**Logo, company name

Description automatically generated**

**IRP**

**Title:** Facilitatory Effect of Unimodal Sound Symbolism in the Naming of Novel Objects

**Author:** Muthukumar Arjun Kumar

**Abstract**

Systematic mapping between sound and meaning have been well established. However, most of these studies have focused on cross-modal mapping between sound and other sensory domains. Onomatopoeia or unimodal mapping of sound hasn’t been widely investigated. This study provides novel evidence that unimodally mapping formant shifts in a nonverbal sound to phonemes in speech has a facilitatory effect on naming. By using animal sounds, this study also provides behavioural evidence for neuroimaging studies in the past that have shown that onomatopoetic words are processed by brain regions that process both nonverbal (animal sounds) and verbal sounds (nouns). The implications of this study don’t confine to unimodal sound symbolism, but extends to sounds symbolism as a whole.

**Introduction**

Numerous studies suggest that there are systematic sound to meaning association cross-linguistically. These findings have been labelled under the general term of ‘sound symbolism’ or ‘iconicity’ (Lockwood & Dingemanse, 2015). This directly challenges the dictum of arbitrariness in language that was the prevalent view for much of the last century (Magnus, 2000). However, the focus has largely been on cross-modal mapping between a particular dimension of the referent and sound (e.g. shape-sound, size-sound, taste-sound; for review Lockwood & Dingemanse, 2015). Onomatopoeia, which is unimodal, and a direct mapping between sound to sound, has largely been ignored. And even in the case of onomatopoeia, studies have not focused on specific elements of the sound that is mapped but have rather focused on onomatopoeia from the perspective of language acquisition (e.g. Laing, 2014, 2017, 2019; Motamedi, et al., 2020). As such, the goal for this study was to investigate if pitch related information can be unimodally mapped to a referent.

**Cross-Modal Mapping to Sound**

The age of the Saussurean dictum of arbitrariness (1916) is over. Symbolic mapping between sound and meaning are now being accepted increasingly as a general part of language (Perniss et al., 2010; Perniss & Vigliocco, 2014). Most studies on sound symbolism have so far focused on cross-modal mapping between the referent and its sound. Among these cross-modal dimensions, shape and sound have received the greatest attention. The exploration of shape-sound symbolism was initiated by Kohler (1929) when he observed that participants consistently chose ‘takete’ for objects with a spiky shape and ‘maluma’ for objects with a rounded shape . Since then, a large body of research has found similar shape-sound associations (Irwin & Newland, 1940; Kim, 1977; Ramachandran & Hubbard, 2001; Styles & Gawne, 2017). The fact that it holds even for infants (Ozturk et al., 2012) and neonates (Peña et al., 2011) seems to suggest that these associations, and sound symbolism in general, could be inherent and not just learned. Size-sound symbolism was first shown when Sapir (1929) demonstrated that objects with high front vowel names tend to be judged as smaller whereas those with large back vowel tend to be judged as larger. This association has since then been shown to hold for speakers of various languages [e.g. English (Ohala, 1994), Korean, Japanese and mandarin (Shinohara & Kawahara, 2016), Russian and Cantonese (Shih, et al., 2019)]. Apart from the shape and size, sound symbolism has also been extended to other varied cross-modal perceptual dimensions such as brightness (Hirata et al., 2011), taste (Simner et al., 2010), color (Moos et al., 2014) and motion (Cusky, 2013).

**Unimodal Mapping to Sound - Onomatopoeia**

Onomatopoeia are non-arbitrary words that imitate the sound of the referent (e.g. beep, knock, click). In focusing on cross-modal sound symbolic associations, onomatopoeia has largely been overlooked. The scarce but existing studies on onomatopoeia have largely focused on it in the context of language acquisition among infants (e.g. Laing, 2014, 2017, 2019; Motamedi, et al., 2020). This is surprising because of three reasons. Firstly, it is the most simplest and direct type of mapping between form and meaning (sound to sound) (Motamedi, et al., 2020). This is opposed to cross-modal mapping which involves perceptual elements from other sensory dimensions to sound, like shape-sound, size-sound, taste-sound etc. Secondly, onomatopoeia is among the most common words that are produced by young children (Laing, 2014). Tardif et al. (2008) showed that 29.5% of the first ten words spoken are onomatopoetic for American English and 40.6 % for Cantonese according to a parental vocabulary questionnaires for infants aged 8–16 months. Thirdly, and particularly relevant to this investigation, onomatopoeia provides a special avenue for research as it can show how acoustic elements of sound could be mapped to phonemes through imitation. This is reflected at the neural level as well, onomatopoeia words were processed by brain regions involved in both verbal (noun) and nonverbal (animal) sounds (Hashimoto, et al., 2006), leading the authors to posit that it could provide a bridge between nouns and animal sounds. However, the exact mechanism of this connection is unknown as the imitation of specific acoustic elements of sound in onomatopoetic mapping remains largely uninvestigated. Assaneo et al. (2011) proposed the definition that onomatopoeia is produced through the transformation of sounds into speech elements that minimize their spectral difference within the constraints of the vocal system. This was verified through computational modelling and perceptual evidence based on how ‘knock’ and ‘click’ sounds match to phonemes. However, it remains uninvestigated on whether such mapping could also occur between formants. Formants are an important sound element that refer to the spectral maximum resulting from resonances in vocal apparatus. Humans have been shown to be sensitive to changing formant patterns. For example, the study by Myers-Schulz et al. (2013) found that participants chose downward formant shifting names for negative valanced images more so than chance compared to upward formant shifting names. Given the evidence that arbitrary sounds like door knocks and beeps could be mapped to phonemic speech elements based on spectral similarity, it seems possible that elements like formants (which are direct speech elements) could also be mapped to phonemes.

**Present Study**

The present study decided to investigate this using a naming task based on the bouba/kiki paradigm (Ramachandran & Hubbard, 2001). Originally developed by Kohler (1929), the paradigm involves a two-alternative forced choice task. This paradigm has been widely used to investigate if there is symbolic association between a sound and its referent (Lockwood & Dingemanse, 2015). By using latency of response for congruent and incongruent conditions based on the participant’s choice of nonword, the effect of unimodal mapping was measured. This will be further elaborated in the methods section. The hypothesis for this study is that unimodal mapping would lead to a facilitatory effect on naming. More specifically, the latency of response for congruent responses would be lesser than incongruent responses. This hypothesis is predicated on three main findings. Firstly and as noted previously, humans have been shown to be sensitive to the direction of formant shifts on naming tasks (Myers-Schulz et al., 2013). Secondly, vocal imitation functions by maximising spectral similarity between the perceived and produced sound (Assaneo et al., 2011). Though Assaneo et al. did not specifically test this for formant shift, it is likely that this rule would still apply as it is an important speech element. Lastly, onomatopoeic word perception was suggested to have neural links to both non-verbal (animal sound) perception and verbal (noun) perception (Hashimoto, et al., 2006). This neural link could possibly provide a facilitatory effect in the naming task.

**Methods**

**Design**

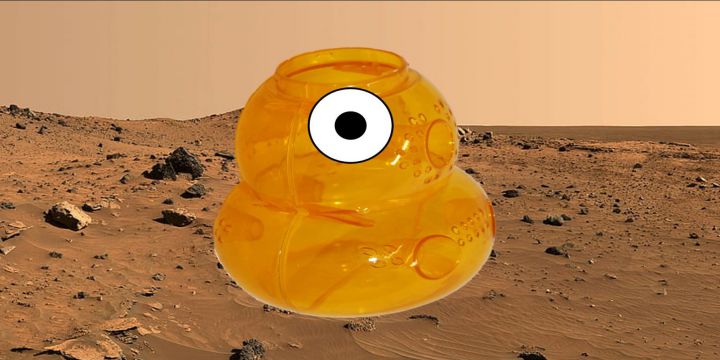
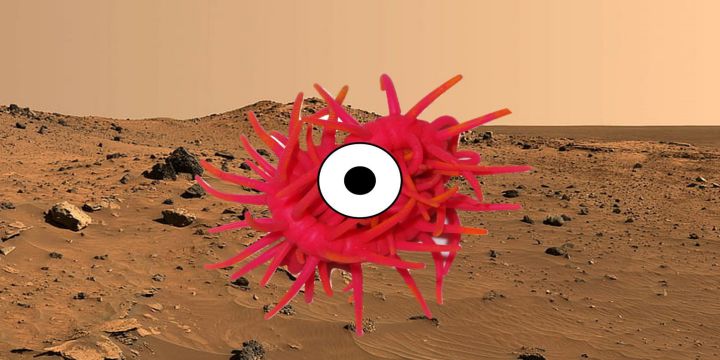
Participants in the study were asked to view an object along with the sound it was said to make. The sound either had an upward shifting or a downward shifting formant pattern. After hearing the sound, they were asked to choose a name for the object from two choices. One of the names had a downward shifting formant pattern at the level of phonemes and the other had an upward shifting formant pattern. Each of their response were then coded as congruent or incongruent, under the ‘Match’ variable. By nature of the design, this was a repeated measure. The latency of response was measured for each trial. As such, the IV in the study was ‘match,’ which had two levels – congruent and incongruent. e.g. match would be congruent when choice of word for the object shown is downward formant shifting in response to a downward formant shifting sound. The DV was latency, which denoted the time taken to respond in each trial.

**Participants**

Healthy English-speaking adults (n=101) between the ages of 18 and 28 (mean = 20.26 ± 1.79) without any hearing and vision problems. All of them were university students who participated in the study as part of their course requirements. They were rewarded with 1 participation point after successful completion. Participants agreed to participate in this study voluntarily.

**Stimuli**

Novel objects: 32 objects were selected from the Novel Object and Unusual Names (NOUN) Database (Horst & Hout, 2015). The objects of this database have been experimentally verified for their novelty. These objects were chosen so that previous naming judgement doesn’t affect naming response in this study. I added simple eyes to the objects to make them animate as they were later told these were Martian aliens discovered as part of a mission. The image of a realistic Mars background was added to make the study more naturalistic (Figure 1).



**Figure 1**. Example novel objects.

*Figure 1. Novel objects*

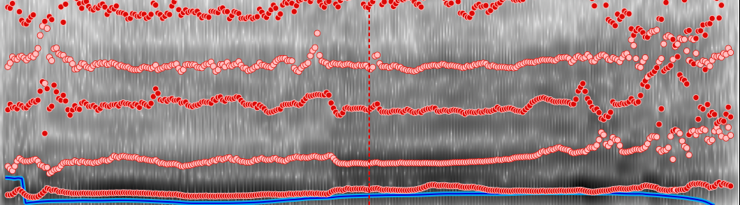
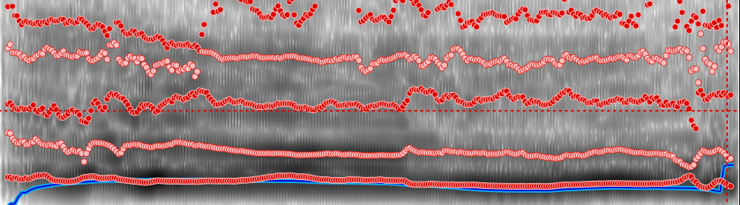
**A**

Audio: A royalty free audio of red deer roar was obtained from freesound.org. The sound was cleaned using Audacity (Version 3.0.0) and clipped to 3 s. I decided to use an animal sound so as to connect to past research by Hashimoto et al. (2006) which found neural links between onomatopoetic words and animal sounds. There were two main types of audio stimuli, upward formant shifting pattern and downward formant shifting pattern. Red deer roar has a very clear downward shifting formant pattern (Fitch & Reby, 2001). I reversed this in audacity to get the upward formant shifting pattern (Figure 2). Additionally, I manipulated the frequency of the audio by increasing and decreasing in steps of 5% of the original sound for both the audio files until there were 16 files in each of the audio category. This was mainly to make the study more generalizable across a range of frequencies.

**A**

**B**

**Figure 2**. Formants in a deer’s roar – a) original downward shifting sound: downward shifts in the formants. B) reversed audio: upward shifting formants.



A picture containing text, athletic game, sport

Description automatically generatedA picture containing text, athletic game, sport

Description automatically generatedNon-Word List: Non-word list created by Myers-Schulz et al. (2013) was used. 32 pairs of nonwords were obtained. Each pair contains one word with upward formant shifting phonemes and the other has downward formant shifting phonemes (Figure 3). These non-words have already been matched for various acoustic and speech featured (plosives, fricatives, nasals, vowels, number of syllables, number of phonemes, number of graphemes).

**B**

**A**

**Figure 3**. Formant shifts in nonwords at the level of phonemes – a) dugada: downward shifting phonemes, b) bupaba: upward shifting phonemes (source: Myers-Schulz et al, 2013)

**Apparatus**

The study was designed using jsPsych (de Leeuw, 2015), a JavaScript library for psychological experiments, and conducted online through JATOS deployed on a Mindprobe server. Participants were asked to wear headphones and do the experiment in a quiet environment. They also went through a headphone screening task to ensure that they wore their headphones (Milne, Bianco, Poole et al.)

**Procedure**

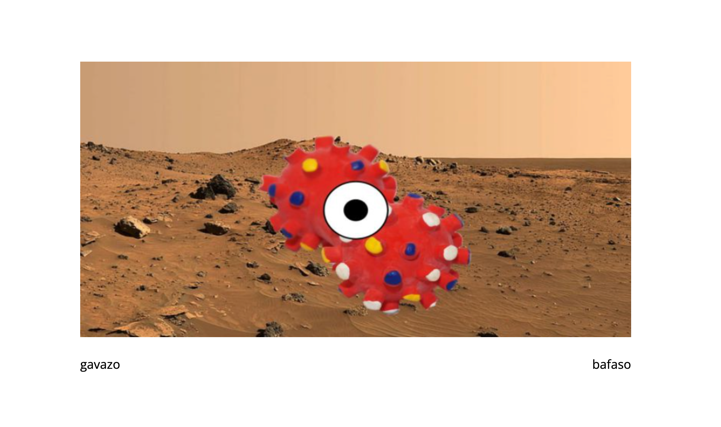
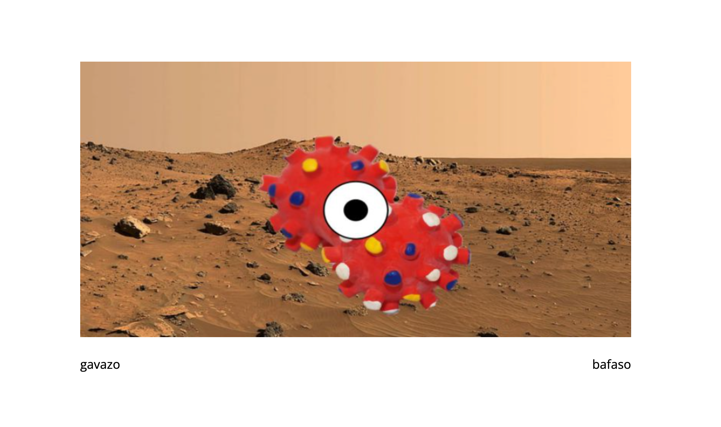
The study consisted of 32 trials in total which were completely randomised. Before the trials began, participants were told they were part of a space exploration mission and they were going to choose names for Martian aliens. Participants were also asked to put on their earphone/headphones and perform the experiment in a quiet environment. Each trial consisted of an initial ‘READY’ screen followed by the simultaneous presentation of the image along with the audio stimulus which was either upward or downward shifting red deer vocalisation (3 s). After 3 seconds, when the audio finished playing, two non-words were shown at either ends of the bottom of the screen, one upward shifting and one downward shifting (Figure 4). The

**Figure 4**. Presentation of trials. The novel object along with the deer’s roar was presented for 3000 ms. This was followed by the two alternatives forced naming task where latency was measured.



3000 s

latency



positions of the words were balanced and randomised during presentation as well. Participants were then asked to choose the word using keypress and the latency of response was measured .

**Results**

The online nature of the experiment provided the study with very little experimental control. Therefore filtering of the data became necessary. All responses less than the 10th percentile of response – 350ms and those after the 90th percentile – 6024ms were removed. After this cleaning process, there were 2610 trials (from 3264) from 99 participants. The distribution of latency was checked for normality using the Shapiro-Wilk test, the result were non-significant (p = 0.149) indicating that there wasn’t a significant deviation from normality. A paired sample student t-test was conducted between the latency for congruent responses and incongruent responses. There was a significant difference between the latency for congruent response (M = 2148.42 ms, SD = 943.33 ms) and incongruent response (M = 2214.13 ms, SD = 1001.24 ms); t (98) = -1.79, p = 0.078 (Figure 5).

**Figure 5**. descriptive plot of means and their 95% confidence intervals

Chart, line chart, box and whisker chart

Description automatically generated

**Discussion**

The hypothesis of this study was that naming would be facilitated when formant shifts in sound is unimodally mapped to the phonemes in a referent’s name. This was measured through latency of responses for congruent vs incongruent mapping. This hypothesis was supported by the results – the latency of unimodally congruent response was lower than that for incongruent response. For example, when participants chose a nonword with downward shifting formants for a referent after listening to a downward formant shifting animal sound, response was faster when compared to the contrary. This clearly suggests a facilitatory effect of unimodal mapping between the formant shift of the animal sound to the formant shift of the phonemes in the nonword choice. Our findings on the facilitatory effect seen on latency is in support of the findings by Assaneo et al. (2011) and Hashimoto et al. (2006). Assaneo et al. suggested that we tend to imitate sounds to vocally produce a sound with the greatest spectral similarity. In this case, it is the direction of formant shifts. When participants chose a congruent response, the closer spectral similarity when compared to an incongruent response could have facilitated their response and thus decreased their latency. It should also be noted that though the individual phonemes in the words are downward shifting, it could be facilitated with a sound that is globally downward shifting (shifting across its entire time period). This points to complex ways in which information could be unimodally mapped. This aspect needs further investigation. As for the neuroimaging study by Hashimoto et al. (2006), they found that brain regions involved in perception of animal sounds as well as those involved in the perception of verbal sounds (nouns) are simultaneously activated when an onomatopoetic word is heard. Our findings thus provide preliminary behavioural evidence for this neural findings. The authors had posited in their study that onomatopoetic words might act as a bridge between nouns and animal sounds. This suggested neural ‘bridge’ could have facilitated response when it was congruent, thus leading to a faster response. At a wider level, this study adds on to the synesthetic origins of language hypothesis (Ramachandran & Hubbard, 2001) which has predominantly focused on cross-modal associations. This study shows that unimodal sound-sound associations of specific acoustic elements are also possible.

There is one main limitations to this study. To simulate upward and downward formant shifting sounds, the originally downward shifting roar of a red deer was directly reversed. However, it is unsure if a simple sound reversal as such could be still possible during the mechanic production of sound. As such this could have introduced some unexpected confounds by virtue of the upward shifting sound being possibly ‘unnatural’. Future studies should test how these results would be affected if the number of downward shifting and upward shifting phonemes are varied in the nonwords. The study by Thompson & Estes (2011) had found a graded association between number of back phonemes and size of a nonobject. Would a similarly graded relationship be observed here in the case of unimodal mapping of formant shift? Future studies should also investigate the effect of gender on latency. A study by Baron-Cohen et al. (1996) found that synaesthesia was more prevalent among women as compared to men at a 6:1 ratio. As such would their latency of for congruent responses be lesser than what it is for men?

**Conclusion**

The goal of this paper was to find if formant shift related information could be unimodally mapped to a referent’s name. I found preliminary support for this as it is observed through a facilitatory effect of unimodally congruent response in the latency of response. This extends and supports the finding by Assaneo et al. (2011) who suggested that vocal imitation aims to produce a sound that is as spectrally close to the sound perceived. However, it is unsure how a general pattern of the sound could be mapped to the level of individual phonemes of the non-word. This facilitatory effect on latency also provides behavioural evidence to the neuroimaging study by Hashimoto et al. (2006) who posited that onomatopoetic words could be the bridge between nouns and animal sounds. This study also adds to the synesthetic origins of language hypothesis, which has mainly focused on cross-modal associations, by showing that unimodal sound-sound associations of specific acoustic elements are also possible.

/number of pages = 10 (not including cover page and abstract)

References

Assaneo, M. F., Nichols, J. I., & Trevisan, M. A. (2011). The anatomy of onomatopoeia. PloS one, 6(12). doi:https://doi.org/10.1371/journal.pone.0028317

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J., & Bolton, P. (1996). Synaesthesia: Prevalence and Familiality. Perception, 25(9), 1073–1079. doi:https://doi.org/10.1068/p251073

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code. A perceptual investigation. Phonetica(65(4)).

Cuskley, C. (2013). Mappings between linguistic sound and motion. Public J. Semiot. 5, , 39–62. doi:10.37693/pjos.2013.5.9651

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. [doi:10.3758/s13428-014-0458-y](https://doi.org/10.3758/s13428-014-0458-y).

de Saussure, F. (1916). Cours de linguistique générale. Payot.

Fitch, W. T. (2000). The evolution of speech: a comparative review. Trends in cognitive sciences, 4(7) , 258–267. doi:https://doi.org/10.1016/s1364-6613(00)01494-7

Fitch, W., & Reby, D. (2001). The descended larynx is not uniquely human. Proceedings. Biological sciences / The Royal Society, 268. doi:10.1098/rspb.2001.1704.

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & al., e. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. Neuro Image, 31, 1762–70.

Hinton, L., Nichols, J., & Ohala, J. (1994). Sound symbolism. doi:https://doi.org/10.1017/CBO9780511751806

Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner’s speeded classification task. Percept. Mot. Skills , 113, 929–940. doi:10.2466/15.21.28.PMS.113.6.929-940

Hockett, C. F. (1960). The origin of speech. Scientific American, 203.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. Behavior research methods, 48(4),, 1393–1409. doi:https://doi.org/10.3758/s13428-015-0647-3

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Phil. Trans. R. Soc. B. doi:http://dx.doi.org/10.1098/rstb.2013.0298

Irwin, F. W., & Newland, E. (1940). A genetic study of the naming of visual figures. J. Psychol. , 9, 3–16. doi:10.1080/00223980.1940.9917674

Kim, K.-O. (1977). Sound symbolism in Korean. J. Linguist, 13, 67–75. doi:10.1017/

Köhler, W. (1929). Gestalt psychology. Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. Cognition , 114, 19–28. Retrieved from 10.1016/j.cognition.2009.08.016

Laing, C. (2019). A role for onomatopoeia in early language: Evidence from phonological development . Language and Cognition,, 11(2), 173-187. doi:10.1017/langcog.2018.23

Laing, C. E. (2014). A phonological analysis of onomatopoeia in early word production. First Language, 34(5), 387–405. doi:https://doi. org/10.1177/0142723714550110

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. Journal of experimental child psychology(161), 32–45. doi:https://doi.org/10.1016/j.jecp.2017.03.017

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

Magnus, M. (2000). What’s in a word?: Studies in phonosemantics. .

Milne, A.E., Bianco, R., Poole, K.C. *et al.* An online headphone screening test based on dichotic pitch. *Behav Res* **53,**1551–1562 (2021). https://doi.org/10.3758/s13428-020-01514-0

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder. i-Perception , 5, 132–142. doi:10.1068/i0626

Morton, E. (1977). On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds . The American Naturalist, 111(981), 855-869.

Motamedi, Yasamin, Murgiano, Margherita, Perniss, P., Wonnacott, E., . . . Gabriella. (2020). Linking language to sensory experience: Onomatopoeia in early language development. Developmental Science. doi:10.1111/desc.13066

Myers-Schulz, B., Pujara, M., Wolf, R. C., & Koenigs, M. (2013). Inherent emotional quality of human speech sounds. Cognition & emotion, 27(6), 1105–1113. doi:https://doi.org/10.1080/02699931.2012.754739

Newmeyer, F. J. (1992). Iconicity and generative grammar. Language,, 68(4), 756–796. doi:https://doi.org/10.2307/416852

Nuckolls, J. (1999). The Case for Sound Symbolism. Annual Review of Anthropology - ANNU REV ANTHROPOL(28), 225-252. doi:10.1146/annurev.anthro.28.1.225.

Ohala, J. J. (1994). The frequency codes underlies the sound symbolic use of voice pitch,” in Sound symbolism, eds L. Hinton, J. Nichols, and J. Ohala. 325–347. doi:10.1017/cbo9780511751806.022

Ozturk, O., Krehm, M., & Vouloumanos, A. (2012). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month olds. J. Exp. Child Psychol.(114), 173–186. doi:10.1016/j.jecp.2012.05.004

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. Psychol. Sci. , 22, 1419–1421. doi:10.1177/ 0956797611421791

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. Philos. Trans. R. Soc. B Biol. Sci.(369). doi:doi: 10.1098/rstb.2013.0300

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

Ramachandran, S., & Hubbard, E. M. (2001). Synaesthesia — A window into perception, thought and language. J. Consc. Stud. , 8, 3–34.

Sapir, E. (1929). A study in phonetic symbolism. J. Exp. Psychol. 12, 225–239. doi:10.1037/h0070931

Shih, S. S., Ackerman, J., Hermalin, N., Inkelas, S., Jang, H., Johnson, J., & al., e. ( 2019). Cross-linguistic and language-specific sound symbolism: Pokémonastics. Retrieved 10 4, 2021, from https://ling.auf.net/lingbuzz/004725

Shinohara, K., & Kawahara, S. (2016). A cross-linguistic study of sound symbolism: The images of size. Proceedings of the 36th Annual Meeting of the Berkeley Linguistics Society(396). doi:10.3765/bls.v36i1.3926

Shinohara, K., Kawahara, S., & Tanaka, H. (2020). Visual and Proprioceptive Perceptions Evoke Motion-Sound Symbolism: Different Acceleration Profiles Are Associated With Different Types of Consonants. Frontiers in psychology, 11. doi:https://doi.org/10.3389/fpsyg.2020.589797

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Crossmodal mappings across gustation and audition. Perception , 39, 553–569. doi:10.1068/p6591

Styles, S. J., & Gawne, L. (2017). When Does Maluma/Takete Fail? Two Key Failures and a Meta-Analysis Suggest That Phonology and Phonotactics Matter. i-Perception, 8(4). doi:https://doi.org/10.1177/20416695177248072041669517724807

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby’s first 10 words. Developmental Psychology, 44, 929–938.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. Quarterly journal of experimental psychology , 64(12), 2392–2404. doi:https://doi.org/10.1080/17470218.2011.605898

Võ, M. L., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. . Journal of vision,, 9(3), 1–15. doi:https://doi.org/10.1167/9.3.24