels, which was matched with speed. This is evident in the fact that consonant only reduplication was rated faster than total reduplication (although this was only marginally significant); if reduplication of consonants alone led to in- creased speed ratings, totally reduplicated items would not have been rated any differently, since the consonants in these items were also reduplicated. Although the specific hypothesis with regards to reduplication (that total reduplication would be rated as faster than other forms of reduplication) was not borne out, the data do show that partial reduplication had significant effects on speed ratings. Considerably more work is needed to investigate how reduplication relates to different types of motion. Crucially, this experiment examined the speed of a bouncing ball, and it is possible that the action of bouncing introduced an additional dimension. For example, perhaps vowel alternation with reduplicated consonants was especially suited to increased speed to encode both the change in direction which took place when the ball bounced as well as the overall speed of the ball. Indeed, Childs (1994:184) suggests that mixed vowel quality and consonants may be associated with irregularity, and bouncing could be considered a form of irregular motion[15].

In terms of vowel quality, this experiment found that back vowels were rated slower than front vowels or mixed vowels, in accordance with the hypothesis. This pattern was pre

_____

[15] The author thanks an anonymous reviewer for this suggestion.

dicted based on other studies in sound symbolism, particularly size sound symbolism, which shows that high and front vowels are associated with smaller size. This study extends that finding to speed, as predicted by Berlin (2006), showing that back vowels in particular are perceived as slow. While the predicted slow quality of back vowels was borne out, participants did not rate front vowels to be particularly fast.

5. Conclusions

This paper has presented evidence of shared cross-modal associations between linguistic sound and motion. The experiment showed that reduplication of consonants paired with alternation of vowels in particular led to higher ratings in speed, and back vowels were rated as being particularly slow. These associations are limited to one particular type of motion, specifically, the speed of a bouncing ball, a type of bi-directional motion. Future work should focus on looking at other aspects of visual motion. Even within the domain of visual speed, it may be valuable to assess associations between the features of linguistic sound studied here and uni-directional motion. Such an assessment would be able to test the hypothesis outlined above that the pattern of consonant reduplication and vowel alternation may be especially suited to encoding bouncing. If this is indeed the case, associations between uni-directional speed and linguistic sound may indeed favour total redupli- cation for higher speed.

There also remain several unexplored avenues with regards to associations between motion and linguistic sound. The current study focused on a very particular feature of motion in the form of speed, which is not necessarily known to be specifically encoded in language. However, other manners and paths of motion are ripe for exploration, for example, rotation around an axis appears to be a recurring theme among Japanese ideophones (Kita, 1997). Associations could also be tested more implicitly, using learnability or reaction time measures. Future explorations should examine a broader set of linguistic features perhaps including variation in place and manner of articulation, as well as considering other aspects of motion using a wide range of methodological approaches.

References
Ahlner, F. & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38(1), 298– 348.
Aitchison, J. (2000). *The seeds of speech: Language origin and Evolution*. Cambridge University Press: Cambridge, UK.
Beavers, J., Levin, B. & Tham, S.W. (2010). The typology of motion expressions revisited. *Journal of Linguistics*, 46(2), 331–377.
Berlin, B. (2006). The first congress of ethnozoological nomenclature. *Journal of the Royal Anthropological Institute* 12, S23–S24.
Bickerton, D. (1990). *Language and Species*. Chicago, IL: Chicago University Press.
Brown, J.E. (2012), September. *The evolution of symbolic communication: An embodied perspective*. Ph.D thesis, University of Edinburgh.

Buhrmester, M, Kwang, T. & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science* 6(1), 3–5.

Childers, J.B. & Tomasello, M. (2006). Are nouns easier to learn than verbs? Three expeirmental studies. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs*, pp. 311–335. New York: Oxford University Press.

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

Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. New York: Praeger.

Colonius, H. & Arndt, P. (2001). A two-stage model for visual and auditory interaction in saccadic latencies. *Perception and Psychophysics,* 63, 126–146.

Crisinel, A-S, Jones, S. & Spence, C. (2012). The sweet taste of maluma: Cross-modal associations between tastes and words. *Chemical Perception*, 5, 266-273.

Curtin, F. & Schulz, P. (1998). Multiple correlations and Bonferroni's correction. *Biological Psychiatry* 44(8), 775–777.

Cuskley, C. (2013, March). *Shared cross-modal associations and the emergence of the lexicon*. Ph.D. thesis, University of Edinburgh.

Cuskley, C. & Kirby, S. (2013). Synaesthesia, cross-modality, and language evolution. In J. Simner & E. Hubbard (Eds.), *Oxford Handbook of Synaesthesia,* pp. 869-907. Oxford, UK: Oxford University Press.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology* 52 (3), 259–268.

de Sousa, H. (2011). Changes in the language of perception in Cantonese. *Senses and Society* 6(1), 38–47.

Deutscher, G. (2005). *The Unfolding of Language*. London: William Heinemann.

Dingemanse, M. (2011). Ideophones and the aesthetics of everyday language in a West-African society. *Senses and Society* 6(1), 77–85.

Fitch, T.W. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America* 102, 1213–1222.

Gallace, A., Boschin, E. & Spence, C. (2011). On the taste of 'bouba' and 'kiki': An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience* 2(1), 34–46.

Gasser, M. (2004). The origins of arbitrariness in language. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pp. 434–439. Hillsdale, NJ: LEA.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj, (Ed.), *Language development Vol. 2*, pp. 38-62. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gosling, S.D., Vazire, S., Srivastava, S. & John, O. (2004). Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist* 59, 93–104.

Green, K.P. (1997). Acoustic cues to place of articulation and the McGurk effect: The role of release bursts, aspiration, and formant transitions. *Journal of Speech, Language, and Hearing Research* 40, 646-665.

Hamano, S. (1998). *The Sound-Symbolic System of Japanese.* Stanford, CA: CSLI Publications.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 325–336.

Henrich, J., Heine, S.J. & Norenzayan, A. (2010). The weirdest people in the world?*Behavioral and Brain Sciences* 33, 62–135.

Hinton, L., Nichols, J. & Ohala, J.J. (1994). Introduction: Sound symbolic processes. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound Symbolism*, pp. 1–14. Cambridge, UK: Cambridge University Press.

Hockett, C. (1960). The origin of speech. *Scientific American* 203, 88–96.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences* 79, 2554-2558.

Hurch, B., Ed. (2005). *Studies on Reduplication*. Berlin: DeGruyter.

Ibarretxe-Antunaño, I. (2006). 'Ttipi-ttapa, ttipi-ttapa. . . korrika!!!': Motion Sound Symbolism in Basque. *ASJU* 40(1-2), 499–518.

Imai, M., Kita, S., Nagumo, M. & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition* 109, 54–65.

Imai, M., Li, L, Haryu, E., Okada, H., Hirsh-Pasek, K. & Golinkoff, R. (2008). Novel noun and verb learning in Chinese, English, and Japanese speaking children. *Child Development* 79, 979–1000.

Irwin, F.W. & Newland, E. (1940). A genetic study of the naming of visual figures. *Journal of Psychology* 9, 3–16.

Iwasaki, N., Vinson, D.P. & Vigliocco, G. (2007a). How does it hurt, 'kiri-kiri' or 'siku-siku'? Japanese mimetic words of pain percieved by Japanese speakers and English speakers. In *Applying theory and research to learning Japanese as a foreign langauge*, pp. 2–19. Cambridge Scholars: Cambridge, UK.

Iwasaki, N., Vinson, D.P. & Vigliocco, G. (2007b). What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese-Language Education around the Globe* 17, 53–78.

Kantartzis, K., Imai, M. & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science* 35, 575–586.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35, 379–415.

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

Köhler, W. (1947). *Gestalt Psychology (2nd Edition)*. New York, NY: Liveright.

Kovic, V., Plunkett, K. & Westerman, G. (2010). The shape of words in the brain. *Cognition* 114, 19–28.

Macdonald, J. & McGurk, H. (1978). Visual influences on speech perception processes. *Attention, Perception and Psychophysics* 24(3), 253– 257.

Majid, A. & Levinson, S.C. (2011). The senses in language and culture. *Senses and Society* 6(1), 5–18.

Massaro, D.W. (2004). From multisensory integration to talking heads and language learning. In G. Calvert, C. Spence, and B. Stein (Eds.), *Handbook of multisensory processes*, pp. 153–176. Cambridge, MA: MIT Press.

Maurer, D., Pathman, T. & Mondloch, C.J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science* 9(3), 16–322.

Mikone, E. (2001). "Ideophones in the Balto-Finnic langauges". In F.K.E. Voeltz & C. Killian-Hatz (Eds.) *Ideophones*, pp. 223–233. Amsterdam: John Benjamins.

Monaghan, P., Christiansen, M.H. & Fitneva, S. (2011). The arbitrariness of the sign: learning advantages from the structure of the vocabulary. J*ournal of Experimental Psychology: General* 140, 325–347.

Monaghan, P., Mattock, K. & Walker, P. (2012). The role of sound symbolism in word learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 38, 1152–1164.

Nagumo, M., Imai, M., Kita, S., Haryu, E. & Kajikawa, S. (2006). Sound iconicity boostraps verb meaning acquisition. Paper presented at the XVth International Conference of Infant Studies, Kyoto, Japan.

Navarra, J. & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of L2 sounds. *Psychological Research* 71, 4–12.

Nielsen, A. & Rendall, D. (2011). The sound of round: Evaluating the sound symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psycholgy* 65(2), 115– 124.

Nielsen, A. & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition* 4(2), 115–126.

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

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 University Press.

Peña, M., Mehler, J. & Nespor, M. (2011). "The role of audiovisual processing in early conceptual development." *Psychological Science* 22 (11), 1419–1421.

Perniss, P., Thompson, R.L. & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* 1, 1–15.

Poirier, C., Collignon, O., Devolder, A., Renier, L., Vanlierde, A., Tranduy, D. & Scheiber, C. (2005). Specific activation of the V5 brain area by auditory motion processing: an fMRI study. *Cognitive Brain Research* 25(3), 650–658.

Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., Veraart, C. & Volder, A.D. (2006). Auditory motion perception activates visual motion areas in early blind subjects. *NeuroImage* 31(1), 279 – 285.

Ramachandran, V. & Hubbard, E. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies* 8(1), 3–34.

Rogers, S.K. & Ross, A.S. (1975). A cross-cultural test of the maluma-takete phenomenon. *Perception* 5(2), 105–106.

Rubino, C. (2005). Reduplication: Form, function and distribution. In B. Hurch (Ed.), *Studies on Reduplication*, pp. 11–30. Berlin: De- Gruyter.

Saji, N., Akita, K., Imai, M., Kantartzis, K. & Kita, S. (2012). The Internal Structures of Sound-Symbolic Systems: the Universal and Language-Specific Portions of Sound Symbolism. Paper presented at CogSci 2012, Sapporo Convention Center, Japan.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology* 12, 239–255.

Saussure, F. (1959). *Course in general linguistics*. New York, NY: Philosophical Library.

Sears, D.O. (1986). College sophmores in the lab: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology* 51, 515–530.

Shams, L., Kamitani, Y. & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research* 14 (1), 147–152.

Simner, J., Cuskley, C. & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition *Perception* 39, 553–569.

Slobin, D.I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Stromqvist & L. Verhoeven (Eds.) *Relating events in narrative: Vol. 2. Typological and contextual perspectives*, pp. 219–257. Mahwah, NJ: Lawrence Earlbaum.

Sonesson, G. (1997). The ecological foundations of iconicity. In I. Rauch & G.F. Carr (Eds.), *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS, Berkeley, June 12-18 1994*, pp. 739-742. Berlin & New York: Mouton de Gruyter.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C. & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research* 14 (1), 139 – 146.

Talmy, L. (2000). *Toward a cognitive semantics: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Thompson, P. & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology* 64(12), 2932–2404.

Trask, R.L. (1997). *The History of Basque*. New York, NY: Routledge.

Tufvesson, S. (2011). Analogy-making in the Semai sensory world. *Senses and Society* 6(1), 86–95.

Voeltz, F.K. & Killian-Hatz, C. (Eds.) (2001). *Ideophones*. Amsterdam: John Benjamins.

Walker, P., Bremner, J.G., Mason, U., Spring, J., Mattock, K., Slater, A. & Johnson, S.P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science* 21(1), 21–25.

Werker, J., Cohen, L.B., Lloyd, V., Casasola, M. & Stager, C.L. (1998). Acquisition of word-object associations by 14 month old infants. *Developmental Psychology* 34, 1289–1309.

Werner, H. (1957). *Comparative Psychology of Mental Development (Revised Edition)*. New York: International Universities Press.

Werner, H. & Wapner, S. (1952). Toward a general theory of perception. *Psychological Review* 59, 324–338.

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93, 10–19.

Wienold, G. (1995). Lexical and conceptual strucutures in expressions for movement and space: with reference to Japanese, Korean, Thai, and Indonesia as compared to English and German. In U. Egli, P.E. Pause, C.Schwarze, A. von Stechow, & G. Wienold (Eds.), *Lexical knowledge in the organization of language*, pp. 301–340. Amsterdam: John Benjamins.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication* 18, 47–67.

Yeo, I-K. & Johnson, R.A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* 87, 954–959.

Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognition and Development* 13(2), 232–265.