Sound Symbolism in Sighted and Blind
The role of orthography and vision in sound-shape correspondences

Roberto Bottini[1,2*], Marco Barilari[1,2] & Olivier Collignon[1,3]

[1]Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy

[2]International School for Advanced Studies (SISSA), Trieste, Italy

[3]Institute of Psychology (IPSY) and Institute of Neuroscience (IONS), University of Louvain

Word count

Abstract: 161

Article: 5937

*Corresponding Author:

Roberto Bottini, Ph.D.

CIMeC - Center for Mind/Brain Sciences, University of Trento

Via delle Regole 101, 38123, Mattarello (TN), Italy.

Phone: +390461283010

bottini.r@gmail.com

Abstract

Non-arbitrary sound-shape correspondences (SSCs) have been consistently observed across languages and together with other sound-symbolic phenomena facilitate language acquisition and may be important for language evolution. Yet, it is unclear what makes a sound “round” or “spiky” to the human mind. One possibility is that we automatically map shapes onto the motor representations used to articulate speech-sounds. However, some SSCs deviate from predictions based on articulation only, calling for additional mechanisms. We hypothesize that the shape of written letters (spontaneously activated by spoken words) may influence SSCs. We tested this hypothesis, in early blind people, who are exposed to an orthography (Braille) in which letters do not have spiky or round outlines. As predicted, early blind showed SSCs that reflected a pure articulatory approach better than sighted people. Conversely, sighted’ SSCs were better explained by combining articulation and orthography. This result shed new light on the experiential basis of sound-symbolism, and demonstrates that vision is not necessary to develop sound-shape correspondences.

Keywords: Sound symbolism, bouba-kiki, blindness, orthography.

1.Introduction

Sound symbolism is the ability to associate verbal acoustic stimuli (sounds or phonemes) with non-acoustic features. A popular example is the shape-sound correspondence. In the previous century Wolfgang Khöler (Kohler, 1947) asked people to decide which of two shapes was *maluma* and which one was *takete* (Figure 1). The large majority of the participants associated the two words with the round and pointy shape respectively. More recently, this paradigm was popularized by Ramachandran and Hubbard (Ramachandran & Hubbard, 2001) using the (non)words *Bouba* and *Kiki*. Since then, the *bouba-kiki effect* has been replicated in many experiments (Ahlner & Zlatev, 2010; Aveyard, 2012; Maurer, Pathman, & Mondloch, 2006; Nielsen & Rendall, 2011; Westbury, 2005) across different cultures (Bremner et al., 2013; Chen, Huang, Woods, & Spence, 2016).

Despite extensive investigation, it is still unclear what drives the existence of sound-shape correspondences. This question is of particular interest since sound-symbolic associations (e.g. onomatopoeia, phonestemes, iconicity) are quite frequent across human languages (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016) and defies the classic dictum of the arbitrariness of the sign (de Saussure, 1959). Moreover, Sound symbolism facilitates word learning in both children (Imai et al., 2015; Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Imai, & Kita, 2011) and adults (Lockwood, Dingemanse, & Hagoort, 2016; Monaghan, Mattock, & Walker, 2012; Nygaard, Cook, & Namy, 2009) and may shed light on the processes of language evolution, where non-arbitrary relationships between sounds and objects may have served as a bootstrap for a more complex and largely arbitrary reference-system (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Imai & Kita, 2014; Perlman, Dale, & Lupyan, 2015; Ramachandran & Hubbard, 2001).

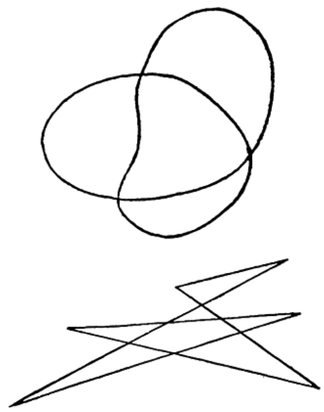

Figure 1. Maluma and Takete in Köhler's original drawings.

What makes a sound "spiky" or "round"? It has been suggested that sound-shape associations are mediated by articulatory movements of the mouth and tongue during speaking (Ramachandran & Hubbard, 2001). Phonemes are mapped onto the motor representations needed to produce them, which in turn are mapped onto a shape. For consonants, the iconic ground for this mapping may be provided by the manner of articulation, i.e. the way the airstream is affected as it flows from the lungs and out to the mouth and nose. For instance, sonorant consonants (/n/ /m/ /l/) that are produced with smooth articulatory movements and a continuous and unobstructed airflow should be associated with round continuous shapes; stop consonants (/d/ /g/ /b/ /t/ /k/ /p/) that are produced with sharp inflections of the tongue on the palate and completely obstructed airflow should map onto irregular shapes; and fricatives/affricates (/v/ /z/ / d͡ʒ/, /f/, /s/, /t͡ʃ/), with a partially obstructed airflow should be somehow in the middle (Aveyard, 2012; Kovic, Plunkett, & Westermann, 2010; Mccormick, Kim, List, & Nygaard, 2015; Westbury, 2005). A similar iconic association can be made for vowels, but on the bases of lips movements, with /o/ and /u/ produced by a curvy movement of the lips, and associated with round shapes

(Nielsen & Rendall, 2011; Ozturk, Krehm, & Vouloumanos, 2012). This articulatory theory[1] of sound symbolism (Ramachandran & Hubbard, 2001) may explain why this phenomena appears to be universal across cultures (Blasi et al., 2016; Bremner et al., 2013) and emerges very early in life (Ozturk et al., 2012).

Yet, the predictions of the articulatory theory are not straightforwardly met by the data. For instance, in several experiments, stop consonants such as /d/ and /b/ are mostly associated with round than spiky shapes, as in the famous case of *bouba* and *kiki* (Bremner et al., 2013; Fryer, Freeman, & Pring, 2014; Ramachandran & Hubbard, 2001). This is true also when participants are not asked to choose between stop consonants (/b/ Vs /k/), and when the effect of vowels is controlled (D'Onofrio, 2014; Mccormick et al., 2015). It has been suggested that the roundedness of the /b/ and /d/ phonemes can be explained by the fact that they are voiced (produced with the vibration of the vocal cords) compared to the unvoiced /k/ and /t/ (D'Onofrio, 2014; Kovic et al., 2010; Mccormick et al., 2015). Indeed, voicing may be another aspect of articulation that influence the association with shape (D'Onofrio, 2014; Kovic et al., 2010): Unvoiced consonants are usually associated with a greater formant (F1) transition and sharper articulatory movements compared to voiced ones (Mccormick et al., 2015), which may provide the iconic ground for mapping continuous and angular shapes. However, it is unclear why voicing seems to have a prominent effect between stop consonants but not between fricatives/affricates.(Mccormick et al., 2015)

[1] It is unclear to what extent the relationship between manner of articulation and shape is based on the activation of motor programs (Ramachandran & Hubbard, 2001) or onto direct mapping of acoustic features that covary with articulation. Throughout the text we will refer to this set of associations as the "articulatory model", although we do not exclude that direct sound-shape mapping can also take place (see Discussion).

Thus, a pure articulatory theory it is not enough to explain the patterns of sound symbolism reported in several previous studies. Other factors must play a role in sound-shape associations. For instance, several authors have suggested that cross-modal sound-shape mappings may reflect statistical regularities of our experience in the physical world. For example, pointed objects may tend to produce sounds that are less tonal and have more abrupt transitions (Fryer et al., 2014; Mccormick et al., 2015). Yet, statistical regularities in the environment that can account for different instances of sound-shape correspondences have yet to be empirically demonstrated (Deroy & Spence, 2013). Here, we suggest that the deviation from a "pure" articulatory approach can be largely explained by the influence of the shape of written letters.

Consonants and vowels are not only associated with a sound, but, at least in the mind of literate people, also with an orthographic sign. The role of orthography in sound symbolism has been mostly approached as a source of confounds to be ruled out (Hung, Styles, & Hsieh, 2017; Westbury, 2005), showing that shape-sound associations *can* emerge independently of orthography. Studies on pre-literate children (Maurer et al., 2006), babies (Ozturk et al., 2012) and illiterate adults (Bremner et al., 2013) successfully proved this point.

Yet, there is also evidence that letter shape can modulate sound-symbolic judgments (Doyle & Bottomley, 2011). For instance, in one of the few systematic studies of orthographic influence on sound symbolism, Cuskley and colleagues (Cuskley, Simner, & Kirby, 2015) showed that orthographic spykeness (i.e. the angularity of letters shape) was actually the strongest predictor of the bouba-kiki effect when compared to voicing and manner of articulation, even with aurally presented words. Nevertheless, the relative contribution of orthographic and phonological factors cannot be clearly drawn from Cuskley et al., (Cuskley et al., 2015) given that: (i) only 8 items were used in their experiment; (ii) voicing and manner of articulation were not orthogonalized with respect to orthography; and (iii) the articulatory spectrum was not fully covered (e.g. there were not sonorant ('round') consonants). Moreover, the hypothesis that orthography is the main drive behind sound-shape correspondences can hardly explain the results with toddlers, babies and illiterate adults (Bremner et al., 2013; Maurer et al., 2006; Ozturk et al., 2012).

In this paper we suggest a different approach according to which: (i) Both voicing and manner of articulation contribute to sound-symbolic associations and are sufficient for the emergence of the typical bouba-kiki effect; (ii) When listening to words people spontaneously activate the orthographic representation of these words (Chéreau, Gaskell, & Dumay, 2007; Ziegler & Ferrand, 1998), and the shape of written letters influences cross-modal mappings between shape and sound. Thus, previously reported sound-symbolic patterns that deviate from a pure articulatory account can be explained by the effect of letter shape. We call this the Blending Orthography and Articulation Hypothesis (BOAH).

Early blindness was used as a model system to test this hypothesis. Although EBs speak, read and write using the same language than their sighted counterparts (in our case, Italian), they are exposed to a very different orthography provided by the tactile alphabet Braille. Contrary to written letters, Braille letters are composed by dots arranged in a 3X2 grid, without spiky or round edges as letters in the written alphabet (see Fig 2). Since braille alphabet cannot provide letter representations in terms of roundness or spikiness, we predict that the pattern of sound-symbolic association shown by blind people would be closer to what a pure articulatory account predicts.

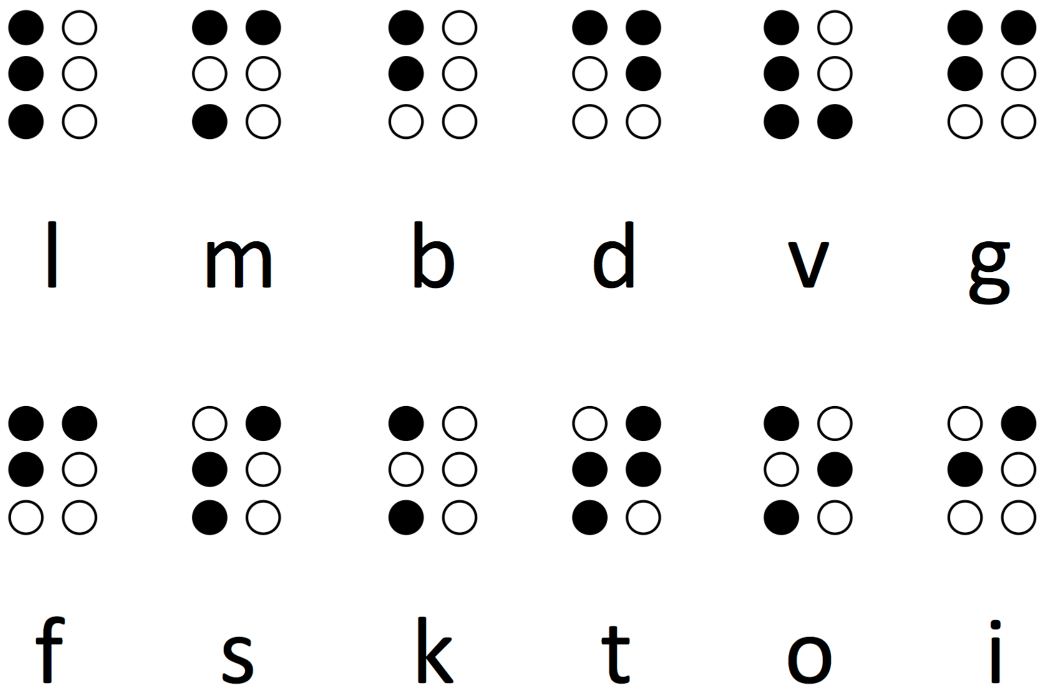

Figure 2. Comparison between letters in Braille and Latin alphabet.

The study of blind people allows us to answer also another question: Is vision necessary to develop sound symbolism? Distinguishing shapes is normally a highly visual process (Denys et al., 2004), thus vision may play an important role in the establishment of shape-sound correspondences. But what happens when visual experience is not available, as in the case of early blind people? According to the articulatory approach, shape-sound correspondences should emerge anyway. Articulatory movements for speech are learned also in the absence of vision and the shape of objects can be conveyed by haptic experience (Peelen, He, Han, Caramazza, & Bi, 2014; Pietrini et al., 2004). Moreover, cross-modal associations between acoustic features, articulatory movements and shape can be largely innate (Ramachandran & Hubbard, 2001). Yet, recent experiments seem to suggest that functional vision may be necessary for the emergence of the bouba-kiki effect, challenging *de facto* the articulatory hypothesis. Fryer and colleagues (Fryer et al., 2014) tested sighted people and a heterogeneous group of visually impaired subjects (including early blind, late blind and partially sighted) in a haptic version of the bouba-kiki task. Although a sound-symbolic effect emerged in both populations, the effect was significantly lower in the visually impaired group. Interestingly, the sub-group of 6 early blind did perform at chance in the task, showing no effect at all (Fryer et al., 2014). Having experienced functional vision may be necessary to develop shape-sound symbolism.

The extant literature suggests at least two reasons why this might be the case. First, it has been suggested that multisensory integration may be impaired in the absence of developmental vision (Champoux et al., 2011; Collignon, Charbonneau, Lassonde, & Lepore, 2009; Putzar, Goerendt, Lange, Rösler, & Röder, 2007). If this is the case, and if sound-symbolism is based on multisensory integration (Lockwood & Dingemanse, 2015; Occelli, Esposito, Venuti, Arduino, & Zampini, 2013; Ozturk et al., 2012; Spence, 2011), blind people may show a reduced bouba-kiki effect (Fryer et al., 2014). Second, sight may give access to regular associations between shape and sounds in the environment that are precluded to haptic experience (Fryer et al., 2014). Some examples may be the association between shape and sounds in commercial brands (Doyle & Bottomley, 2011) or the shape of someone else's lips when we see her speaking (Ramachandran & Hubbard, 2001).

Although the statistical regularities that link shapes and sounds in the environment are still unspecified (Deroy & Spence, 2013), the lack of sound-symbolism in the blind would be a powerful evidence against the innatist view. Most importantly for this study, if early blindness prevents shape-sound associations, it may not constitute an ideal test-bed for our hypothesis. Thus we first performed a close replication of Fryer et al. experiment (Fryer et al., 2014) with a larger number of early blind participants (n=30) to test whether EB show a bouba-kiki effect.

2. Experiment 1

Experiment 1 was a fateful replication of Fryer et al., experiment (Fryer et al., 2014), with one important exception. Contrary to Fryer et al., who use the words *bouba* and *kiki*, we preferred to use Köler's original words *maluma* and *takete.* Both bouba and kiki contain stop consonants (/b/ and /k/) and, according to our hypothesis, the association between /b/ and round shapes could be inflated by the shape of the written letter 'b'. If this is the case, bouba and kiki may sound pretty similar in terms of spikeness to EB participants who are not influenced by canonical orthography. This might explain, at least in part, Fryer et al. results. Since Experiment 1 was designed to test whether EB develop sound symbolism independently of orthography, the stimuli *maluma* and *takete* seemed more appropriate, being at the two poles of the articulatory spectrum, at least in terms of consonants.

2.1 Methods

2.1.1 Participants

Sixty participants completed the experiment in exchange of payment: 30 early blind (EB) and 30 sighted controls (SC). Participants in the EB and SC group were matched pairwise for gender (17 females), age (EB = 36.27, sd= 10.86; SC = 36.70, sd= 11.62) and years of education (EB = 14.10, sd= 3.09; SC = 14.43, sd= 2.92; see also Supplementary Information). All participants were Italian native speakers and were blindfolded during the task. Participants in the EB group lost their sight at birth of before 4 years of age and were totally blind or had only rudimentary sensitivity for brightness differences. All of them could never perceived shapes and/or have no visual memory of shapes. In all cases, blindness was attributed to peripheral deficits with no additional neurological problems. All our blind participants had no experience of visual reading and were fluent in Braille.

The sample size was decided based on the amount of early blind participants we could reach and the strength of the effect. Although the experiment had only 4 trials, the bouba-kiki effect is know to be fairly strong, with ~80-90% of participants choosing the typical shape-sound associations. Moreover, we substantially increased the number of early blind participants (from 6 to 30) compared to the experiment that we intend to conceptually replicate (Fryer et al., 2014). The ethical committee of the University of Trento approved this study and all participants were naive with respect to the purpose of the experiment.

2.1.2 Material

Stimuli were four pairs of shapes modeled after Fryer et al. 2014. They were custom made in solid resin with a 3-D printer (Figure 3). Pair A and pair B were designed to mimic Köler line drawings of *maluma* and *takete*. Objects in Pair A were 3D shapes, one smooth and bulb-shaped and the other spiky and irregular in all dimensions. Pair B consisted of 2D shapes, one rounded and one spiky all over. In Pair C there were two discs identical in shape but different in texture (smooth Vs. cross-hatched) and in pair D we had two spheres, one smooth and the other spiky all over. Each pair of objects was contained in a box measuring 300 mm x 17 mm x 12 mm.

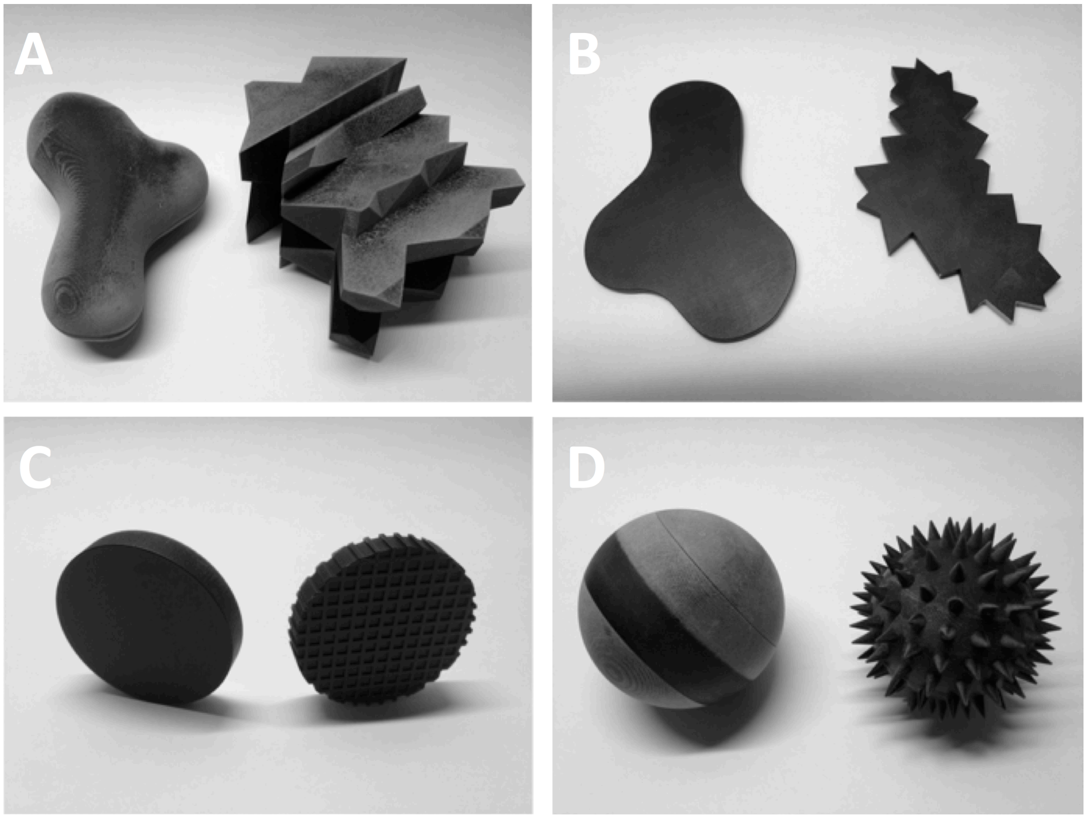

Figure 3. Object pairs used in Experiment 1.

2.1.3 Procedure

Instructions were provided to the subjects orally, according to a pre-defined script (see Supplementary material). The four boxes were presented on the table in front of the person, one at the time. Participants, who were blindfolded before entering the room, were asked to explore the two objects inside the box with both hands and to bring out either *maluma* or *takete.* The experiment consisted in four trials, one for each object pair. The order of trials was determined according to four different lists (ABCD, BCDA, CDAB, DABC) counterbalanced across subjects. In half of the trials participants were asked to bring out "*maluma*" and in the other half to bring out "*takete*" (alternated across trials). The counterbalancing led to a total of 8 (4 X 2) different rotations. For each trial, subjects scored 1 point if they mapped word and shape in the expected way (e.g. spiky object as *takete*), otherwise they scored zero. The total score, thus, ranged from 0 to 4. All participants were debriefed after the experiment. The interview was conducted orally and we asked: "How did you choose which one was *maluma* and which one was *takete*?" (see supplementary material).

2.2 Results

Both groups showed a sound symbolic effect mapping shapes to words in the expected manner (Figure 4). The SC group scored on average 3.63/4. Twenty-five out of thirty (83.33%) chose as expected for all 4 pairs, selecting rounded objects as *maluma* and spiky objects as *takete*; one subject choose consistently in the opposite direction and the remaining 4/30 (13.33%) were inconsistent, choosing a rounded shape sometimes as *takete* and sometimes as *maluma*. The EB group had an average score of 3.13/4. Twenty-two participants out of thirty (73.33%) choose as expected in all the 4 trials, 5/30 (16.67%) chose consistently in the opposite way and the 3/30 (10%) were inconsistent.

A one sample *t*-test showed that the score of each group was significantly higher than chance, for the SC group: $t\ (29) = 9.227$, p = < .001, Cohen's d = 1.69, and for EB group: $t\ (29) = 3.954$, p = < .001, Cohen's d = 0.72. For consistency with Fryer et al. (Fryer et al., 2014) analysis, we ran a one way ANOVA showing that the main effect of group (SC Vs. EB) was not significant ($F\ (1, 58) = 2.21$, p = .14, Cohen's d = 0.39). To further qualify this null results we run Bayesian repeated measure ANOVA (JASP Team, 2017) which revealed very little (anecdotal) evidence against the null hypothesis (BF10 = 0.659; Kass & Raftery, 1995).

Since ANOVAs assume a normal distribution, which is hardly the case for binomial variables, we confirmed this null result with a non-parametric permutation test resampling our data 1000 times. Also in this case the difference between groups was not significant (p=.17, Confidence interval = [.14, .18]).

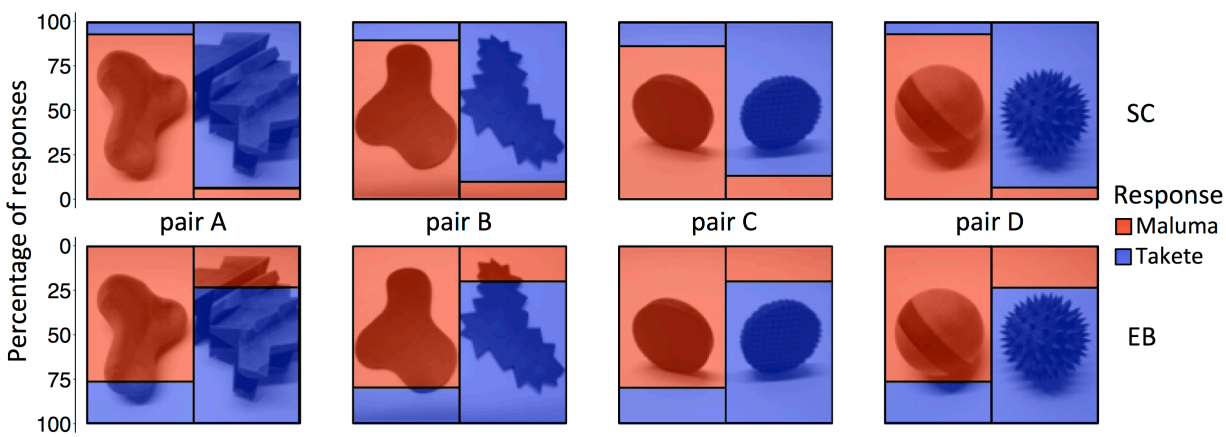

Figure 4. Percentage of maluma and takete responses for each object in the SC (first row) and EB (second row) group.

2.3 Discussion

Early loss of vision does not prevent the development of sound symbolism. Early Blind individuals showed a significant bouba-kiki effect that was statistically indistinguishable from their sighted counterpart. This result is in keeping with the articulatory account, given that the mapping from articulatory movements to object shape, mediated by phonological and acoustic maps, can in principle take place without vision (Ramachandran & Hubbard, 2001).

Our results contrast with the data reported by Fryer and colleagues, who found a significantly reduced effect of sound symbolism in a heterogeneous population of visually impaired people (Early blind, late blind and partially sighted; Fryer et al., 2014a). This discrepancy can be explained (at least in part) by the effect of orthography on sound-shape associations. Fryer and colleagues used the nonwords *bouba* and *kiki*, that are both composed by stop consonants and should be quite similar in term of angularity according to a pure articulatory approach. The "round" sound attributed to voiced stop consonants such as /b/ and /d/ may be a product of the activation of orthographic representations during phonological processing. The activation of orthographic information may not be as automatic in visually impaired people who have limited reading experience, and/or may concern a different orthographic format (without angles and curvatures) if their primary orthographic system is Braille. Thus, in Fryer et al experiment, *bouba* may have sound quite spiky and not so different from *kiki* to blind participants. This state of affairs can explain why Fryer's results differ from ours, and is in line with BOAH.

3. Experiment 2.

Having demonstrated that sound-shape mapping can emerge in EB, in Experiment 2 we proceeded for a direct test of BOAH. EB and SC listened to several pseudowords and decided whether each word sounded "round" or "spiky". We predicted that EB (who have limited experience with canonical orthography) should show a pattern of sound symbolism closer to the prediction of a pure articulatory account. On the other hand, sighted responses should be explained better by a combined effect of articulation and orthography.

3.1 Methods

3.1.1 *Participants*

Thirty-six participants completed the experiment in exchange of payment: 18 early blind (EB) and 18 sighted controls (SC). Participants in the EB and SC group were matched pairwise for gender (9 females), age (EB = 38.56, sd= 13.07; SC = 39.56, sd= 12.91) and years of education (EB = 15.22, sd= 3.26; SC = 15.83, sd= 2.96; see also Supplementary Information). All participants were Italian native speakers and were blindfolded during the tasks. Half of the participants (9 EB and 9 SC) participated previously to Experiment 1. Participants in the EB group were totally blind or had only rudimentary sensitivity for brightness differences. In all cases, blindness was attributed to peripheral deficits with no additional neurological problems. All our blind participants had no experience of visual reading and were fluent in braille.

Given the lack of previous estimation of the effect size of orthographical factors on sound symbolism (especially when the articulatory factors are taken into account), we did not have a precise estimate on which we could base our sample size. Given the relatively large number of trials (n= 240) we thought that a standard sample size of 18-20 participants could be acceptable. Moreover a similar sample size (n=20) has been recently used to detect subtle subliminal effects of sound-symbolic associations (Hung et al., 2017). Finally, as it is often the case for research with special populations, our sample size choice was constrained by the amount of early blind participants that we could reach and that would agree to complete a relatively long experiment (~100 minutes). The ethical committee of the University of Trento approved this study and all participants were naive with respect to the purpose of the experiment.

3.1.2 Material

Stimuli consisted in 240 pseudo-words audio recorded by a male Italian native speaker and edited into separated audio files with the same auditory properties (44100 Hz, 32 bit, mono, 78 dB of intensity). We systematically sampled the acoustic and phonemic space using pseudo-words that were composed by one of 5 classes of consonants (see McCormick et al. 2015 for a similar design). These 5 classes can be arranged from round-sounding to spiky-sounding according the manner of articulation and voicing: (1) Sonorant; (2) Voiced Af/Fricative; (3) Unvoiced Af/Fricative; (4) Voiced Stop; (5) Unvoiced Stop (Aveyard, 2012; Kovic et al., 2010; Mccormick et al., 2015).

All words have a CVCVCV structure based on the two pseudo-words used in Exp.1 (see supplementary information for the complete list). Each stimulus contained consonants from one of five different classes: Sonorants (/n/, /m/, /l/), Voiced Af/Fricative (/v/, /z/, /d͡ʒ/), Unvoiced Af/Fricative (/f/, /s/, /t͡ʃ/), Voiced Stop (/d/, /g/, /b/) or Unvoiced Stop (/t/, /k/, /p/). There were 48 words for each consonant class, meaning that all the consonants in a given word came from the same class. All words had equal consonants in the first and third position, and a different consonant in the second position (as in *maluma* and *takete*). Each consonant appeared an equal number of times across the stimuli.

For each consonant class, half of the stimuli contained back/rounded vowels (/o/, /u/), and the other half contained front/unrounded vowels (/e/, /i/). In half of the stimuli the same vowel appeared in the first and third position (e.g. mulomu), and for the other half in the second and third position (e.g. tekiti).

3.1.3 Procedure

All the participants completed the survey on-line using the platform SurveyGizmo™. The survey was made accessible for visually impaired people and its accessibility was validated by two blind participants (not included in the sample) before starting data collection. Participants were strongly encouraged to listen to the stimuli through headphones for a better sound quality. Each stimulus in the set was presented aurally one by one and participants were asked to focus only on the sound of the pseudo-word and to classify it as round or pointy. The order of the stimuli was pseudo-randomized according to two different lists counterbalanced across subjects. The position of the two responses (i.e. round – pointy or pointy – round) was also counterbalanced across subjects. The counterbalancing led to a total of 4 (2 X 2) different versions of the survey.

3.1.4 Analyses and predictors

Analyses were conducted using a Mixed Effect GLM with binomial distribution.[2] Data were analyzed using R (R Core Team, 2013) and the package lme4 (Bates, Mächler, Bolker, & Walker, 2014). Random effects were the intercepts of subjects and items, whereas fixed effects were Consonant articulation (5 levels, ordered for manner of articulation and voicing), Vowel articulation (round, unround), Orthographic Spikeness, and Group (EB, SC).

[2] It is still debated whether or not to include random slopes in mixed effect models. Although it has been suggested that using only random intercepts may increase Type I error (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013), this it is not always the case (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017) and maximal models with random slopes are highly conservative, increasing Type II error (Matuschek et al., 2017). Moreover, they may over-fit the data (Matuschek et al., 2017) and fail to converge. In the text we decided to report results from the model with intercepts only, whereas in the supplementary material we reported results from the highly conservative maximal model. The crucial main effects and interactions remain significant also when the most conservative statistics are adopted (see supplementary materials).

The ordered values of Articulatory Spykeness (AS from here) and the magnitude of each step was chosen based on previous results in the literature (D'Onofrio, 2014; Kovic et al., 2010; Mccormick et al., 2015) and the theoretical assumption that voicing contributes less than manner of articulation in determining spikeness associations (see Introduction). AS increased by 1 point for change in manner of articulation and 0.5 for change in voicing, resulting in the following scale: Sonorant = 1; Af/Fricative Voiced = 2; Af/Fricative Unvoiced = 2.5; Stop Voiced= 3; Stop Unvoiced = 3.5. Values of AS were standardized before running statistics to allow a meaningful comparison between odd ratios.

The regressor Orthographic Spikeness (OS from here) was created as follows. Forty additional sighted participants completed a survey on-line in which they had to rate the visual spikeness of single letters from 1 to 7 (1= not spiky, 7= very spiky). Half of the participants saw the letters in their capital notation (e.g. A, B, C) and the other half in minuscule (e.g. a, b, c). We then averaged the ratings to obtain the mean spikeness rating for each letter.

In a different survey, 10 additional sighted participants listened to some of the pseudowords used in the main experiments. These were pseudowords that according to Italian orthography could have been written in more than one way (e.g. /takete/ can be *takete* or *tachete*). Participants were asked to listen each pseudoword and then write it down, in order to confirm empirically the orthographical alternatives.

Based on the rating of OS for each letter we then computed the average orthographic spikeness of each pseudoword in our experiment. When a pseudoword had more than one orthographic alternative, we computed its final OS by averaging the OS of all the possible alternatives (as indicated by the survey). Values of OS were standardized before running statistics to allow a meaningful comparison between odd ratios.

Importantly, our measure of OS did not include the shape of vowels. That is because vowel's OS is highly correlated with their articulatory spikeness: Back vowels graphemes ("o", "u") are more round than front vowels ones ("e", "i"). In our attempt to make orthographic and articulatory spikeness as much orthogonal as possible, including vowels in the count of OS seemed counterproductive and possibly misleading. Yet, the effect of vowels' orthography on sound symbolism was tested in another way: We predicted that the sound symbolic associations between vowels and shape would be greater in sighted than blind. That is because in sighted the effect of articulation might be enhanced by the congruent orthographic spikeness (or roundness) of vowels.

3.2 Results

Initially, we tested the effect of the articulation factors only, both in sighted and blind. "Spiky" and "Round" responses were entered in a Mixed Effect Generalized Linear Model with binomial distribution. Random effects were the intercept of subjects and items, whereas fixed effects were Consonant articulation (5 levels, ordered for manner of articulation and voicing), Vowel articulation (round, unround) and Group (EB, SC). Odd-ratio and Upper/Lower levels of the confidence interval are reported as a measure of effect size.

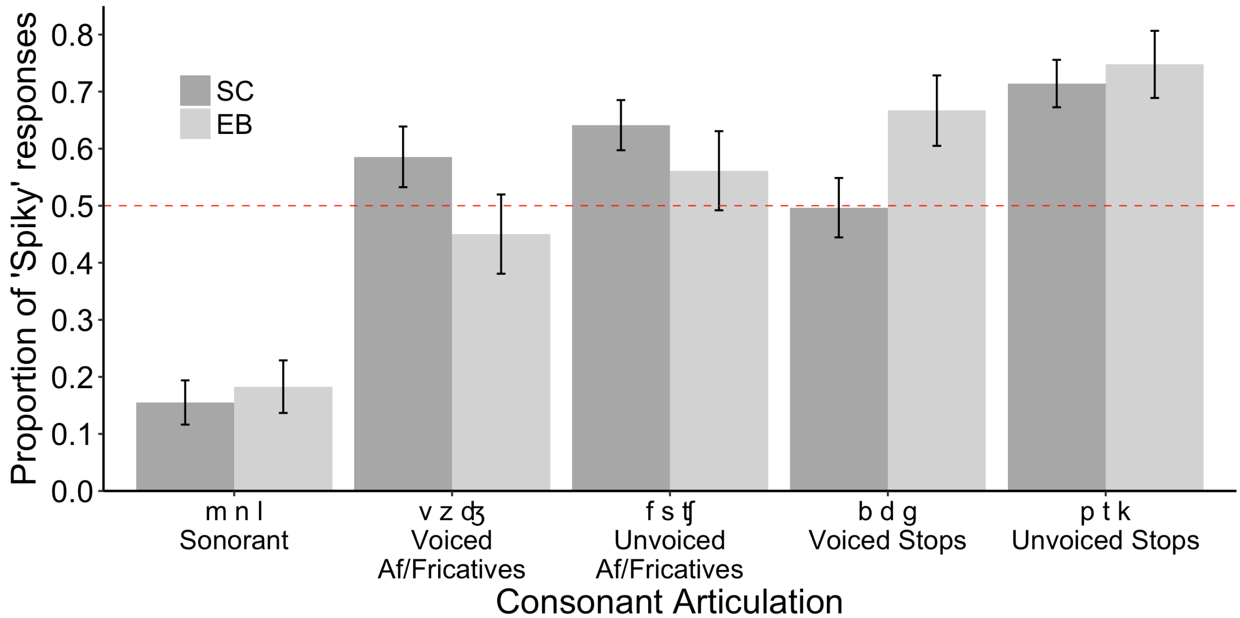

Figure 5. Proportions of ‘spiky’ responses in Experiment 2, arranged by consonant class, for sighted (SC) and blind (EB). The response pattern provided by EB follows monotonically the prediction of the articulatory approach. That is not the case for SC responses, that show a “spikeness” profile very similar to the one reported in McCormick et al.(Mccormick et al., 2015) Dashed line is chance level. Error bars represent ±1 SEM

We found, as expected, a large effect of Consonant articulation (z = 20.80, p< .001, Odd-ratio=3.07, LL=2.76, UL=3.41) and Vowel articulation (z = 13.58, p< .001, Odd-ratio= 3.84, LI= 3.16, HI= 4.66). As shown in Figure 5, sonorants consonants were associated with round responses, stop consonants with spiky responses, and fricative/affricatives were somewhat in the middle. Similarly, as shown in Figure 6, back vowels (/o/, /u/) were associated with round responses and front (/e/, /i/) vowels with spiky responses. The main effect of Group was not significant (z= -1.08, p= .28, Odd-ratio= 0.74, LL= 0.46, UL= 1.24).

As predicted, the interaction between Consonant Articulation and Group was significant ($z = -2.99$, $p = .002$, Odd-ratio= 0.83, LL= 0.74, UL= 1.94). EB followed more closely than SC the prediction of the articulatory theories (see Fig. 5). Indeed, in EB the proportion of spiky responses increased monotonically according to the predicted steps based on manner of articulation and voicing. Such a monotonic increase was not present in the sighted group.

Moreover, we predicted that the effect of sound symbolism for vowels would be stronger in sighted (where the articulatory roundness and spikiness of /o/, /u/, /e/, /i/ is accompanied by a congruent orthographic spikeness). This prediction was confirmed by the presence of an interaction between Vowel articulation and Group ($z = 5.19$, $p < .001$, Odd-ratio= 1.75, LL= 1.41, UL=2.17) (see also Fig 6).

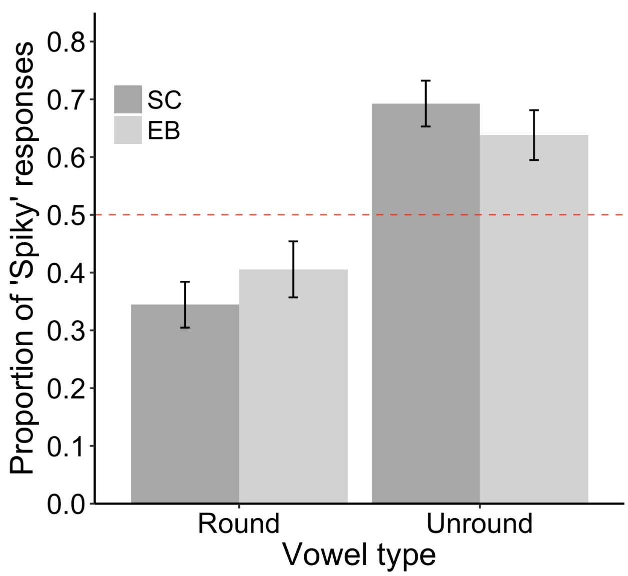

Figure 6. Proportion of ‘spiky’ responses in Experiment 2 aggregated by vowel type. Dashed bar is chance level. Error bars represent ±1 SEM.

Then, to test to what extent the differences between the two groups were due to the influence of orthography we added to the model the factor Orthographic Spikeness. Interestingly, the interaction between Consonant Articulation and Group became non significant (z = -1.25, p= .21, Odd-ratio= 0.92, LL= 0.81, UL= 1.05). On the other hand, the interaction between Orthographic Spikeness and Group was highly significant (z = 5.00, p< .001, Odd-ratio= 1.30, LL= 1.17, UL=1.45), suggesting that the different pattern of sound-symbolic association between sighted and blind was mostly due to the influence of letter shape.

To better articulate this latter interaction we divided the database in sighted and blind groups and we ran two separate models. In the EB group the factor Orthographic Spikeness did not reach significance (z = 0.97, p= .33, Odd ratio= 0.96, LL=.89 UL=1.03). Indeed, compared to a simpler model with Consonant Articulation and Vowel Articulation as predictors, adding the factor Orthographic Spikeness did not provide a better fit of the blind data (Chisq(1) = 0.94, p= .33).

In contrast, in the SC group the factor Orthographic Spikeness was highly significant (z = 3.94, p< .001, Odd ratio= 1.29, LL=1.13 UL=1.46). A regression including Consonant Articulation, Vowel Articulation, and Orthographic Spikeness as predictors provided a better fit than a simpler model without Orthographic Spikeness (Chisq(1) = 15.14, p< .001), suggesting that letter shape was an important factor in determining the judgments of sighted people.

3.3. Discussion

Overall, articulatory spikeness predicted shape-sound associations better in EB than SC. As predicted, sound-symbolic pattern were better explained in the blind by a pure articulatory model and in the sighted by a model that integrates articulatory and orthographic factors. As suggested in the introduction of this experiment, the reduced bouba-kiki effect previously reported in visually impaired people may be due, at least in part, to the role orthography plays in the emergence of this effect. Figure 7 clearly exemplifies this by reporting difference in spikeness ratings between stop unvoiced (e.g. /k/ as in kiki) and stop voiced (e.g. /b/ as in bouba) consonants, for each subject. This difference is normally small in the EB group (as the articulatory approach would predict) and larger in the SC group (as the effect of orthography would predict).

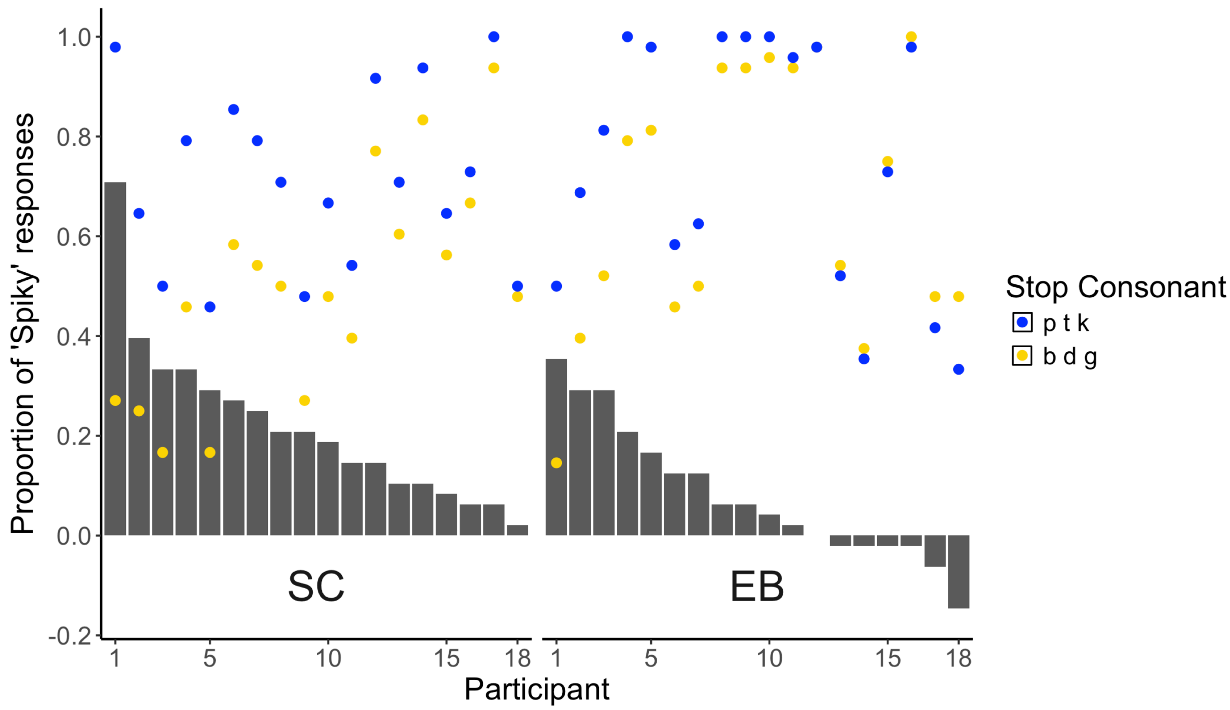

Figure 7. Spikeness difference (bars) between unvoiced (blue dots) and voiced (yellow dots) stop consonants for each participant in the two groups. From left to right, the participants are ordered according to the strength of the effect.

4. General Discussion

Sound-symbolic association between shape and phonemes is a robust cross-cultural phenomenon that defies the arbitrariness of the sign traditionally endorsed by linguistic theories (de Saussure, 1959). The mechanisms of this cross-modal association remain debated. According to the articulatory hypothesis, phonemes are mapped onto the mouth-tongue motor representations needed to produce them, which in turn are mapped onto a shape (Ramachandran & Hubbard, 2001). Yet, across several experiments, shape-phonemes associations deviate more or less greatly from the prediction of a pure articulation account calling for the effect of other factors.

We suggest that previously reported pattern of shape-phoneme association can be better explained by an account that blends articulatory and orthographic information (BOAH). Phonological processing of words activates word's orthographic representations (Chéreau et al., 2007; Ziegler & Ferrand, 1998) that can influence the mapping between phonology and shape.

We tested this hypothesis in a population of early blind participants who are primarily exposed to Braille letters that lack angular and round edges. In line with our predictions we found that shape-sound associations in EB were better explained by a pure articulatory account compared to SC. EB judgments of word spikeness increased monotonically with the change in manner of articulation and voicing. On the other hand, sighted judgments were better predicted by combining the effect of articulation and that of orthography.

This result shed new light on the origin of shape-sound correspondences, supporting the articulatory hypothesis originally proposed by Ramachandran and Hubbard.(Ramachandran & Hubbard, 2001) Deviations from a pure articulatory account reported in previous experiments could be attributed to the effect of orthography.

Our data do not necessarily exclude that other factors (e.g. cultural, phonological; Chen et al., 2016; Knoeferle, Li, Maggioni, & Spence, 2017) may contribute to specify patterns of sound symbolic associations, however, future studies should take into account the orthographic shape of written words (instead of simply dismiss it) while testing for other possible cross-modal mechanisms. Interestingly, the mechanisms providing the iconic ground between manner of articulation, voicing and shape are still unclear. This association may be mediated by the activation of motor representations of speech movements as suggested by Ramachandran and Hubbard (Hubbard & Ramachandran, 2005; Ramachandran & Hubbard, 2001). Yet, a more direct cross-modal mapping between shape and acoustic features (that covary with articulatory movements) may also take place. Alternatively, motor representations may be more active in childhood and being replaced with more direct mapping in adults. Although neuroimaging evidence to date provides some support for the motor-mediated mapping hypothesis (Hubbard & Ramachandran, 2005; Kanero, Imai, Okuda, Okada, & Matsuda, 2014; Ramachandran & Hubbard, 2001; Revill, Namy, DeFife, & Nygaard, 2014), future studies using multivariate analysis with orthogonalized acoustic and motor predictors may help discriminate the contribution of these factors to sound-symbolic correspondences.

Last but not least, our study suggests that early loss of vision does not prevent the development of shape-phoneme sound symbolic associations. This result is in keeping with the articulation approach, and suggests that vision it is not necessary to map representations of speech sounds onto representations of objects' shape. Nevertheless, lack of vision might impair other types of cross-modal correspondences (for instance height-pitch associations; Deroy, Fasiello, Hayward, & Auvray, 2016) that involve different experiential basis. However, by demonstrating a solid bouba-kiki effect in early blind, our results open to the intriguing possibility that EB may be more sensitive to (some) iconic features of language compared to sighted. One of the major problem in language learning is referential ambiguity, meaning the difficulty of attributing uttered words to their referents in highly cluttered environments. One can imagine, for instance, the number of possible referents for a 1 y.o. language-learner when someone says the word "napkin" on a fully set table. Recent evidence suggests that the ambiguity problem may be reduced, from the first year of life, by the frequency distribution of objects in infants' egocentric vision. The statistics of infants' visual experience guide referential processing as that high-frequency objects coincide with the object names normatively learned first (Clerkin, Hart, Rehg, Yu, & Smith, 2017). This and other resources to reduce referential ambiguity such as eye movements and ostensive display (Smith, Colunga, & Yoshida, 2010; Waxman & Gelman, 2009) are not accessible or highly limited to blind children. In these conditions, sound symbolic cues may play an important role in solving referential ambiguity (Imai & Kita, 2014) scaffolding form-to-meaning mapping and word learning.

Acknowledgements

This work was supported by a European Research Council starting grant (MADVIS grant #337573) attributed to OC. We are also extremely thankful to our participants and the Unione Ciechi e Ipovedenti in Milano and Trento, Savona, Trieste and Cuneo.

References

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

Aveyard, M. E. (2012). Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory & Cognition*, *40*(1), 83–92. http://doi.org/10.3758/s13421-011-0139-3

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(December), 1–2. http://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *Eprint arXiv:1406.5823*, 51. http://doi.org/10.18637/jss.v067.i01

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

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). " Bouba" and " Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172. http://doi.org/10.1016/j.cognition.2012.09.007

Champoux, F., Collignon, O., Bacon, B. a, Lepore, F., Zatorre, R. J., & Théoret, H. (2011). Early- and late-onset blindness both curb audiotactile integration on the parchment-skin illusion. *Psychological Science*, *22*(1), 19–25. http://doi.org/10.1177/0956797610391099

Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When "Bouba" equals "Kiki": Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, *6*(1), 26681. http://doi.org/10.1038/srep26681

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition*, *102*(3), 341–360. http://doi.org/10.1016/j.cognition.2006.01.001

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. http://doi.org/10.1098/rstb.2016.0055

Collignon, O., Charbonneau, G. G. G., Lassonde, M., & Lepore, F. (2009). Early visual deprivation alters multisensory processing in peripersonal space. *Neuropsychologia*, *47*(14), 3236–3243. http://doi.org/10.1016/j.neuropsychologia.2009.07.025

Cuskley, C., Simner, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba – kiki effect. *Psychological Research*, *81*(1), 119–130. http://doi.org/10.1007/s00426-015-0709-2

D'Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. *Language and Speech*, *57*(3), 367–393. http://doi.org/10.1177/0023830913507694 de Saussure, F. (1959). *Course of general linguistics*. New York: Philosophical Library.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D. C., & Orban, G. A. (2004). The Processing of Visual Shape in the Cerebral Cortex of Human and Nonhuman Primates: A Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *24*(10), 2551–2565. http://doi.org/10.1523/JNEUROSCI.3569-03.2004

Deroy, O., Fasiello, I., Hayward, V., & Auvray, M. (2016). Differentiated audio-tactile correspondences in sighted and blind individuals. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1204–1214. http://doi.org/10.1037/xhp0000152

Deroy, O., & Spence, C. (2013). Are we all born synaesthetic? Examining the neonatal synaesthesia hypothesis. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1240–1253. http://doi.org/10.1016/j.neubiorev.2013.04.001

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. http://doi.org/10.1016/j.tics.2015.07.013

Doyle, J. R., & Bottomley, P. A. (2011). Mixed Messages in Brand Names: Separating the Impacts of Letter Shape from Sound Symbolism. *Psychology & Marketing*, *28*(7), 749–762. http://doi.org/10.1002/mar

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough : How visual experience influences haptic – auditory associations in the "' Bouba – Kiki '" effect. *Cognition*, *132*(2), 164–173. http://doi.org/10.1016/j.cognition.2014.03.015

Hubbard, E. M., & Ramachandran, V. S. (2005). Neurocognitive mechanisms of synesthesia. *Neuron*, *48*(3), 509–520. http://doi.org/10.1016/j.neuron.2005.10.012

Hung, S., Styles, S. J., & Hsieh, P. (2017). Can a Word Sound Like a Shape Before You Have Seen It ? Sound-Shape Mapping Prior to Conscious Awareness. *Psychological Science*, *28*(3), 263–275. http://doi.org/10.1177/0956797616677313

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130298-. http://doi.org/10.1098/rstb.2013.0298

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

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE*, *10*(2), 1–17. http://doi.org/10.1371/journal.pone.0116494

JASP Team. (2017). JASP (Version 0.8.1.2)[Computer software].

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS ONE*, *9*(5), 1–8. http://doi.org/10.1371/journal.pone.0097905

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*(3), 575–586.

http://doi.org/10.1111/j.1551-6709.2010.01169.x

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*(1), 5562. http://doi.org/10.1038/s41598-017-05965-y

Kohler, W. (1947). *Gestalt psychology (2nd ed.)*. New York, NY: Liveright Publishing.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. http://doi.org/10.1016/j.cognition.2009.08.016

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 1–14. http://doi.org/10.3389/fpsyg.2015.01624

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-Symbolism Boosts Novel Word Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1274–1281. http://doi.org/10.1037/a0028642

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. http://doi.org/10.1016/j.jml.2017.01.001

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322.

http://doi.org/10.1111/j.1467-7687.2006.00495.x

Mccormick, K., Kim, J. Y., List, S., & Nygaard, L. C. (2015). Sound to Meaning Mappings in the Bouba - Kiki Effect. *CogSci 2015*, 1565–1570.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(5), 1152–64. http://doi.org/10.1037/a0027747

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/Revue Canadienne de Psychologie Expérimentale*, *65*(2), 115–124. http://doi.org/10.1037/a0022268

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186. http://doi.org/10.1016/j.cognition.2009.04.001

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The takete-maluma phenomenon in autism spectrum disorders. *Perception*, *42*(2), 233–241. http://doi.org/10.1068/p7357

Ozturk, O., Krehm, M., & Vouloumanos, A. (2012). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. http://doi.org/10.1016/j.jecp.2012.05.004

Peelen, M. V, He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(1), 163–70. http://doi.org/10.1523/JNEUROSCI.1114-13.2014

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

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. C., Cohen, L., … Haxby, J. V. (2004). Beyond sensory images : Object-based representation in the human ventral pathway. *Pnas*, *101*(15), 5658–5663.

Putzar, L., Goerendt, I., Lange, K., Rösler, F., & Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nature Neuroscience*, *10*(10), 1243–1245. http://doi.org/10.1038/nn1978

R Core Team. (2013). R: A language and environment for statistical computing.

Ramachandran, V. S., & Hubbard, E. H. (2001). Synaesthesia—A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34. http://doi.org/10.1111/1468-0068.00363

Revill, K. P., Namy, L. L., DeFife, L. C., & Nygaard, L. C. (2014). Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language*, *128*(1), 18–24. http://doi.org/10.1016/j.bandl.2013.11.002

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as Process: Contextually Cued Attention and Early Word Learning. *Cognitive Science*, *34*(7), 1287–1314.

http://doi.org/10.1111/j.1551-6709.2010.01130.x

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. http://doi.org/10.3758/s13414-010-0073-7

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. http://doi.org/10.1016/j.tics.2009.03.006

Westbury, C. (2005). Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language*, *93*(1), 10–19. http://doi.org/10.1016/j.bandl.2004.07.006

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, *5*(4), 683–689. http://doi.org/10.3758/BF03208845