Sonority, Size and Shape in Sound Symbolism

Research Project/Dissertation (Paper 9265)

Submitted on: Wednesday 19th September 2012

Candidate No:750749

Cover Page

ABSTRACT

Sound symbolism refers to the systematic mappings between phonological properties of labels and perceptual properties of their referents (Maurer et al., 2006). Despite a large body of evidence on its existence, the exact phonological properties that cause such symbolic effects are less clear. The present study tested whether vowels or consonants are more salient in driving the size and shape symbolic effects, and whether sonority, a phonetic parameter related to the overall acoustic energy of a speech sound (Gussenhoven & Jacobs, 2005, p.138) reliably predicts size and shape symbolism. The results showed different perceptual patterns for the two types of sound symbolism. For size symbolism, sonority was the determining factor in that: (1) the more sonorous vowels were more powerful in driving size symbolism than the less sonorous consonants; (2) consonant sonority linearly predicted size symbolism. For shape symbolism, however, sonority was less effective because: (1) vowels and consonants are equally salient in predicting shape symbolism; (2) consonant sonority influenced roundness, but the effect was non-linear. These results contradict Nielsen and Rendall's (2011) claim that consonants are the driving force behind shape symbolic effects; and the discrepancy between size and shape symbolism results was also not reported in previous research. It is suggested that size and shape symbolism may be different in terms of their underlying sound symbolic mechanisms.

Keywords: sound symbolism, sonority, size, shape

CONTENTS

1. Introduction

1.1 Background concerning sound symbolism research

The debate regarding the relationship between the sound and the meaning of words dates back at least as far as Plato, who recorded a dialogue between Hermogenes and Socrates: Hermogenes argued that *"there is no name given to anything by nature; all is convention and habit of the users"*, while Socrates claimed that *"names have by nature a truth"*. At the end of the dialogue, both of the debaters find themselves in a stalemate: some sound forms do indeed seem to 'resemble' their meanings, while many others do not (cf. Jowett, 2009). The debate continues today among linguists, anthropologists, psychologists, marketers, and philosophers. Following Ferdinand de Saussure (1959), the *"father of modern linguistics"* (see Lyons, 1968), it is held as a doctrine of mainstream linguistics that the link between the signifier (sound) and the signified (concept) is arbitrary. The idea of a "tree" bears no inner relationship to the sound sequence "t-r-i" which serves as its signifier in English, and the signifier is not universal cross-linguistically: it is "arbre" in French, "tsurī" in Japanese and "shu" in Chinese, etc. Saussure also rejected onomatopoeia and interjections as challenges to the arbitrariness of linguistic signs because they are *"never organic elements of a linguistic system"* (p. 69), and are not universal either: The French and English onomatopoeic words for a dog's bark are "Oua Oua" and "Bow Bow", and the French and English pain interjections are "Aie" and "Ouch". Besides, some words such as French *fouet* "whip" or *glas* "knell" may sound "suggestive" in meaning, but they actually have evolved from non-onomatopoeic origins: *fouet* is derived from Latin *fagus* "beech-tree", and *glas* is from *classicum* meaning "the sound of a trumpet". Thus the quality of their present sounds is just a "fortuitous" result of phonetic evolution.

Still, some others stand by Socrates. Saussure's contemporary Otto Jespersen (1922) surveyed a list of words for "little", "child or young animals", "small things", and diminutive suffixes in a number of languages, and concluded that the high-front-unround vowel "i" was commonly used cross-linguistically to indicate *"what is small, slight, insignificant, or weak"*. This claim was later supported by Sapir's (1929) experimental study, in which 74.6%-96.4% of the 500 participants associated non-words containing the vowel "a" (e.g., "mal") with objects of larger size, and non-words containing the vowel "i" (e.g., "mil") with objects of smaller size. Sapir argued that although the "referential symbolism" of linguistic forms was arbitrary over *"an uncontrollably long period of historical development"*, there was an "expressive symbolism" of language which was *"more fundamental"* and *"psychologically primary"*. In the same year, Köhler (1929, pp. 224-225) noted that language did not use words to describe sensory experiences *"in a haphazard way"*: in primitive languages, *"names of things and events often originate according to the similarity between their properties in vision or touch and certain sounds or auditory wholes"*. He reported a study in which people *"decide with ease"* that the curvy shape was "baluba" ("maluma" in Köhler, 1947) and the angular shape "takete" (see Figure 1).

Figure 1. The visual stimuli used by Köhler (1929).

Other early experimental studies on sound/phonetic symbolism include those reported by Fox (1935), Newman (1933), Uznadze (1924), etc., and their results univocally suggested a form of intrinsic connection between sounds and meanings. This has led to a constant stream of research on sound symbolism over the past century (see Hinton et al., 1994; Nuckolls, 1999; Spence, 2011, for reviews), mainly focusing on: (1) the existence of particular sound-meaning associations (e.g., Boyle & Tarte, 1980; Chastaing, 1958; Holland & Wertheimer, 1964; Johnson, 1967; Lindauer, 1990; Taylor, 1963); (2) the cross-linguistic universality of sound-meaning associations (e.g., Brown et al., 1955; Davis, 1961; Gebels, 1969; Osgood, 1960; Roger & Ross, 1975), and (3) the early development of sound-meaning associations (e.g., Braaten, 1993; Irwin & Newland, 1940). However, the results of such studies often triggered scepticism or were viewed as marginal in linguistics and psycholinguistics at that time (e.g., see Hamano, 1998; Nuckolls, 1999). For example, Newmeyer (1993, p. 758) stated that *"the number of pictorial, imitative, or onomatopoetic non-derived words in any language is vanishingly small"*. The experimental methods of the early studies (e.g., forced-choice, English-foreign-pairs[1]) were also criticised for being biased and inappropriate (e.g., Bentley & Varon, 1933; Brackbill & Little, 1957; Brown, 1958; Brown & Nutall, 1959). Bentley and Varon used a method of "free association" by asking participants to give the synonym or describe the meaning of a given sound, and they found no indication of symbolism in the responses. Brackbill and Little asked participants to decide whether a pair of words from two different languages was the same or different in meaning, and the results showed much lower success rate than were obtained by Brown et al. (1955), using a English-foreign-pairs method.

The recent resurgence of research interest on sound symbolism is partly indebted to a couple of publications by Ramachandran and Hubbard (2001, 2003). They reported an anecdotal study similar to Köhler's (1929), but with different non-word stimuli ("bouba" and "kiki") and slightly different shapes (see Figure 2), and claimed that 95% of the participants associated "bouba" with the rounded shape and "kiki" with the jagged shape[2] (see also Bremner et al., in press). Ramachandran and Hubbard speculated that this bias towards mapping certain sound contours onto certain vocalizations might stem from cross-activation between a sensory (i.e., auditory) and a motor map (i.e., Broca's area), which may explain the

[1] The participants were shown pairs of English antonyms and pairs of equivalent foreign language antonyms, and were asked to match the two English words with their foreign language equivalents (e.g., Brown et al., 1955).

[2] Though note no proper description of the underlying research was given in the paper.

origin of language, and may be linked to synaesthesia, a neurological condition in which people experience sensations in one modality when a second modality is stimulated (see Cytowic, 2002).

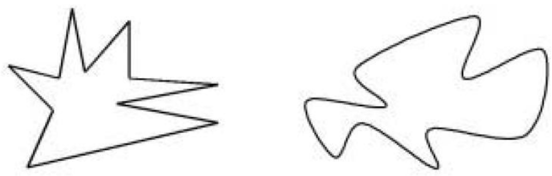


Figure 2. The visual stimuli used by Ramachandran and Hubbard (2001).

The "Bouba-Kiki effect" thus became a famous example of sound symbolism, and provoked a number of studies on multisensory sound-meaning correspondences, not only between sound and shape or sound and size, but also between sound and colour (e.g., Ward et al., 2006; Wrembel, 2007; Wrembel & Rataj, 2008), sound and taste / smell (e.g., Crisinel & Spence, 2009; Gallace et al., 2011; Mesz et al., 2011; Simner et al., 2010; see Spence, 2012; and Knöferle & Spence, in press, for reviews), sound and touch (e.g., Yau et al., 2009), sound and spatial position (e.g., Loomis et al, 2002), sound and emotion (e.g., Hasada, 1998; Chuenwattanapranithi et al., 2008), etc. For example, Ward et al. (2006) reported that both normal participants and synaesthetes display a tendency to associate lower pitched sounds with darker colours and higher pitched sounds with lighter colours, suggesting a common underlying neural mechanism; Simner et al. (2010) found that when manipulating the formant frequencies F1, F2, the spectral balance, and the volume of a vowel sound from low to high, the associated tastes change from "sweet" to "salty", "bitter", and "sour" on a continuum; Chuenwattanapranithi et al., meanwhile, reported that vowels synthesized with a dynamically lengthened vocal tract and lowered fundamental frequency F0 were associated with anger and that the opposite was true for the perception of happiness.

In terms of experimental methods, many of the recent studies on sound symbolism still used Köhler's (1929) and Sapir's (1929) forced-choice methodology (e.g., Berlin, 2006; Maurer et al., 2006; Nielsen & Rendall, 2011; Ramachandran & Hubbard, 2001), which has often been criticised as: (1) relying on a limited number of carefully selected stimuli, and (2) biasing the participants with transparent manipulations (e.g., Kovic et al., 2009; Westbury, 2005). More subtle behavioural studies (e.g., using the Implicit Association Task (IAT), or the Implicit Interference Task) and neuroimaging studies (e.g., electroencephalography, EEG, functional magnetic resonance imaging, fMRI) have been conducted in order to demonstrate sound symbolism. For example, Westbury asked the participants to determine whether the word/non-word inside of a jagged or rounded frame (see Figure 3) was a real word or a non-word. The word/non-word stimuli consisted of stop consonants (e.g., "d, k, p, t"), continuant consonants (e.g., "m, l, n"), or both. The reaction times (RTs) showed interference patterns consistent with Köhler's claims: the RTs were longer when the stops were presented within the rounded frames and the continuants were presented within the jagged frames.

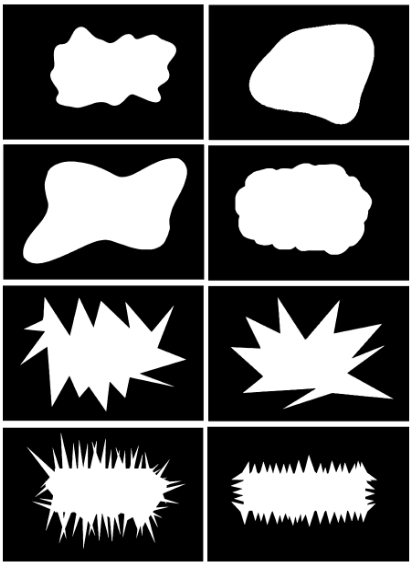


Figure 3. Examples of frames used in Westbury (2005).

Kovic et al. (2009) used a categorization task to measure sound symbolism neurologically with EEG. Participants first "learned" that the rounded stimuli were "dom" and the angular stimuli were "shick" (congruent condition) or the rounded stimuli were "shick" and the angular stimuli were "dom" (incongruent condition), they then indicated whether the stimuli were categorized correctly in the test phase. The analysis of the event-related potential (ERP) results demonstrated that congruent audio-visual stimuli elicited a stronger, negative-peaking wave between 140-180 ms after visual onset than did the incongruent stimuli at both the O1 and O2 electrode sites (located above the primary visual cortex). Additionally, a recent fMRI study by Peiffer-Smadja (2010) revealed that during passive perception, there were increased deactivations for congruent sound-shape pairs (e.g., "teki"- angular shapes; "lujo"-round shapes) as compared to incongruent pairs (e.g., "teki"-round shapes; "lujo"-angular shapes) in the left and right lateral frontal gyri. The neuroimaging evidence again supported that participants were sensitive to the bouba-kiki effect.

One primary benefit of sound symbolism appears to be its facilitation of language acquisition, both for adults and for children (Maurer et al., 2006). For example, Nygaard et al. (2009) showed that adult English participants were more accurate and rapid in identifying the previously learned English translations of Japanese words when they were paired with the English equivalents or antonyms than when they were paired with random English words. Imai et al. (2008) demonstrated that 3-year-old Japanese children were able to generalise the meaning of novel verbs to the same-manner action performed by a different actor when the verbs were sound-symbolic, but were unable to generalize the meaning of the novel verbs when they were not sound-symbolic.

The principles of sound symbolism can also benefit advertisers and marketers when it comes to creating more suitable brand names (Abel & Glinert, 2008; Klink, 2000, 2001; Klink & Athaide, 2012; Lowrey & Shrum, 2007; Ngo et al., 2011; Yorkston & Menon, 2004; see Spence, 2012, for a review). For example, Klink (2000) surveyed 31 pairs of invented brand names (e.g., "Nidax"-"Nodax") regarding their ability to communicate particular product attributes (e.g., the thickness of ketchup) and found that labels containing front vowels were perceived as *"smaller, lighter (relative to darker), milder, thinner, softer, faster, colder, bitter, more feminine, friendlier, weaker, lighter (relative to heavier), and prettier*" (p. 10); Yorkston and Menon demonstrated that an ice-cream named "Frosch" was expected to be creamier than one named "Frisch"[3]; Lowrey and Shrum showed that brand names containing front vowels were preferred when the symbolic meaning (e.g., small, sharp) was positive for a product category (e.g., two-seater convertible, knife), but were preferred less when the symbolic meaning was negative for the product (e.g., sport utility vehicle, hammer). Ngo et al. reported that chocolate with 70% and 90% cocoa content are associated with the sounds "takete" and "tuki", whereas 30% cocoa chocolate are more likely to be associated with "lula" and "maluma". These results suggest that sound symbolism is an ecologically valid phenomenon that may affect our everyday behaviour and decision-making.

1.2 Hypotheses for sound symbolism

1.2.1 The "what" question

With converging evidence, the existence of sound symbolism is firmly established, but the exact correspondence between sound features and symbolic meanings still remains something of a puzzle. A sound unit such as a phoneme, stress, pitch, duration, or tone have all been reported to involve symbolic meanings (see Table 1).

Table 1. Summary of the crossmodal correspondence of different sound units.

Sound unit	Crossmodal correspondence	Study
Phoneme	[o u], [l m n b]—round	Kovic et al. (2009); Peiffer-Smadja (2010); Westbury (2005)

[3] Though note it is creaminess expectations not creaminess experience that were tested.

	—large	Berlin (2006); Maurer et al. (2006); Newman (1933); Taylor & Taylor (1962)
	—bright (colour)	Wrembel (2007); Wrembel & Rataj (2008)
	[i e], [p t k]—angular	Kovic et al. (2009); Peiffer-Smadja (2010); Westbury (2005)
	—small	Berlin (2006); Maurer et al. (2006); Newman (1933); Taylor & Taylor (1962)
	—dark (colour)	Wrembel (2007); Wrembel & Rataj (2008)
Pitch	high—high (elevation)	Ben-Artzi & Marks (1995); Bernstein & Edelstein (1971); Evans & Treisman (2010); Melara & O'Brien (1987); Patching & Quinlan (2002)
	—bright (colour)	Marks (1987); Ward et al. (2006)
	—light (weight)	Marks (1987); Martino & Marks (1999); Melara (1989)
	—angular	Marks (1987)
	—small	Evans & Treisman (2010); Gallace & Spence (2006)
	—sour	Crisinel & Spence (2009); Simner et al. (2010)
Stress	stressed—intensified meaning	Kawahara et al. (2008)
Tone	high—small	Matisoff (1994)
Duration	lengthened—intensified meaning	LaPolla (1994)
Amplitude	loud—bright (colour)	Bond & Stevens (1969); Marks (1987)

These sound units can be classified into segmental features (characteristics of the individual phonemes) and suprasegmental features (characteristics of the syllable, the utterance, or the sentence) in phonological theory (Snow, 2001), as shown in Table 2.

Table 2. Segmental and suprasegmental features in phonology.

Supra-segmental feature			Stress Pitch Tone Duration Amplitude	
Segmental feature	Consonant	Major class	[+/−syllabic]	vowels / consonants
			[+/−consonantal]	liquids, nasals, fricatives, stops / vowels, glides
			[+/−approximant]	vowels, glides, liquids / nasals, fricatives, stops
			[+/−sonorant]	vowels, glides, liquids, nasals / fricatives, stops
		Laryngeal	[+/−voice]	vibration / non-vibration of the vocal folds
			[+/−spread glottis]	glottal friction / no frication
			[+/−constricted glottis]	vocal folds together and tense / no constriction
		Manner	[+/−continuant]	stops / fricatives, nasals, liquids, glides, vowels
			[+/−strident]	[s, z, tʃ, dʒ, ʃ, ʒ] / others

		[+/−nasal]	nasals / others
		[+/−lateral]	airflow through the tongue / not
	Place	[LABIAL]	[w, m,, b, p, f, v]
		[CORONAL]	[l, r, n, z, ð, ʒ, d, s, ʃ, θ, t, tʃ, dʒ]
		[DORSAL]	[y, k, g, ŋ]
		[GLOTTAL]	[ʔ, h]
Vowel		[+/−high]	[i I u ʊ] / others
		[+/−low]	[æ a ɒ] / others
		[+/−back]	[o ɔ u ʊ ʌ ɒ] / others
		[+/−front]	[i I e ɛ æ] / others
		[+/−tense]	[o i e æ u] / others
		[+/−round]	[o ɔ u ʊ ɒ] / others

Despite consistent findings on the association between suprasegmental features and symbolic meanings, the association between segmental features and symbolic meanings is currently less clear. Since Sapir (1929), various phonological classifications of the two phoneme groups ([o u m b] and [i e t k]) have been proposed in the literature on sound symbolism, but there is a lack of consensus as to which classification is the most reliable one and/or has the highest impact in terms of sound symbolism[4] (see Table 3).

Table 3. The phonological classification of sound symbolic phonemes in some recent studies.

Study	"big/round" sound	"small/angular" sound	Segmental feature
Abel & Glinert (2008)	voiced consonants	voiceless consonants	[+/−voice]
Klink (2000)	back vowels high frequency consonants	front vowels low frequency consonants	[+/−back][+/−front] [+/−continuant] [+/−voice]
Lowrey & Shrum (2007)	back vowels	front vowels	[+/−back][+/−front]
Nielsen & Rendall (2011)	sonorant consonants	strident consonants	[+/−sonorant] [+/−strident]
Peiffer-Smadja (2010)	rounded back vowels voiced continuants	unrounded front vowels voiceless stops	[+/−round] [+/−voice] [+/−continuant]
Sweeney et al. (2012)	rounded sounds	unrounded sounds	[+/−round]
Westbury (2005)	continuants	stops	[+/−continuant]
Yorkston & Menon (2004)	low back vowels	high front vowels	[+/−high][+/−low] [+/−back][+/−front]

[4] Note that some of the segmental features are highly correlated (e.g., in English, all the [+round] vowels are also [+back], and all the [-round] vowels are also [+front]; see Table 2), making it difficult to orthogonally manipulate the features.

Another unanswered question is whether vowels or consonants are more salient in terms of driving sound symbolism effects. Following Sapir (1929), a number of studies have claimed that sound symbolic effects are primarily attributable to the phonological features of the vowels (Lowrey & Shrum, 2007; Maurer et al., 2006; Newman, 1933; Ramachandran & Hubbard, 2001; Yorkston & Menon, 2004), yet many others emphasize the role of consonants (Abel & Glinert, 2008; Westbury, 2005). A recent study by Nielsen and Rendall (2011) specifically investigated this question. They presented participants with pairs of incongruent sounds consisting of "round" consonants and "angular" vowels (e.g., "maleme") or "angular" consonants and "round" vowels (e.g., "takuta"), and asked them to choose which one mapped on to a curvy shape and which one mapped to an angular shape. The results suggested that participants were tracking the sound symbolic meaning of the consonants when they were incongruent with the vowels in terms of their predictions of "roundness" (e.g., "maleme" is curvy and "takuta" is angular).

Yet their conclusion should be interpreted with some caution for several reasons: (1) Nielsen and Rendall excluded the vowels "o" and "i", which are the most "round" and "angular" vowels reported in the literature (see Nuckolls, 1999 for a review); (2) they only tested a limited set of consonants ("p, t, k" and "m, l, n"), which makes the results less conclusive; (3) their non-word stimuli included two-syllable and three-syllable strings (e.g., kaatoo, muhbeemee), which introduced the confound of stress (so, for example, [`kaatoo] and [kaa`too] may elicit rather different sound symbolic effects), and (4) they formed some of their sound stimuli by swapping the original stimuli pairs in Köhler (1929) and Maurer et al. (2006) (e.g., "maluma-takete" became "maleme-takuta"), yet it is uncertain whether the vowel quality remained the same for the changed pairs. For example, the angular-sounding "a [ʌ]" in the original word "takete" may become the round-sounding "a [a]" in the swapped word “maleme”. It is thus unknown whether “maleme” being round is due to the round consonants “m, l” or the round vowel “a [a]”.

1.2.2 The “why” question

The explanation for sound symbolism is also a subject of intense debate currently. To date, the evidence seems to be consistent with a neurological cross-connections account as suggested by Ramachandran and Hubbard (2001; 2003). They postulated that the motor cortical activation for the act of articulation drove the “Bouba-Kiki effect”. For example, articulating “kiki” involves sharp inflections of the tongue on the palate, while pronouncing “bouba” involves the rounding of the lips, hence they relate to the sharpness and roundness of the images respectively. But acoustic features of the sound itself could also be the driving force, as argued by Ohala (1996). This author suggested that motor articulation is not necessarily involved in sound perception, since infants and animals who are incapable of articulating speech sounds, can nevertheless differentiate them.

Despite numerous possible causes for a structural correspondence in the human brain, it is also possible that crossmodal correspondence between speech sounds and visual shapes is purely statistical (see Spence, 2011, for a review). Since semantic contrasts are generally marked by phonetic contrasts, these markings lead to the non-random distribution of the words of the language (Berlin, 2006). Namely, language starts with arbitrary pairings of sounds and meaning, only that two phonetically contrastive sounds are used to convey two semantically contrastive meanings, for example, in any given language, the front high vowel “i” can be used for “smallness” or “largeness” while the low back vowel “a” on the other end of the spectrum is used for the semantically contrasted “largeness” or “smallness”. This contrastive explanation is supported by Brackbill and Little’s (1957) study in which they presented antonymic word pairs from two different languages, and asked participants to guess which words had the same meanings. They found that *"where meaning contrasts are not as great, correct guessing becomes much more difficult"* (Brackbill & Little, 1957, p. 318). Farmer et al. (2006) also reported that participants' RTs differed when presented with noun-sounding and verb-sounding nouns, indicating differing sound patterns for nouns and verbs. Yet there is a lack of evidence to suggest that a hardwired noun-verb symbolism exists in the human brain. Thus, the claimed neurological correspondence may just be the result of statistical observation of our language environment.

One way in which to differentiate between the neurological and statistical accounts is to examine whether sound symbolism occurs in young children who are relatively inexperienced with language. Maurer et al. (2006) investigated the bouba-kiki effect with preschool children (mean age = 2.8 years), and found that children showed the same cross-sensory matching tendencies as adults. This was then attributed to infants' synaesthesia-like, cross-modal linkages which these authors proposed to be a general feature of neural organization in infants. They, in turn, argued that the cross-modal linkages are especially functional in infancy as they are thought to be gradually pruned during development to yield the relative independence of different sensory systems of adults (Maurer & Mondloch, 2004; though see Deroy & Spence, submitted). But even if pre-linguistic children also exhibit sound symbolic mappings (Walker et al., 2010), it could still be argued that they have already had sufficient time to learn the statistical correspondence in their language environment (see Aslin et al., 1998).

Another way in which to distinguish between the two accounts is to examine the cross-linguistic observations concerning sound symbolism. If language starts with arbitrary pairings of phonemes and meanings, then the sound symbolic effects may not be universal. Unfortunately, cross-linguistic studies have documented mixed findings (see Hinton et al., 1994; Nuckolls, 1999, for reviews). On the one hand, sound symbolism, especially size symbolism (e.g., /i/ conveys physical smallness), appears to be universal across languages (e.g., Jespersen, 1922; Ohala, 1994; Sherzer, 1976). On the other hand, considerable variance in sound symbolism has also been observed (e.g., Atzetd & Gerard, 1965; Diffloth, 1994; Taylor & Taylor, 1965). For example, Diffloth reported a reverse pattern in Bahnar where high vowels (e.g., /i/) are associated with large size and low vowels (e.g., /a/) are associated with small size (but see Tsur, 2006, for alternative interpretations of the results).

Thompson and Estes (2011) proposed a new way in which to distinguish the neurological and the statistical accounts. They argued that the statistical account can only mark two contrastive values of the given physical property, such as "small" and "large". The neurological explanation, however, may mark continuous degrees of the given physical property, such as "small", "medium", and "large" because gestures and sound properties are both continuous. For example, a mid-sized object could be indicated by a moderate hand or mouth gesture and by a midrange amplitude. Thompson and Estes then tested whether the size of an object (small, medium-sized, and large) could linearly predict the number of "small-sounding" and "large-sounding" phonemes in their names. Their name stimuli include sets of five consonant-vowel-consonant-vowel-consonant-vowel (CVCVCV) strings such as "wodolo", where all the six phonemes are "large-sounding", "tibudo" with four "large-sounding" phonemes ("b, u , d, o"), "kuloti" with three "large" phonemes ("u, l, o"), "bitiku" with two "large" phonemes ("b, u"), and "kitete" with no "large" phonemes. The participants had to match the five sounds to three computer-generated novel objects ("greebles") of three different sizes. The results showed that as the size of the greebles linearly increases, so too is the number of the large-sounding phonemes in their preferred names, thereby supporting a continuant neurological explanation instead of a contrastive statistical account of size symbolism.

Given the review of the literature summarized here, it can be argued that the continuous-categorical perception method appears to constitute a promising means of evaluating two different explanations of sound symbolism, and we also tested participants' continuous perception of sound-size and sound-shape mappings. The next section outlines the research questions of the current study.

1.3 Goals of the present study

The present study investigates both the "what" and the "why" questions, with a focus on the two most robustly established sound symbolic effects: sound-size symbolism effect and sound-shape symbolism effect. Regarding the "what" question, the present study was first designed to test whether the consonants or the vowels are more salient in terms of driving size and shape symbolism. The first hypothesis (Hypothesis 1), contrary to the claim put forward by Nielson and Rendall (2011), is that vowels are more salient because: (1) vowels constitute the nucleus (or the peak) of a syllable that can bear stress and pitch[5]; (2) children have been found to assimilate the PLACE feature (see Table 2) of the consonants to that of the vowels, for instance, in child initial speech, "pick /p-i-k/" with labial consonant "p", coronal vowel "i" and dorsal consonant "k" are usually pronounced as "tit /t-i-t/", where all the consonants are changed to the coronal consonant "t", following the PLACE feature of the vowel "i" (see Fikkert & Levelt, 2008); (3) vowels are, in general, more sonorous than consonants (sonority can be viewed as the "intrinsic loudness" of individual phonemes, see below for a detailed discussion);

[5] A syllable (e.g., "mal") generally consists of an optional onset (consonant "m"), an obligatory nucleus (vowel "a") and an optional coda (consonant "l"). The sonority of the syllable peaks at the nucleus, which could bear stress (as in the disyllable word "`mal.let") and pitch, tone, etc. Consonants cannot bear stress, pitch, tone, etc. as they are usually the onsets or codas of syllables (see Gussenhoven & Jacobs, 2005, for an introduction).

Second, the present study aims to provide a better phonological classification of the "large"/"round" sounding phonemes and the "small"/"angular" phonemes. The second hypothesis (Hypothesis 2) is that sonority scale linearly predicts size symbolism. In particular, phonemes can be ranked along a sonority scale, which follows their acoustic intensity relative to the energy involved in producing the sound, namely a sound's *"loudness relative to that of other sounds with the same length, stress and pitch"* (Ladefoged, 1993, p. 245). Despite some minor variations, the generally acknowledged (e.g., Gussenhoven & Jacobs, 2005) sonority scale is:

Vowels: Low Vowels > Mid Vowels > High Vowels

Consonants: Glide > Liquids > Nasals > Voiced Fricatives > Voiced Stops > Voiceless Fricatives > Voiceless Stops (see Parker, 2002).

Thus, for English, the ranking is as follows:

Vowels: [a] > [o e] > [u i]

Consonants: [w y] > [r l] > [m n] > [z v ð ʒ] > [b d g] > [s f θ ʃ] > [p t k]

We could consider "sonority" as the "intrinsic loudness" of the phonemes. Since the suprasegmental "loudness" (regardless of the phonemes) and frequency of sounds (also relates to loudness) has been generally reported to suggest symbolic meaning (e.g., Bond & Stevens, 1969; Marks, 1987; Ohala, 1994; Tsur, 2006), it is naturally assumed that the "intrinsic loudness" of different phonemes would also be associated with figure size. According to Hypothesis 2, then, more sonorous sounds should correspond to larger-sized objects, and this effect should be continuous instead of contrastive. The question, however, as to whether the differing sonorities can be perceived by ear. Evidence from phonology and phonetics suggests that they can: (1) monosyllabic words such as "snug" are well formed since the sonority slope rises progressively from "s" to "n", and peaks at the nucleus "ʌ" before dropping to the coda “g” (see Footnote 5, for the notion of onset, nucleus and coda), but words like “nsug” are less common as the onset cluster “ns” has a decreasing sonority. This principle (Sonority Sequencing Principle, SSP) was found in most languages, thus sonority must be perceptible (Parker, 2002); (2) in English reduplicative rhyming words (e.g., “roly-poly, willy-nilly, loosey-goosey”), the onsets of the first morpheme is generally more sonorous than the second morpheme; (3) in sound perception tests (participants heard a sound (e.g., CV) and indicated which sound it was on the screen), participants were more easily confused between phonemes with same sonority (see Komatsu et al., 2002, for a multi-dimensional analysis of the confusion matrix).

We further test the sonority hypothesis in the case of shape-symbolism, namely, whether more sonorous sounds are perceived as more rounded in shape? Reports from the sound symbolism literature suggest that “large sounding” phonemes also “sound round”, and “small sounding” phonemes also “sound angular” (see Table 3), but no previous studies have, as far as we are aware, investigated the symbolic effects of the “medium sounding” phonemes. The current study fills the gap by examining whether the “medium-sounding” phonemes would also predict “medium” size and “medium” rotundity/angularity.

The results of the study may also have implications for the “why question” concerning sound symbolism. Similar to Thompson and Estes (2011), we tested whether sound symbolism is continuous or contrastive, namely, whether it follows the sonority scale continuously such that the most sonorous sounds are matched with the largest and most rounded shapes, medium sonorous sounds with medium large and rounded shapes, and least sonorous sounds with the smallest and least rounded shapes. If the hypothesis was confirmed, this could be interpreted as evidence against the contrastive account in which only two extreme sensory features are marked. However, it is unknown whether size and shape symbolism would both follow the sonority scale continuously. As noted by Stevens (1957), there are two kinds of perceptual continua, one being “prothetic (quantitative)” and the other being “metathetic (qualitative)”. Size belongs to the prothetic dimension where “more” and “less” can be measured quantitatively. Shape, on the contrary, is often regarded as the methathetic dimension where “round” is qualitatively different from “angular”, but not “more” than the other. Thus we may find continuous marking for size symbolism but contrastive marking for shape symbolism.

To further investigate the relationship between size and shape symbolism, a third experiment was conducted in order to determine whether rounded objects are perceived as larger in size. If the unimodal correspondence is observed, then it is also possible that shape symbolism is mediated by size symbolism.

2. Experiment 1—Size symbolism

2.1 Methods

Experiment 1 investigated sound-size symbolism. The study comprised two tasks; Task 1 was designed to address Hypothesis 1 and Task 2 addressed Hypothesis 2.

Participants: 30 Oxford University students participated in the study (19 female, mean age = 22.3 years), all with normal or corrected-to-normal vision. 7 of the participants were non-native English speakers, and the rest were native English speakers. The whole experiment lasted for 30-40 minutes, and was approved by the Oxford University ethics review. Participants gave their informed consent before taking part in the experiment, and were reimbursed £10 for attending the study.

Stimuli: The visual stimuli for Task 1 consisted of pictures of five pairs of Russian dolls differing in the pattern on their bodies. The pattern was the same for each pair; the small and larger dolls in the pair subtended 2.45° (H) x 4.53° (V) and 4.90° (H) x 9.05° (V) of visual angle at a distance of 40 cm (see Figure 4). The Russian dolls were chosen as the visual stimuli as they inherently convey the meaning of differing sizes (the smaller dolls can be put into the larger dolls). As a consequence, there is no need for an extra reference object as used by Thompson & Este (2011) to indicate that object size varies (rather than distance). For Task 2, the same five sets of dolls were used, each with 5 varying sizes to form a 5-point scale (see Figure 5 for an example). The heights of the 5 dolls subtended 2.45° (H) x 4.53° (V), 3.09° (H) x 5.66° (V), 3.69° (H) x 6.79° (V), 4.30° (H) x 7.92° (V) and 4.90° (H) x 9.05° (V) of visual angle at a distance of 40 cm. There were 5 scales in total (see APPENDIX for a full list of the visual stimuli used in Experiments 1, 2 and 3).

Figure 4. The visual stimuli used in Experiment 1, Task 1.

Figure 5. The visual stimuli used in Experiment 1, Task 2.

The auditory stimuli used in Task 1 included 24 non-words, each consisting of one consonant and one vowel (CV string; see Table 4). The 6 "large-sounding" consonants chosen were "m, b, w, r, l, n", as they are consistently reported to be linked to large size in past research (e.g., Berlin, 2006; Maurer et al., 2006; Newman, 1933; Thompson & Estes, 2011). Likewise, the 6 "small-sounding" consonants chosen were "s, f, θ, ʃ, t, k"; the "large-sounding" vowels were "o, u", and the "small-sounding" vowels were "i, e". The sounds were recorded as 16-bit stereo files at a sampling frequency of 44,100 Hz (CD quality) by a female phonologist, who is a native English speaker. The duration of each sound was adjusted to 1 second and the average intensity was scaled to 70 decibels (dB(A)) by the acoustic analysis software Praat[6].

Table 4. The auditory stimuli used in Task 1.

"Large" C	**mi**	**bi**	**wi**	**ri**	**li**	**ni**
"Small" V	me	be	we	re	le	ne
"Small" C	so	fo	θo	ʃo	to	ko
"Large" V	su	fu	θu	ʃu	tu	ku

The auditory stimuli used in Task 2 included 100 CV non-words (20 consonants x 5 vowels; see Table 5). Using one-syllable CV strings rather than multiple-syllable CVCVs ("bouba/kiki") or CVCVCVs ("maluma/takete") as stimuli eliminates the potential confounding influence of stress (for example, [`buba] and [bu`ba] may elicit rather different sound symbolic effects). CVCs ("mal/mil") with one syllable were not used either because (1) we were testing the relative salience of vowels and consonants, so the number of consonants and vowels in the sound sequence needed to be equal; (2) CVC strings are the most common word formation pattern in English (Charles-Luce & Luce, 1990), whereas CV words are comparatively rare (examples are "key, saw, go", etc.), thus the chance that any of our stimuli resemble real words is relatively low as compared to CVC words; (3) CVC stimuli may induce phonological assimilation, that is, the phonological features of the final consonant (coda) may influence the features of the preceding vowel. For example, nasal codas "m, n" will nasalize the preceding vowel "ɑ" into "ɑ̃". Thus a CVC stimulus is not suitable in the context of the present study.

[6] Note that this intensity was the acoustic features of the sound files. The actual sound volume of the auditory stimuli through the loudspeakers was measured using a decibel meter.

Table 5. The auditory stimuli utilized in Task 2, Experiment 1 and 2.

wa	ya	la	ra	ma	na	za	va	ʒa	ða	ba	da	ga	sa	fa	ʃa	θa	pa	ta	ka
wu	yu	lu	ru	mu	nu	zu	vu	ʒu	ðu	bu	du	gu	su	fu	ʃu	θu	pu	tu	ku
wo	yo	lo	ro	mo	no	zo	vo	ʒo	ðo	bo	do	go	so	fo	ʃo	θo	po	to	ko
we	ye	le	re	me	ne	ze	ve	ʒe	ðe	be	de	ge	se	fe	ʃe	θe	pe	te	ke
wi	yi	li	ri	mi	ni	zi	vi	ʒi	ði	bi	di	gi	si	fi	ʃi	θi	pi	ti	ki

Procedure: The participants were seated in front of 17-inch screen in a small and silent testing room. The centre of the screen was positioned at the participants' eye-level at a distance of 40 cm. The average volume of all sounds was in the comfortable listening range of 60 to 70 dB(A).

In Task 1, the participants heard a CV non-word from the loudspeaker. At the same time, the participants saw two dolls of differing sizes aligned horizontally on the screen. The participants then pressed the "z" key on the keyboard if they thought the sound matched the left image, or the "m" key if they thought the sound matched the right image. The participants were instructed to respond as rapidly and accurately as possible. Each of the 24 CV sounds was presented 3 times, giving rise to a total of 72 trials. The presentation order of the auditory and visual stimuli was completely randomized across trials, so was the position of the small and big sized dolls.

In Task 2, the participants heard a CV sound and saw a scale on the screen. They then had to indicate, by pressing a number from one to five on the keyboard, which size of the dolls they thought matched the sound. Each sound was presented 3 times, giving rise to a total of 300 trials. The presentation of the auditory stimuli was completely randomized, but the orientation of the scales was preserved (small to large from left to right) across trials. The experimental room remained semi-darkened and silent throughout the course of the experiment. The experimenter remained out of the participant's line of sight during the task. The participants were debriefed after the experiment, and were asked whether they were aware of sound symbolic effects; none of them reported previous knowledge of sound symbolism.

2.2 Results

2.2.1 Task 1

The participants' responses in Task 1 were scored individually using both a vowel and a consonant scheme. In the vowel scheme, a correct score was defined as matching the "large sounding" vowel ("o, u") with the large image, or matching the "small sounding" vowel ("i, e") with the small image. Put differently, the vowel scheme assumes that the non-word's vowel content correctly determines the participant's response, while the consonant content is neglected. In the consonant scheme, a correct score was given for responses matching the "large sounding" consonant ("m, b, w, r, l, n") with the large image, or matching the "small sounding" consonant ("s, f, θ, ʃ, t, k") with the small image. In other words, the consonant scheme assumes that the consonant content determines the participant's response, while the vowel content is ignored. Participants' correct matching scores were averaged across trials, and individual participants' averages were tested for their deviation from chance (50% correct) using one-sample *t*-tests (two-tailed). The results revealed that for size symbolism, the percentage of correct scores according to the vowel scheme was significantly higher than chance at 58.2% ($t(29) = 2.94$, $p = .006$). The percentage of correct scores according to the consonant scheme was 41.8% ($t(29) = -2.94$, $p = .006$; see Figure 6). Analyses on the level of individual vowels revealed that the correct percentages for "i, u" were significantly above chance under vowel scheme ("i": $t(29) = 3.01$, $p = .005$; "u": $t(29) = 3.63$, $p = .001$; see Figure 7); similarly, analyses on the level of individual consonants showed that, the correct percentages for "n, k, t, ʃ, θ, f" were significantly above chance under vowel scheme ("n":

$t(29) = 2.63$, $p = .013$; "k": $t(29) = 2.14$, $p = .041$; "t": $t(29) = 2.07$, $p = .047$; "ʃ": $t(29) = 2.17$, $p = .038$; "θ": $t(29) = 3.4$, $p = .002$; "f": $t(29) = 2.94$, $p = .006$, respectively; see Figure 8).

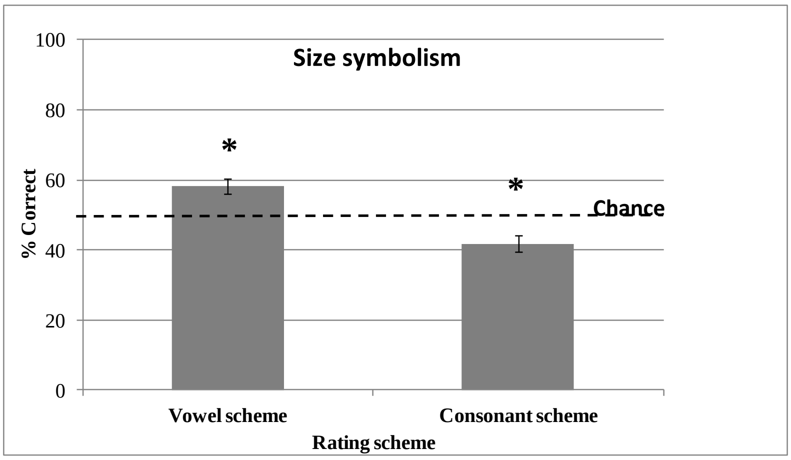


Figure 6. Experiment 1, Task 1: Mean (±SEM) correct size scores by the vowel and the consonant schemes. Under vowel scheme, correct scores were assigned for matching large (small)-sounding vowels with large (small) objects; under consonant scheme, correct scores were assigned for matching large (small)-sounding consonants with large (small) objects. Effects shown were compared to chance level (50%).

*: $p < .05$.

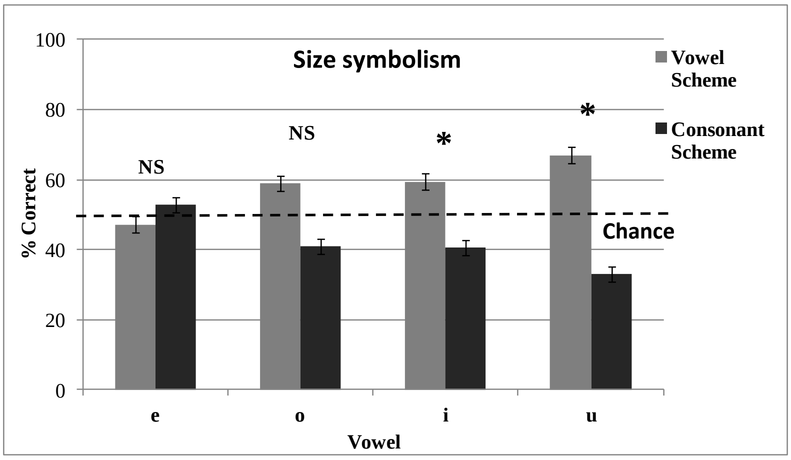


Figure 7. Experiment 1, Task 1: Mean (±SEM) correct size scores under vowel and consonant schemes on the level of individual vowels. Effects shown were compared to chance level (50%).

*: $p < .05$; NS: non-significant.

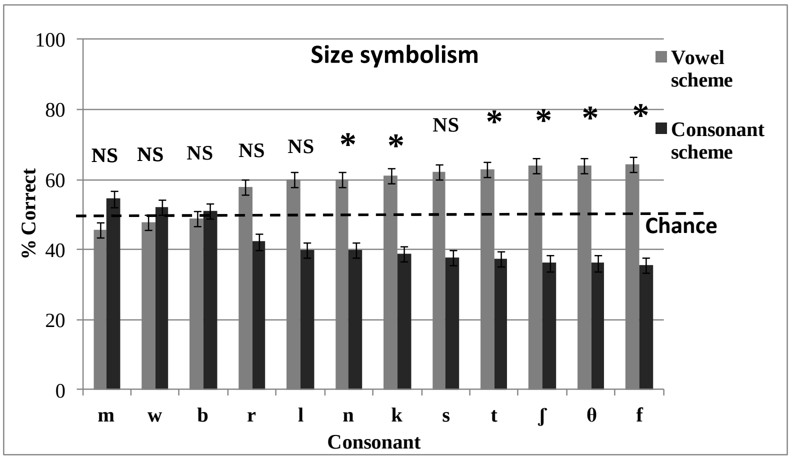


Figure 8. Experiment 1, Task 1: Mean (+SEM) correct size scores under vowel and consonant schemes on the level of individual consonants. Effects shown were compared to chance level (50%). *: $p < .05$; NS: non-significant.

In general, the results support Hypothesis 1 that the vowels are more salient in terms of driving the size symbolism effect when the consonant and vowel contents of the sounds are incongruent in their size symbolic meanings. Specifically, the vowels “i, u” emerged as the strongest drivers of this effect, such that any “Ci (consonant + vowel “i”, e.g. “wi, fi, ki”)” sound tended to be matched to small shapes, and any “Cu” sound was likely to be matched to large shapes. The results also suggest that the vowel “e” is the weakest vowel in terms of predicting size symbolism, such that a “Ce” sound could be either small or large (mid-sized). On the level of individual consonants, the results suggest that a CV sound beginning with “n, k, t, ʃ, θ, f” tends to follow their vowel contents in their size symbolic meanings.

The results contradict Nielsen and Rendall’s (2011) claim that consonants are the dominant driver of sound symbolism. However, their experiments focused on sound-shape symbolism, which may account for the discrepancy in the results observed. To see whether Hypothesis 1 (vowels are more salient in driving sound symbolic effects) still holds for shape symbolism, we conducted an identical experiment with round/angular shapes as the visual stimuli. The results are presented in Section 3.2.1.

2.2.2 Task 2

The average size score for each CV sound was calculated for each participant, with larger values indicating larger size. The 20 consonants were grouped into 7 sonority groups: Glide ("w, y"), Liquid ("l, r"), Nasal ("m, n"), Voiced Fricative ("z, v, ð, ʒ"), Voiced Stop ("b, d, g"), Voiceless Fricative ("s, f, θ, ʃ") and Voiceless Stop ("p, t, k"). According to the established sonority scale (e.g., Parker, 2002), [a] > [o e] > [u i], and Glides > Liquids > Nasals > Voiced Fricatives > Voiced Stops > Voiceless Fricatives > Voiceless Stops in terms of sonority. We therefore expected that the perceived sizes of the vowels and consonants would also follow the scale, namely, [a] > [o e] > [u i], and Glides > Liquids > Nasals >Voiced Fricatives > Voiced Stops > Voiceless Fricatives > Voiceless Stops in terms of the matched visual size. A 5 (vowel type) x 7 (consonant group) repeated measures analysis of variance (ANOVA) was conducted to test the main and interactive effects of consonant sonority group and vowel type on size scale responses. The analysis revealed significant main effects of consonant sonority group ($F(2.26, 65.56) = 26.1$, $p < .001$, $\eta^2_{\text{partial}} = .47$) and of vowel type ($F(2.44, 70.84) = 49.23$, $p < .001$, $\eta^2_{\text{partial}} = .63$). The interactive effect of vowel type and consonant group on size scores was also significant ($F(10.53, 305.42) = 2.05$, $p = .002$, $\eta^2_{\text{partial}} = .07$), suggesting that the effect of vowel type varied across different consonant groups. More specifically, more sonorous consonants were perceived as much larger when followed by vowels "a, o, u" than when followed by "e, i", while less sonorous consonants did not differ greatly when followed by vowels "a, o, u" or by "e, i". To instantiate, "wa" sounded much larger than "wi", but "ka" and "ki" were less different in size (see Figure 9).

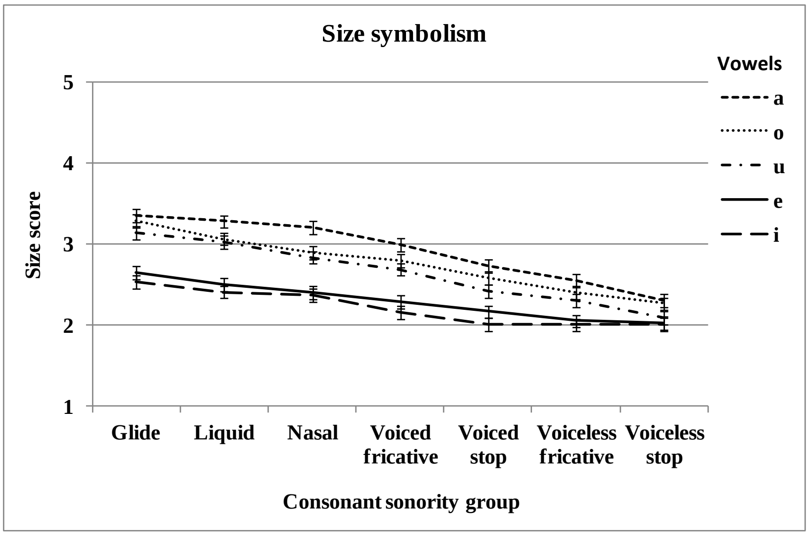


Figure 9. Experiment 1, Task 2: Mean (+SEM) size scores of the consonant groups following the sonority scale.

It can be seen that for all the vowels, the size scores linearly declined along the consonant sonority scale, namely, the scores decreased from the most sonorous glides to the least sonorous voiceless fricatives. The size scores for the vowels ranked as [a] > [o] > [u] > [e] > [i], not following the sonority hierarchy of the vowels. Instead, it appeared that “a, o, u” grouped together and “e, i” grouped together.

The test for a linear trend in the size scores along consonant sonority revealed a significant linearity ($F(1, 29) = 45.16$, $p < .001$, $\eta^2_{\text{partial}} = .61$), suggesting that size effect was continuous along the sonority scale. No linearity was found for the vowel ranking ($F(1, 29) = 0.515$, $p = .479$, $\eta^2_{\text{partial}} = .017$) or the consonant group and vowel interaction ($F(1, 29) = 1.58$, $p = .219$, $\eta^2_{\text{partial}} = .052$). Post-hoc pairwise comparisons also revealed significant differences between the four least sonorous consonant groups (voiced fricative, voiced stop, voiceless fricative, voiceless stop) at the $p < .05$ level, but the differences between the four most sonorous consonant groups (glide, liquid, nasal, voiced fricative) failed to reach statistical significance. For the vowel comparisons, except for “e, i” ($p = .197$), all were significantly different from each other.

3. Experiment 2—Shape symbolism

3.1 Methods

Experiment 2 investigated sound-size symbolism. There were also two tasks, Task 1 was designed to address Hypothesis 1 and Task 2 addressed Hypothesis 2.

Participants: 30 university students (21 female, mean age = 21.4 years) took part in the study, all with normal or corrected-to-normal vision. Six of the participants were non-native English speakers. The whole experiment lasted for 30-40 minutes, and was approved by the Oxford University ethics review. Participants gave their informed consent before taking part in the experiment, and were reimbursed £10 for taking part in the study.

Stimuli: The auditory stimuli for Task 1 and 2 of Experiment 2 were identical with those used in Task 1 and 2 of Experiment 1 (see Tables 1 and 2). The visual stimuli for Task 1 included 5 pairs of round and angular shapes (see Figure 10 for an example). The shapes were all black and white, generated by the online pattern generator "subble" (http://www.subblue.com/projects; downloaded on 4, May, 2012), and all fit into a square subtended 4.26° (H) x 4.27° (V) of visual angle[7].

Figure 10. The visual stimuli used in Experiment 1, Task 1.

For Task 2, the same 5 pairs of shapes were used, only that they were shown at the ends of a 5-point scale (see Figure 11 for an example). The 5 pairs formed 25 scales in total.

[7] Admittedly, there is a confound of shape and size, since although the side lengths of the shapes are the same, the areas covered by the shapes are different. In particular, round shapes were generally larger in areas than angular shapes. This confound, however, exists in most previous studies on shape symbolism (e.g., Köhler, 1929; Maurer et al., 2006; Ramachandran & Hubbard, 2001, 2003; Westbury, 2005).

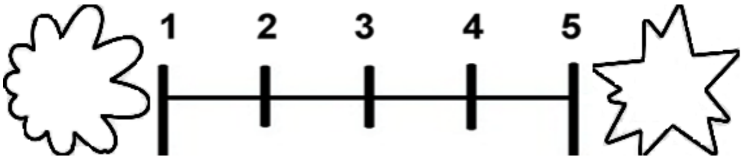


Figure 11. The visual stimuli used in Task 2, Experiment 2.

Procedure: In Task 1, the participants heard a CV non-word and saw two shapes aligned horizontally on the screen, one round and one angular. They then pressed “z” on the keyboard if they thought the sound matched the left image, or “m” if they preferred the right image. Each of the 24 CV sounds was presented 3 times, giving 72 trials in total. The presentation of the auditory and visual stimuli was randomized across trials, as was the position of the round and angular shapes.

In Task 2, the participants heard a CV sound and saw a scale on the screen. They then had to indicate, by pressing a number from one to five on the keyboard, how much they thought that the sound resembled the two shapes. Each sound was presented 3 times, giving rise to a total of 300 trials. The presentation of the stimuli was randomized, as was the orientation of the scales.

3.2 Results

3.2.1 Task 1

The participants’ responses in Task 1 were scored individually using both a vowel and a consonant scheme. In the vowel scheme, a correct score was defined as matching the “round sounding” vowel (“o, u”) with the round image, or matching the “angular sounding” vowel (“i, e”) with the angular image. In the consonant scheme, a correct score was given for responses matching the “round sounding” consonant (“m, b, w, r, l, n”) with the round image, or matching the “angular sounding” consonant (“s, f, θ, ʃ, t, k”) with the angular image. Participants’ correct matching scores were averaged across trials, and individual participants’ averages were tested for their deviation from chance level (50% correct) using one-sample *t*-tests (two-tailed). The results showed that for shape-symbolism, both the percentages of correct scores based on the vowel and the consonant scheme were not significantly different from chance, 50% ($t(29) = .147$, $p = .88$; see Figure 12). It would therefore appear that participants simply ventured a guess when the consonant and the vowel contents were incongruent. However, further scrutiny of the data revealed that this was not the case. Figures 13 and 14 show the detailed percentages of correct scores for each vowel and consonant under both schemes. A clear difference can be seen between the "round " and "angular" sounding phonemes. From the perspective of the vowels, a word containing a "round sounding" vowel ("o, u") was perceived as round for more than 60% of the time ("o":$t(29) = 5.17$, $p < .001$; "u": $t(29) = 3.12$, $p = .004$); a word containing an "angular sounding" vowel ("e, i"), was, however, perceived as angular less than 40% of the time ("e": $t(29) = -4.95$, $p < .001$; "i": $t(29) = -2.44$, $p = .021$). From the perspective of the consonants, all of the "round sounding" consonants ("w, l, r, b, n, m") were associated with round shapes 60% or more of the time (except for "b": $t(29) = -1.94$, $p = .062$, and "n": $t(29) = -1.53$, $p = .136$, all significant at $p = .05$: "w": $t(29)= -5.72$, $p < .001$; "l": $t(29) = -3.25$, $p = .003$; "r": $t(29) = -2.36$, $p = .025$; "m": $t(29) = -2.04$, $p = .05$); all the "angular sounding" consonants ("k, t, θ, ʃ, s, f"), again, predicted angularity less than 40% of the time (except for "k": $t(29) = 0.56$, $p = .582$, all significant at $p = .05$: "t": $t(29) = 2.28$, $p = .03$; "θ": $t(29) = 3.3$, $p = .003$; "ʃ": $t(29) = 2.42$, $p = .022$; "s": $t(29) = 3.57$, $p = .001$; "f": $t(29) = 3.66$, $p = .001$).

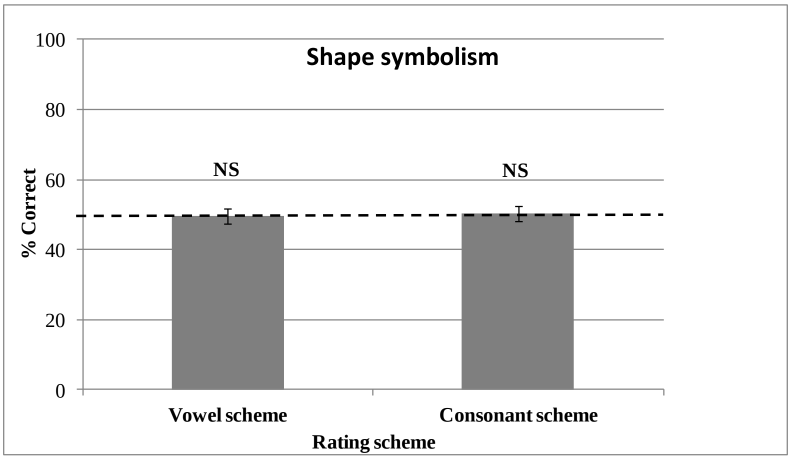


Figure 12. Experiment 2, Task 1: Mean (±SEM) correct shape scores by the vowel and the consonant schemes. Under vowel scheme, correct scores were assigned for matching large (small)-sounding vowels with large (small) objects; under consonant scheme, correct scores were assigned for matching large (small)-sounding consonants with large (small) objects. Effects shown were compared to chance level (50%).

NS: non-significant.

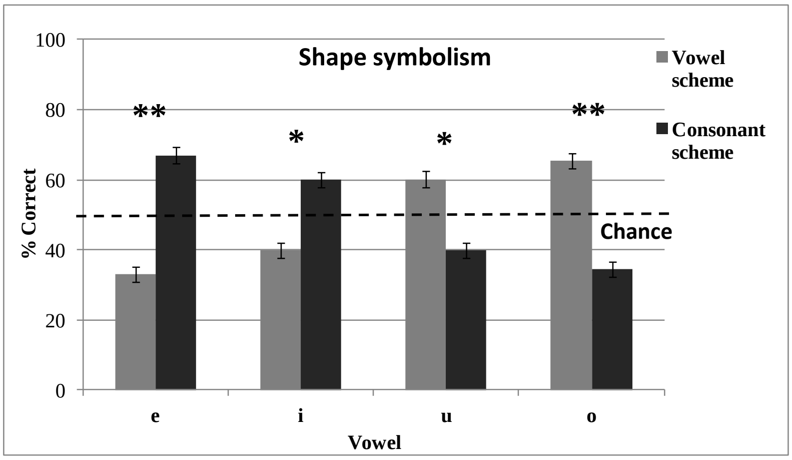


Figure 13. Experiment 2, Task 1: Mean (±SEM) correct shape scores by the vowel and the consonant schemes on the level of individual vowels. Effects shown were compared to chance level (50%).

**: $p < .001$; *: $p < .05$.

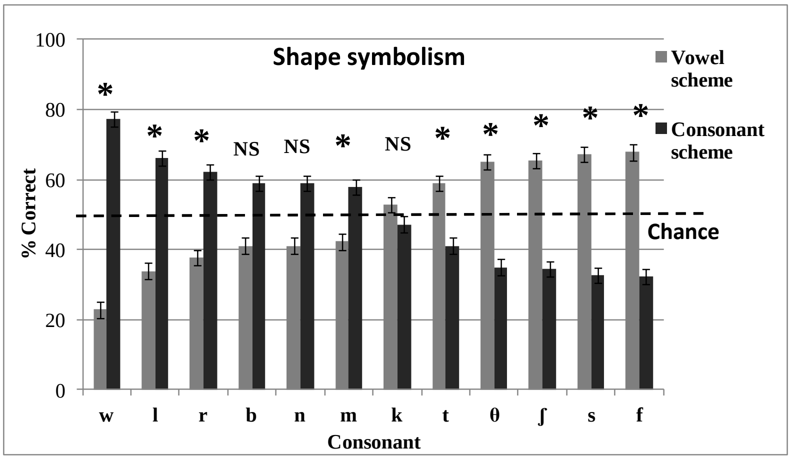


Figure 14. Experiment 2, Task 1: Mean (+SEM) correct size scores under vowel and consonant schemes on the level of individual consonants. Effects shown were compared to chance level (50%). *: $p < .05$; NS: non-significant.

In short, vowels and consonants appear to be equally salient in terms of driving shape symbolism, but the "round sounding" phonemes, be they vowels or consonants, are more salient than "angular sounding" phonemes. Again, contrary to Nielsen and Rendall's (2011) conclusion, we did not find a dominant effect of consonants for shape symbolism; but Hypothesis 1, that vowels are more salient, was not supported either. The results appear to suggest that an incongruent CV sound containing both "large sounding" and "small sounding" phonemes can be either large or small, depending on the vowel content; but that a CV sound containing both "round sounding" and "angular sounding" phonemes is more likely to be rounded. This discrepancy between size and shape symbolism is further illustrated in Figure 15, where the percentage of actual responses for each of the sounds were shown. Of all the 24 incongruent CV sounds (see Table 4), 54.9% were matched to the large size and 45.1% were matched to the small size (significantly different from chance, $t(719) = \pm 3.52$, $p < .001$); 63.2% were matched to the round shape and 36.8% were matched to the angular shape (significantly different from chance, $t(719) = 10.79$, $p < .001$). Thus the difference between the choices for sizes (9.7%) is much smaller than that for shapes (26.4%; see Figure 15). The implications of the discrepancy between size and shape symbolism are discussed in Section 4.

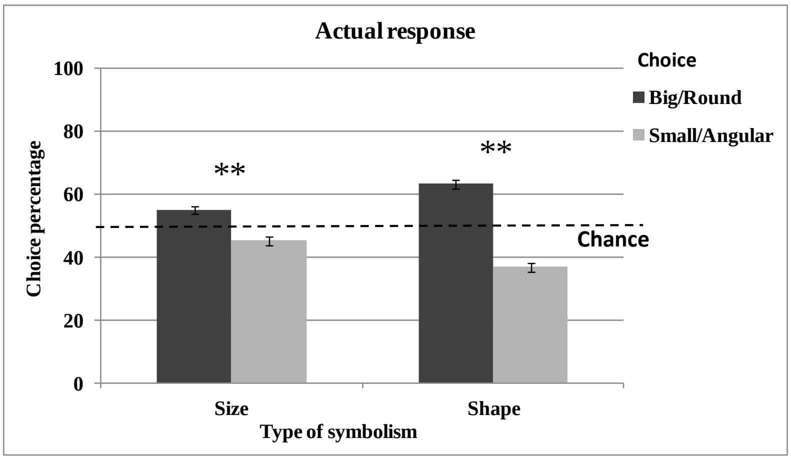


Figure 15. Experiment 1 and 2, Task 1: Mean (+SEM) percentage of actual responses to shape and size for the incongruent CV sounds. Effects shown were compared to chance level (50%).
**: $p < .001$.

3.2.2 Task 2

As for the analysis of Task 2 in Experiment 2, the average roundness score for each CV sound was calculated for each participant, with larger values indicating more rounded shapes. The roundness scores were aggregated across participants (see Figure 16). It can be seen that although there was a general decline of the roundness scores along the consonant sonority scale, the scores from voiced fricatives to voiceless fricatives increased—this pattern remained true for all 5 vowels. The vowel ranking for roundness score was [o] > [u] > [a] > [e] > [i], this pattern clearly followed the [+roundness] distinction rather than vowel sonority.

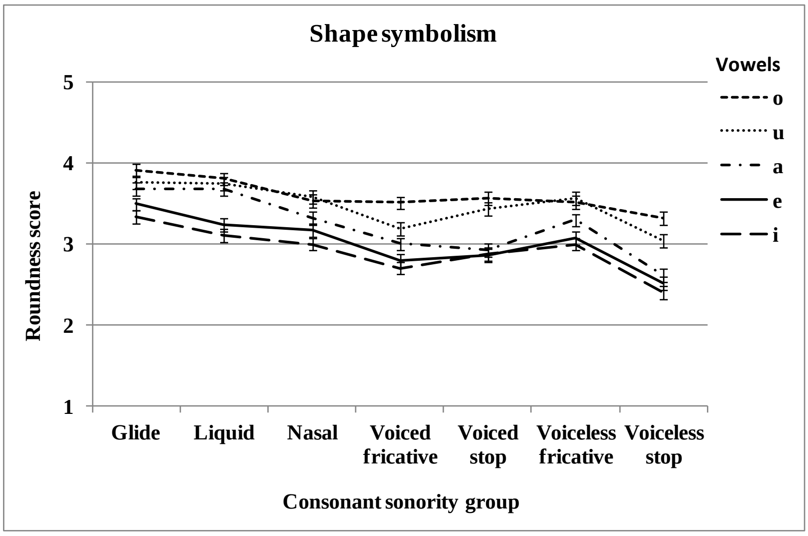


Figure 16. Experiment 2, Task 2: Mean (±SEM) shape scores of the consonant groups following the sonority scale.

The results were entered into a 5 (vowel type) x 7 (consonant group) repeated measures ANOVA. Main effects were found for consonant group ($F(3.03, 87.82) = 26.72$, $p < .001$, $\eta^2_{partial} = .48$), for vowel type ($F(2.56, 74.14) = 64.8$, $p < .001$, $\eta^2_{partial} = .691$), and for the interaction between the consonant and vowel factors ($F(9.91, 287.51) = 2.21$, $p < .018$, $\eta^2_{partial} = .071$).

Post-hoc pairwise comparisons revealed significant differences only between the consonant groups "nasal" and "voiced fricative" ($p = .022$), and between voiceless fricative and voiceless stop ($p < .001$); no significant differences were found for other adjacent consonant groups, indicating a weak linear effect. For the vowel comparisons, the difference between "e, i" again failed to reach statistical significance; other vowel pairs were all significantly different from each other at the $p < .05$ level.

The Pearson's correlation coefficient revealed a strong and significant correlation between the average size score and the average roundness score for the 100 sounds ($r(98) = .61$, $p < .001$). Experiment 3 was conducted to further see whether this correlation might be mediated by the unimodal correspondence between size and shape in vision.

4. Experiment 3—Size-shape symbolism

4.1 Methods

Participants: 20 of the participants (14 female, mean age = 23.4 years) who participated in Experiment 2 also participated in Experiment 3. Six of the participants were non-native English speakers and 14 of them were native. The whole experiment lasted for approximately 10 minutes, and was approved by the Oxford University ethics review.

Stimuli: The visual stimuli consisted of the same 5 size scales used in Task 2, Experiment 2, and 5 pairs of round-angular shapes used in Task 1, Experiment 1. But the shapes all had 5 different sizes (see Figure 17). The sizes from the smallest to the largest subtended 4.52° (H) x 4.53° (V), 5.65° (H) x 5.66° (V), 6.78° (H) x 6.79° (V), 7.91° (H) x 7.92° (V) and 9.04° (H) x 9.05° (V) of visual angle at a distance of 40 cm. There were no auditory stimuli.

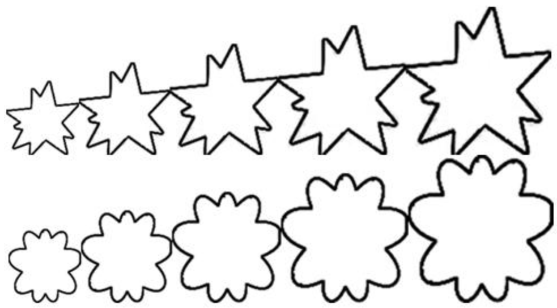


Figure 17. The visual stimuli used in Experiment 3.

Procedure: The participants were presented with a shape of one size on the screen for 200 ms, followed by a size scale. They then indicted on the scale which of the dolls had the same height with the image they just saw (see Figure 18). Since each of the 10 shapes has 5 different sizes, there were 50 images in total, and each image was tested for 3 times, giving rise to a total of 150 trials.

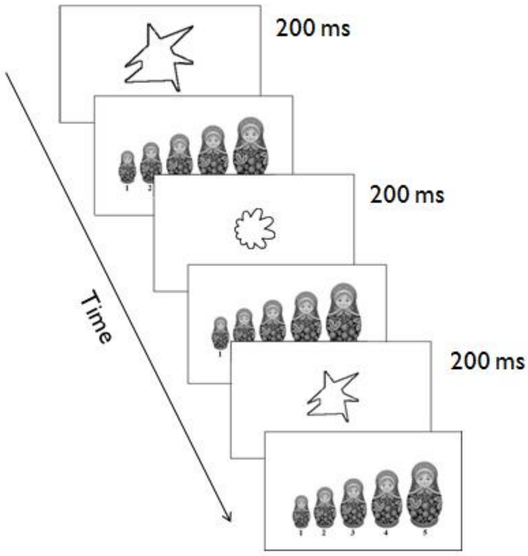


Figure 18. The procedure used in Experiment 3.

4.2. Results

The percentage of overall correct responses was 46.6% (not significantly different from chance, $t(19) = -1.58$, $p = .129$); the correct percentage for round shapes was 44.5% (significantly lower than chance, $t(19) = -2.45$, $p = .024$); the correct percentage for angular shapes was 48.7%, (not significantly different from chance, $t(19) = -0.49$, $p = .63$; see Figure 19). The round and angular shapes displayed different error patterns. For the round shapes, 23.41% of the errors were made by rating them as smaller than their actual sizes, 76.59% were made by rating them as larger than the actual sizes; for the angular shapes, however, 48.11% of the errors were "smaller rating", and 51.89% were "lager rating". An independent-samples *t*-test revealed a significant difference between round and angular shapes in error types ($t(38) = \pm 3.3$, $p = .002$; see Figure 20). The results suggest that round shapes were more likely to be perceived as larger than their actual sizes whereas angular shapes tended be perceived as smaller than their actual sizes.

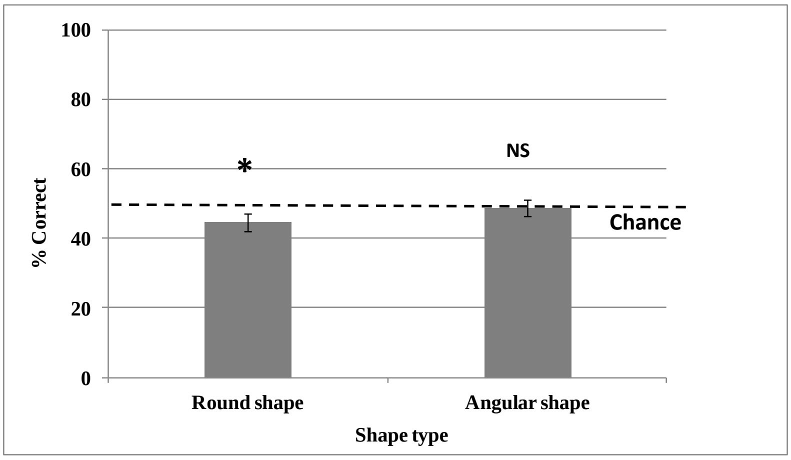


Figure 19. Experiment 3: Mean (±SEM) percentage of correct responses matching the round and angular shapes to their actual sizes. Effects shown were compared to chance, 50%.

*: p < .05; NS: non-significant.

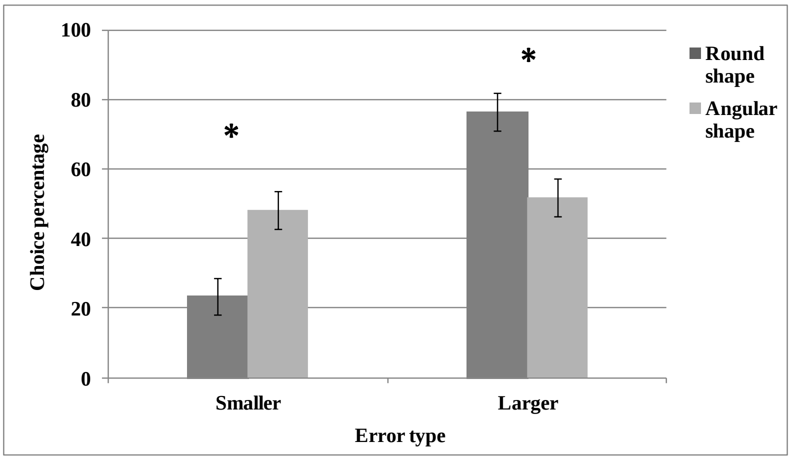


Figure 20. Experiment 3: Mean (±SEM) percentage of correct responses matching the round and angular shapes to their actual sizes. Effects shown were comparisons between round and angular shapes.

*: p < .05.

5. Discussion

The results from both Task 1 and Task 2 of Experiments 1 and 2 reveal different sound symbolic patterns for size and shape symbolism. Task 1 results suggest that for size symbolism, vowels are more salient when the vowel and consonant contents are incongruent in their symbolic meanings. But for shape symbolism, it is the “round sounding” phonemes that count; no matter whether they are vowels or consonants. This discrepancy may be explained by different brain cross-activation underlying different types of sound symbolism. As discussed before, vowels are generally more sonorant than consonants, thus if vowels are what drives any size symbolism effect, it may well be due to the fact that they are intrinsically louder than consonants. This suggests that sonority-based effects are due to an acoustic-based mapping between the visual and auditory cortex. On the contrary, vowels and consonants could be equally important in shape symbolism, but the “round sounding” phonemes play a dominant role. We further examined the so-called “round sounding” phonemes in our stimuli (see Table 4), and found that vowels “o, u” are the [+round] phonemes, and consonants “m, b w, r” can also be [+round] as they are either bilabial or involve protrusion of the lips (Gussenhoven & Jacobs, 2011). Thus, 6 out of 8 “round sounding” phonemes in the set of stimuli utilized here actually have the [+round] phonological feature, which is in effect an articulatory feature. This suggests a motor-based mapping between the mouth and hand areas for shape symbolism.

As argued by Ramachadran and Hubbard (2011), sound symbolism might be caused by direct cross-activation between IT and other visual centres and the size of the auditory cortex, or by motor to motor mappings (synkinesia) between hand gestures and tongue, lip and mouth movements in the Penfield motor homunculus. It is then suggested that neuroimaging studies ought to be conducted in order to determine whether size and shape symbolism elicit different patterns of brain activation. Peiffer-Smadja’s (2010) fMRI study makes a start in this direction: he analysed brain activation during presentation of auditory stimuli, visual stimuli and auditory-visual stimuli, and observed that visual and auditory stimuli activated different premotor areas that could correspond to hand and mouth motor cortex respectively. Unfortunately, he only tested shape symbolism, thus it is unknown whether size symbolism displays the same motor-to-motor activation.

The difference between size and shape symbolism is also observed in the results of Task 2. Size symbolism demonstrated a significant linear decrease of size along the sonority scale of consonants (from the most to the least sonorous), while shape symbolism displayed a strong contrast of roundness between the most sonorous sounds and the least sonorous sounds, but a less clear relationship was observed between the medium rounded shapes and medium sonority. Specifically, the less sonorous voiceless fricatives "s, f, θ, ʃ" were significantly more rounded than the more sonorous voiced fricatives "z, v, ð, ʒ" (p = .008). This continuous-contrastive difference appears to suggest different underlying mechanisms for size and shape symbolism. As argue by Thompsons and Estes (2011), continuous marking reflects neurological correspondence because motor activation is continuous (moderate hand or mouth gestures represent mid-sized object); contrastive marking can only be explained by a statistical learning hypothesis where two contrastive phonemes happened to be statistically associated with two antonyms. Consequently, the size symbolic effect is driven by neurological cross-activation, while shape symbolism is the result of statistical learning.

But this is not necessarily the case. As mentioned in Section 1.3, size and shape reside on different perceptual continua (Stevens, 1957). The size continuum is "prothetic" in that the "more" and "less" ends can be measured quantitatively, while the shape continuum is "metathetic" where the "round" and "angular" ends are qualitatively different (Smith & Sera, 1992). Since sonority scale (mainly based on sound amplitude) also constitutes a "prothetic" continuum, it is of no surprise that it matches well with size scale but not shape scale. According to Stevens (1971), all category scales are nonlinear because the scaling results can be described by a power function with a virtual exponent that is smaller than the actual exponent of the continuum. The difference of the two scaling class is even reflected in the morphosyntax of the degree clauses in our language: the comparatives of prothetic adjectives such as large/big, loud, long, high, fast, etc. are generally formed by adding “-er” to their adjectives (larger/bigger, louder, longer, higher, faster, etc.), but comparatives of metathetic adjectives such as round, sweet, sad, etc. are more than often formed by “more + adjective” (more round, more sweet, more sad, etc. there are exceptions such as “bitter” and “happier”, but the number is small compared to that of the prothetic adjectives). Bresnan (1973), among others, suggested that “more” is underlyingly “-er” + “many/much” in morphosyntax. If this is the case, then our degree clauses actually differentiates the two scaling adjectives by directly or indirectly merging them with the degree head “-er”.

Parallel with the direct/indirect merging in morphosyntax, size, and shape symbolism could also be direct/indirect: size symbolism is neurological either by cross-activation or by the nature of the cognitive structure; shape symbolism is grounded on the prototypical magnitude scale to differentiate rotundity/angularity. The results of Experiment 3 suggest this to be possible: rounded objects tend to be perceived as larger than their actual size, whereas angular shapes are more likely to be seen as smaller than the actual size.

Of course size and shape symbolism can both be neurological. Recall the discussion for Task 1 results, where size symbolism seems to be based on cross-activation between visual and auditory brain areas (perceptual), and shape symbolism is based on cross-activation between mouth and hand motor areas (articulatory). Therefore a sonority scale which is perceptual in nature is undoubtedly better correlated with size symbolism. If we used an articulatory scale instead, such as the degree of lip protrusion (roundness), we may observe a better correlation with shape symbolism.

It is also possible that both forms of symbolism originate from our statistical observation of the world where large-sized animals generally produce low-frequency/low-pitched sounds and small-sized animals usually produce high-frequency/high-pitched sounds

(Parise & Spence, 2009; Spence, 2011). The association between sounds and shape is less clear, but we could still argue in the same fashion that “angular” animals like hedgehogs and crocodiles tend to elicit fear or unpleasant feelings which are, in turn, linked to high-frequency/high-pitched sounds (e.g. scream), whereas “round” animals like rabbits and koala are usually associated with affection and low-frequency/low-pitched sounds (e.g. laughter). But since size-sound associations are more robustly attested than shape-sound associations, the statistical correlation between sound and size is stronger than that between sound and shape.

Two last factors to be considered are the grapheme and semantic interference. Apparently, “s, f” was curvy and “z, v” was angular (“θ, ʃ” may also be more curvy than “ð, ʒ”, but we excluded them from consideration since most of our participants are naive to the underlying phonological representations of the sounds). Therefore, it is possible that the graphemes of the phonemes was activated when hearing the sounds and biased the perception, especially for “zi” and “vi”, which are the exact pronunciation for the letters “z” and “v”. In effect, “si (sea/see)”, “ʃu (shoe)”, “ʃi (she)”, etc. are all homonyms of real words, thus semantic interference, either consciously or not, could not be excluded. As a matter of fact, although most participants reported after the experiment that none of the words resembled real words, five of them did mention that “ʃu” reminded them of “shoe” and hence they chose a rounded image. This may well contribute to the higher average roundness score of the “s, f, θ, ʃ” group (voiceless fricatives) than the “z, v, ð, ʒ” group (voiced fricatives).

6. Conclusions

To conclude, different perceptual patterns were observed for size and for shape symbolism. For size symbolism, sonority appears to be the determining factor in that: (1) the more sonorous vowels are more powerful in terms of driving size symbolism than the less sonorous consonants; (2) consonant sonority linearly predicts the symbolic size[8]. For shape symbolism, however, sonority is less effective because: (1) vowels and consonants could be equally salient in predicting shape symbolism; (2) consonant sonority influences roundness, but the effect is non-linear. This may suggest that perceptual-based cross-activation in the visual and auditory brain areas for size symbolism, and articulatory-based cross-activation in the mouth and hand motor brain areas for shape symbolism (see also Spence & Deroy, 2012; Sweeny et al., 2012, for further evidence on how mouth shapes could alter perception).

Alternatively, following Thompson and Estes's (2011) line of argument, size symbolism must be neurological since it shows continuous marking which cannot be explained by a contrastive statistical learning account where two sets of phonetically contrastive phonemes are used to mark two semantically contrastive meaning. Shape symbolism, however, shows a contrastive effect so it may be explained by statistical learning. We further demonstrated results from Experiment 3 to show that there is also a perceptual bias for rounded shapes to be associated with larger objects, suggesting a mediating role of size symbolism.

Nevertheless, size and shape symbolism need not be necessarily different in terms of their underlying mechanism. Both might be neurological, only differing in their patterns of cross-activation mechanisms. They could also be both statistical, only that medium sonorous phonemes could correspond with midsized objects, but there is no mid-rounded shapes to correspond with. Put it another way, shape symbolism is similar to a "bird-fish" symbolism (Berlin, 1994): we may find one set of phonemes frequently used for bird names ("i") and another set of phonemes frequently associated with fish names ("u", "a"), but we cannot say the phonemes in between the two phonetically contrastive phonemes ("e") correspond to

[8] Though it is probably the relative size, not the absolute size that is predicted.

animal names in between birds and fish. We don’t even know what is in between birds and fish at first.

Further studies should be conducted on distinguishing different sound-sense correspondences, especially between those on the prothetic dimensions (e.g., amplitude, height, speed, weight) and those on the metathetic dimensions (e.g., shape, taste, colour, emotion). Neuroimaging studies comparing the brain activation patterns for different sound symbolisms may be a promising direction.

REFERENCES

Abel, G. A., & Glinert, L. H. (2008). Chemotherapy as language: Sound symbolism in cancer medication names. *Social Science & Medicine, 66*, 1863-1869.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.

Atzet, J., & Gerard, H. B. (1956). A study of phonetic symbolism among native Navajo speakers. *Journal of Personality and Social Psychology, 1*, 524-528.

Ben-Artzi, E., & Marks, L. E. (1995). Visual–auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics, 57*, 1151-1162.

Bentley, M., & Varon, E. (1933). An accessory study of phonetic symbolism. *American Journal of Psychology, 45*, 76-86.

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, N. Johanna & J. Ohala (Eds.), *Sound symbolism*, Cambridge: Cambridge University Press, pp. 76-93.

Berlin, B. (2006). The first congress of ethnozoological nomenclature. *The Journal of the Royal Anthropological Institute, 12*, 23-44.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology, 87*, 241-247.

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*, 337-339.

Boyle, M. W., & Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research, 9,* 535-544.

Braaten, R. (1993). *Synesthetic correspondence between visual location and auditory pitch in infants*. Paper presented at the 34th Annual Meeting of the Psychonomic Society.

Brackbill, Y., & Little, K. B. (1957). Factors determining the guessing of meanings for foreign words. *Journal of Abnormal and Social Psychology, 54,* 312-318.

Bremner, A., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K., & Spence, C. (in press). Bouba and Kiki in Namibia? Western shape-symbolism does not extend to taste in a remote population**.** *Cognition*.

Bresnan, J. (1973). The syntax of the comparative clause construction in English. *Linguistic Inquiry, 4,* 275–343.

Brown, R. (1958). *Words and things*. New York: Free Press.

Brown, R., Black, A., & Horowitz, A. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology, 50,* 388-393.

Brown, R., & Nuttall, R. (1959). Method in phonetic symbolism experiments. *Journal of Abnormal and Social Psychology, 59,* 441-445.

Chan, M. (1996). Sound symbolism and the Chinese language. In Cheng, T., Li, Y., & Zhang, H. (eds.) *Proceedings of the 7th North American conference on Chinese linguistics (NACCL) and the 4th international conference on Chinese linguistics (ICCL). Volume 2.* Los Angeles, CA: GSIL Publications, University of Southern California, pp. 17-34.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17,* 205-215.

Chastaing, M. (1958). Le Symbolisme des Voyelles: Significations des 'i' I & II [Symbolism of the vowels: Meanings of 'i' I & II.]. *Journal de Psychologie, 55,* 402-423; 461-481.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., and Maneewonggvatana, S. (2008). Encoding emotions in speech with the size code. *Phonetica, 65,* 210-230.

Crisinel, A., & Spence, C. (2009). Implicit associations between basic tastes and pitch. *Neuroscience Letters, 464*, 39-42.

Cytowic, R. (2002). *Synesthesia: A union of the senses (2nd Ed.)*. Cambridge, MA: MIT Press.

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

Diffloth, G. (1994). i: Big, a: Small. In L. Hinton, N. Johanna & J. Ohala (Eds.), *Sound symbolism*, Cambridge: Cambridge University Press, pp. 130-147.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*, 1–12.

Farmer, T., Christiansen, M., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103,* 12203.

Fikkert, P., & Levelt, C. C. (2008). How does place fall into place? The lexicon and emergent constraints in the developing phonological grammar. In: Avery, P., Dresher, B. E. & Rice, K. (Ed.), *Contrast in phonology: Perception and acquisition*. Berlin: Mouton, pp.231-268.

Fox, C. W. (1935). An experimental study of naming. *The American Journal of Psychology, 47*, 545-579.

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*, 34-46.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*, 1191-1203.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behavior, 8,* 310-312.

Gussenhoven, C., & Jacobs, H. (2005). *Understanding phonology (2nd Ed.)*. London: Hodder Education .

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, *22*, 383-388.

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Center for the Study of Language and Information, Stanford: Stanford University.

Hasada, Rie. 1998. Sound symbolic emotion words in Japanese. In: A. Athanasiadou & E.Tabakowska (eds), *Speaking of Emotions: Conceptualisation and Expression*. Berlin: Mouton de Gruyter, pp. 83-98.

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

Holland, M., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or maluma, and takete revisited. *Perceptual and Motor Skills*, *19*, 111-117.

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

Jespersen, O. (1922). The symbolic value of the vowel i. *Philologica, 1*, 1-19.

Johnson, R. C. (1967). Magnitude symbolism of English words. *Journal of Verbal Learning and Verbal Behavior*, *6,* 508-511.

Jowett, B. (2009). *Plato's Cratylus with introduction*. Rockville, MD: Serenity Publishers.

Komatsu, M., Tokuma, S., Tokuma, W., & Arai, T. (2002). In Proceedings of interspeech 2002.

Kawahara, S., Kazuko S., & Yumi U. (2008). A positional effect in sound symbolism: An experimental study. *Proceedings of the 8th annual meeting of Japanese cognitive linguistics association*, pp. 417-427.

Kim, K. (1977). Sound symbolism in Korean. *Journal of Linguistics, 13,* 67-75.

Klink, R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters, 11*, 5-20.

Klink, R. (2001). Creating meaningful new brand names: a study of semantics and sound symbolism. *Journal of Marketing Theory and Practice, 9*, 27-34.

Klink, R., & Athaide, G. (2012). Creating brand personality with brand names. *Marketing Letters, 23*, 109-117.

Knöferle, K. M., & Spence, C. (in press). Crossmodal correspondences between sounds and tastes. *Psychonomic Bulletin & Review*.

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

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2009). The shape of words in the brain. *Cognition, 114,* 19–28.

Ladefoged, P. (1993). *A course in phonetics* (3rd Ed.). Fort Worth, London: Harcourt Brace College Publishers.

LaPolla, R. (1994). An experimental investigation into phonetic symbolism as it relates to Mandarin Chinese. In L. Hinton, N. Johanna & J. Ohala (Eds.), *Sound symbolism*, Cambridge: Cambridge University Press, pp. 130-147.

Lindauer, M. (1990). The meanings of the physiognomic stimuli taketa and maluma. *Bulletin of the Psychonomic Society, 28,* 47-50.

Loomis, J., Lippa, Y., Klatzky, R., & Golledge, R. (1998). Spatial updating of locations Specified by 3-D sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 335-345.

Lowrey, T., & Shrum, L. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research*, *34*, 406-414.

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

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384-394.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception, 28*, 903-923.

Matisoff, J. A. (1994). Tone, intonation, and sound symbolism in Lahu: loading the syllable canon. In L. Hinton, N. Johanna & J. Ohala (Eds.), *Sound symbolism*, Cambridge: Cambridge University Press, pp. 115-129.

Maurer, D., & Mondloch, C. J. (2005). Neonatal synaesthesia: A reevaluation. In L. C. Robertson & N. Sagiv (Eds.), *Synaesthesia: Perspectives from cognitive neuroscience*. Oxford: Oxford University Press, pp. 193–213.

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316-322.

Melara, R. D. (1989a). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 69-79.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. Journal of Experimental Psychology: General, 116, 323-336.

Mesz, B., Trevisan ô, M., & Sigman, M. (2011). The taste of music. *Perception, 40*, 209-219.

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

Newmeyer, F. (1993). Iconicity and generative grammar. *Language, 68,* 756-796.

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 Psychology*, *65*, 115-124.

Ngo, M. K., Misra, R., & Spence, C. (2011) Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference, 22,* 567-572.

Nuckolls, J. (1999). The case for sound symbolism. *Annual Review of Anthropology, 28,* 255-282.

Nygaard, L., Cook, A., & Namy, L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*, 181-186.

Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound Symbolism, 2*, 325-347.

Ohala. J. (1996). Speech percept ion is hearing sounds, not tongues. *Journal of the Acoustical Society of America, 99*, 1718-1725.

Osgood, C. E. (1960). The cross-cultural generality of visual–verbal synesthetic tendencies. *Behavioral Science, 5*, 146-169.

Parker, S. (2002). *Quantifying the sonority hierarchy*, Linguistics, University of Massachusetts Amherst: PhD. Dissertation.

Parise, C., & Spence, C. (2009). "When birds of a feather flock together": Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE, 4,* e5664.

Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 755-775.

Peiffer-Smadja, N. (2010). Exploring the bouba/kiki effect: a behavioral and fMRI study. Neuropsychology and Neuroimaging: Université Paris Descartes: MSc Dissertation.

Ramachandran, V. S., & Hubbard, E. (2001). Synaesthesia — a window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3-34.

Ramachandran, V. S., & Hubbard, E. M. (2003). Hearing colors, tasting shapes. *Scientific American, 288*, 43-49.

Rogers, S. K., & Ross, A. S. (1975). A cross-cultural test of the Maluma–Takete phenomenon. *Perception, 4*, 105–106.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12,* 225-239.

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

Seo, H., & Hummel, T. (2011). Auditory-olfactory integration: Congruent or pleasant sounds amplify odor pleasantness. *Chemical Senses, 36*, 301-309.

Sherzer, J. (1976). *An areal-tyological study of American Indian languages north of Mexico*. Amsterdam: North-Holland.

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

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology, 24*, 99-142.

Snow, D. (2001). Transcription of suprasegmentals. *Topics in Language Disorders, 21*, 41-51.

Spence, C. (2011). Crossmodal correspondence: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971-995.

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*, *22*, 37-54.

Spence, C., & Deroy, O. (in press). Hearing mouth shapes: Sound symbolism and the reverse McGurk effect. *i-Perception*.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153-181.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review, 78,* 426-450.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2012). Sounds exaggerate visual shape. *Cognition*, *124*, 194-200.

Taylor, I. K. (1963). Phonetic symbolism re-examined. *Psychological Bulletin, 60*, 200–209.

Taylor, I. K., & Taylor, M. M. (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology, 16*, 344-356.

Taylor, I. K. & Taylor, M. M. (1965). Another look at phonetic symbolism. *Psychological Bulletin*, *64*, 413-427.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology, 64,* 2392-2404.

Tsur, R. (2006). Size–sound symbolism revisited. *Journal of Pragmatics, 38,* 905-924.

Uznadze, D. (1924). Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung [An experimental contribution to the problem of the psychological bases of naming]. *Psychologische Forschung*, *5*, 25-43.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. (2010). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science, 21*, 21-25.

Ward, J., Huckstep B., & Tsakanikos, E. (2006). sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? *Cortex, 42*, 264- 280

Wertheimer, M. (1958). The relation between the sound of a word and its meaning. *The American Journal of Psychology, 71*, 412-415.

Wesson, D. & Wilson, D. (2010). Smelling sounds: olfactory-auditory sensory convergence in the olfactory tubercle. *Journal of Neuroscience, 30*, 3013-3021.

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

Woods, A., Poliakoff, E., Lloyd, D., Kuenzel, J., Hodson, R. Gonda, H., Batchelor, J., Dijksterhuis , G., & Thomas, A. (2011). Effect of background noise on food perception. *Food Quality and Preference, 22*, 42-47.

Wrembel, M. (2007). "Still sounds like a rainbow" - a proposal for a coloured vowel chart. In *Proceedings of the Phonetics Teaching & Learning Language Conference*, *August 2007*.

Wrembel, M., & Rataj, K. (2008) "Sounds like a rainbow" - sound-colour mappings in vowel perception, in Botinis, A. (Ed.): *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, ExLing 2008*, Athens: University of Athens, 237-240.

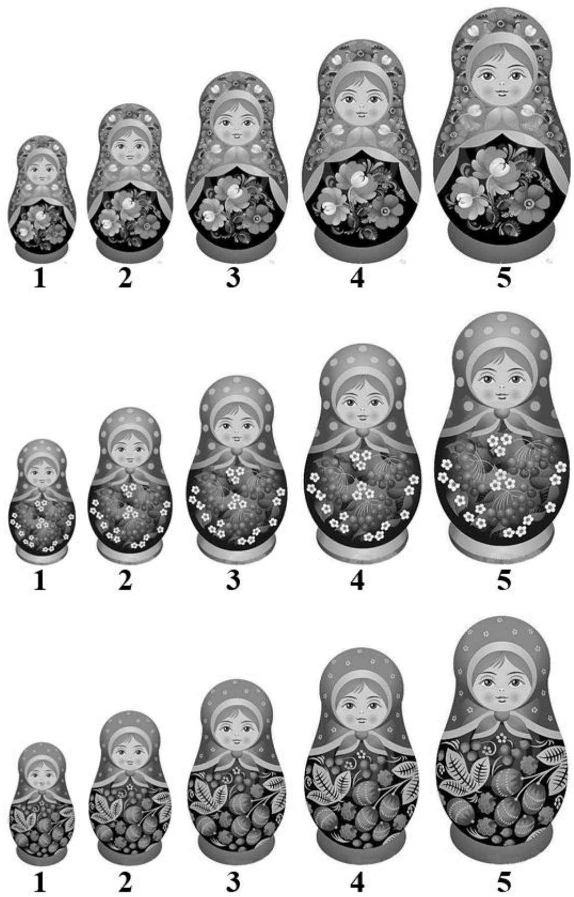
Yau J. M., Olenczak J. B., Dammann J. F., & Bensmaia S. J. (2009). Temporal frequency channels are linked across audition and touch. *Current Biology, 19*, 561-566.

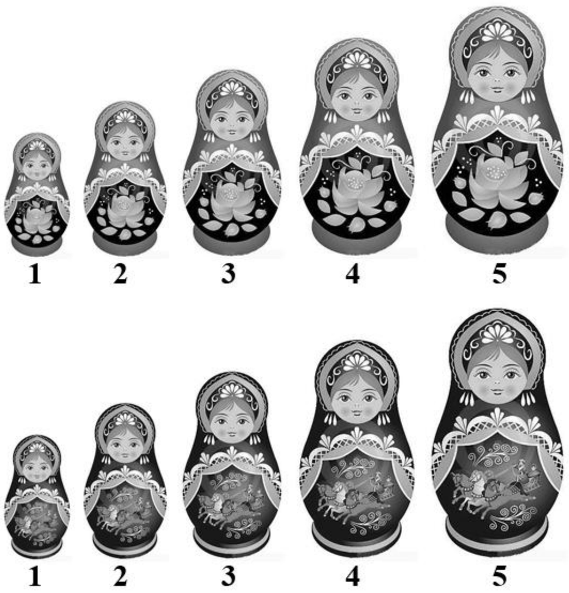
Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research*, *31*, 43-51.

APPENDIX

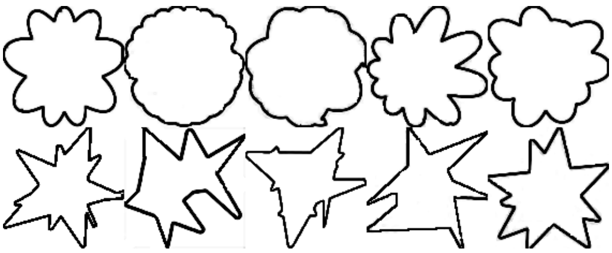
1. The 5 dolls used in Task 1, Experiment 1.

2. The 5 size scales used in Task 2, Experiment 1.

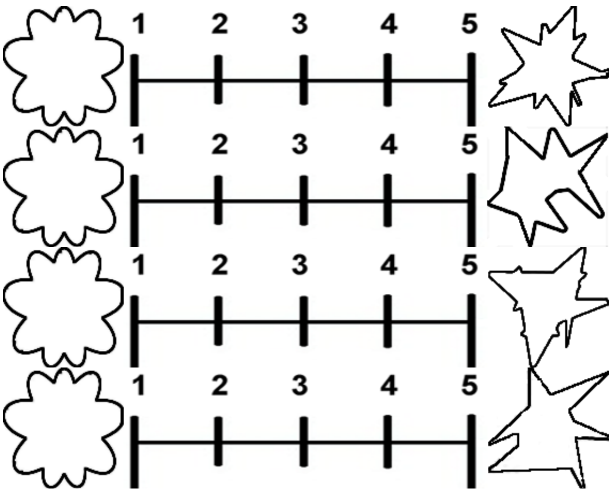


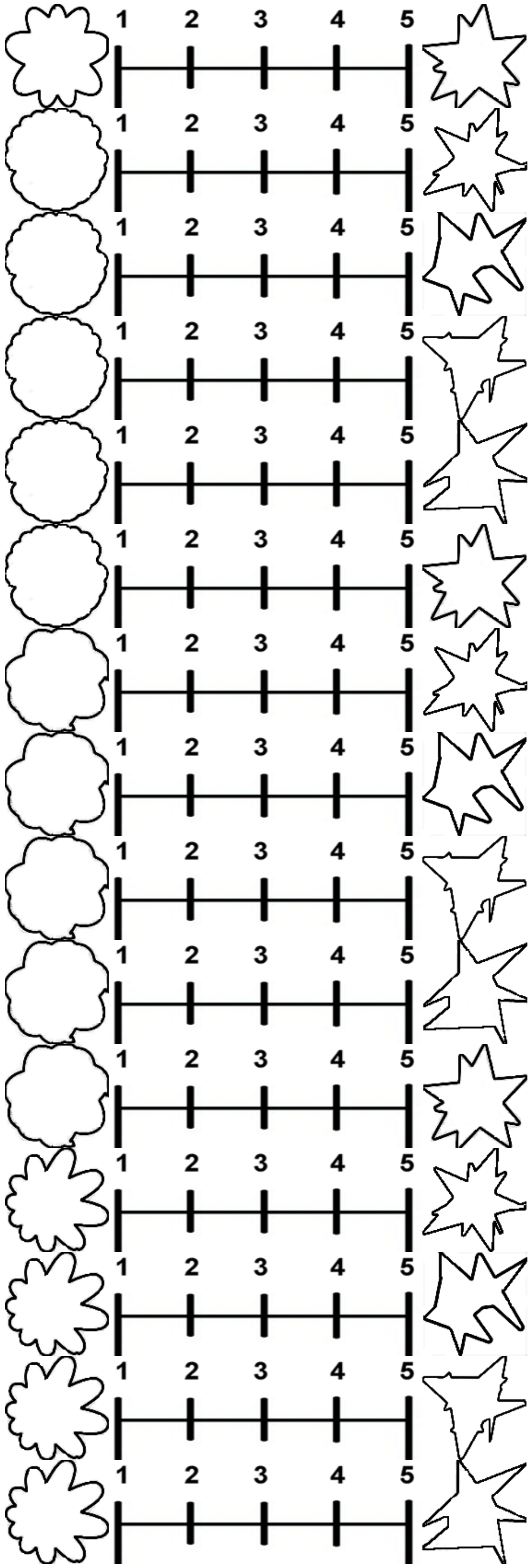


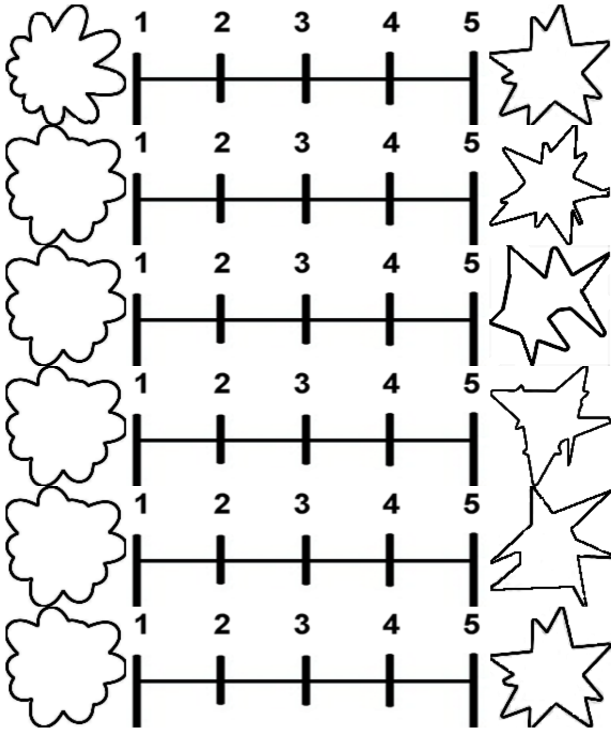
3. The 10 shapes used in Task 1, Experiment 2.



4. The 25 shape scales used in Task 2, Experiment 2.




1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5


1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5

ACKNOWLEDGMENTS

I would like to express my great gratitude to the Department of Experiment Psychology, University of Oxford for funding this study, and to my supervisors Prof. Charles Spence and Dr. Klemens Knöferle for their very detailed guidance and comments on the drafts of the thesis, and their patience and encouragement during the process of writing.

The thesis also benefits a great deal from discussion with Prof. Yi Xu at the Department of Linguistics, University College London, and Prof. Aditi Lahiri at the Department of Linguistics, University of Oxford.

I'm also very grateful to Mr. Naxian Wang at the Department of Statistics, University of Oxford for helping organize and analyse the data.