Sounds Full of Meaning and the Evolution of Language[1]

Susanne Fuchs and Aleksandra Ćwiek

Imagine that you are a child again and smell fresh-baked cake when you come home after school. Maybe it's your grandmother's apple cake or your neighbor's famous cheesecake. You audibly inhale through the nose and your eyes roll because the cake smells delicious. You open the door. There is a piece waiting for you right there, still warm and fresh from the oven! You take a bite. Now, with this memory in place, how could you let someone experience it with you? You could try describing the taste with words. Maybe the cake is mild, maybe zesty, or maybe it's just delightful! However, it might be difficult to describe a sensory experience like that using conventional language.

A different way to describe the experience would be to use depictive rather than descriptive communication. You might do a breathy grunt followed by a long /m/. There can be yumminess all over your body, too: hands touching the smacking lips, eyes and eyebrows frowned in the sense of pleasure. An m-sound might be meaningless without the additional information of the situation, but producing "mmm" with the smell of one's favorite cake in mind clearly delivers the meaning of pleasure.

The intent of this article is to show that speech sounds can be much more than mere meaning-distinguishing units. Through established cross-modal correspondences with other sensory dimensions, human vocalizations can bear meaning that translates to a real-world context. We argue that cross-modal correspondences and the iconic resemblance between the audible form of spoken language and other sensory information create meaning and were essential to get language off the ground at its dawn. In this sense, the world of sounds can be full of meaning.

The World of Sounds

Traditionally, a speech sound (or phoneme in linguistic terms) is the smallest meaning-distinguishing unit of speech. For example, the difference between "hit"/hɪt/ and "hat" /hæt/ is but one vowel sound. Sounds themselves are considered to have no meaning, and they are defined by the conventions of a language. English "hat" and Spanish "sombrero" both refer to an object covering the head but include different sounds. However, some sounds are very stable across languages.

We can observe astonishing examples of sounds bearing meaning in sound symbolism. But what does this mean? Let us look at a different example of a minimal pair than the two words hit and hat. Compare "zig"/zɪg/ and "zag"/zæg/. The difference in vowels is the same as in hit versus hat, but what about the difference in meaning? "Zigzag" paints a picture in our heads of a back-and-forth movement. In this example, /ɪ/ versus /æ/ evoke a feeling of an opposite direction of movement. The word zigzag is iconic; it creates a mapping between aspects of the acoustic signal and features of the action or visual image.

The most comprehensive study on sound symbolism that we are aware of is the one published by Blasi et al. (2016), who investigated almost 4,300 languages. These are about two-thirds of all existing languages! The authors used a list of 40 words from the Swadesh (1955) list, which encompasses a total of 100 concepts that are least likely to be borrowed from other languages. The Swadesh list was created to compare vocabularies cross-linguistically, aiming toward a better understanding of concept stability and change across language histories. Those concepts include body parts (e.g., eye, lips, breast), pronouns (e.g., I, we, you), and motion verbs (e.g., swim, walk), among others.

Blasi et al. (2016) analyzed the relationship between the occurrence or avoidance of sounds within those concepts

1 This paper is dedicated to Mary Wünsch.

https://doi.org/10.1121/AT.2022.18.2.43

across languages. Several consistent associations between sounds and specific meanings were found. One example is an association between the vowel /i/ and the concept of "small," such as French "petite," or Maori "iti." In the following, we discuss possible reasons for further research on the sound-to-size mapping. Another example is the correlation of /m/ and /u/ with the concept of "breast." Both sounds engage the lips during articulation. For /u/, the lips are protruded and for /m/ they are closed. This use of lips has been discussed in terms of a direct relationship to sucking and breastfeeding in babies, therefore aligning with the meaning of breast. We later go into more detail and explore the evidence for sounds to create and, ultimately, bear meaning.

Cross-Modal Correspondence

Sounds can become meaningful when they are fused with other sensory information such as visual shape or touch. The product of this "fusion" is called a cross-modal correspondence. Everyday human life is full of cross-modal experiences. We perceive the world around us through all senses: smell, taste, touch, sight, and sound. Remember the example from the beginning, your favorite cake coming out of the oven. You may recollect the good smell in the air, the delicious taste, the texture in the mouth, and the mmm-sound. For sure, this cake looks good waiting for you on the table!

Sound symbolism is only one specific case of cross-modal correspondence. It is probably more common than we realize. For example, all diaper brand names in Japan and most in Germany include a bilabial consonant, a consonant that is produced using both lips, because we connect bilabial sounds with babies, who use their lips to suck milk and make the first sounds. A whole branch of marketing research deals with designing a perfect brand or product name according to what the product is, how it is used, and the target group. So, in supermarkets, our attention may be caught by a product name that is imposed on specific properties of the product and designed for a specific target group. The vowel /a/ can be correlated with dark beer rather than light beer, and women generally respond more favorably to products containing front vowels like /e/ or /i/ (Klink, 2000; 2009).

Examples of sound symbolism can also be found in fantasy names, such as Pokémon. And so, the strength or size of the Pokémon can be correlated with such things as the name length or the number of certain consonants. Here again, one sensory modality, such as the visual perception of size, might be reflected in the sound of its name.

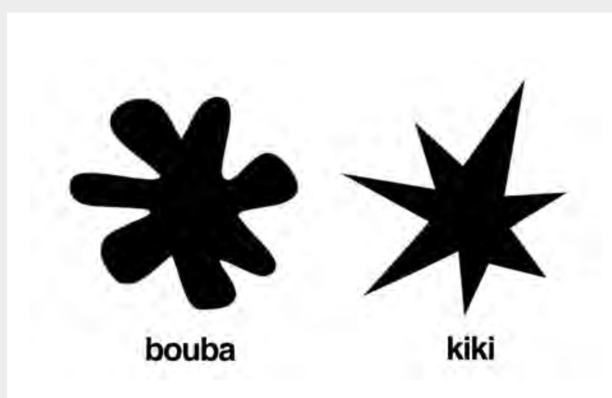
We now introduce a few examples of the fascinating world of cross-modal correspondences (for a review, see Spence, 2011). We picked some of the very popular ones and some that may be relatively less known. We do not wish to imply that every human will perceive these cross-modal correspondences in the same way. It must be borne in mind that sound-meaning relationships can also be specific for a particular language. However, here, we specifically rely on examples, which have been tested cross-linguistically. A large proportion of speakers across the globe would match selected sounds with other sensory properties. One reason for choosing these examples is that the role of sound symbolism in language evolution is discussed in **Cross-Modal Correspondence and Language Evolution**.

Sounds Map to Visual Shapes

One of the most popular cross-modal correspondences is the "bouba"/"kiki" effect, originally known as "baluba"/"takete" (Köhler, 1947). It has been repeatedly shown that when people are asked to match the pseudowords bouba and kiki to a visual shape, most of them will use bouba for the round shape and kiki for the spiky shape (**Figure 1**).

When asked to draw the fantasy characters called bouba or kiki, the artist may end up with characters as shown in **Figure 2**. Those also exhibit rounder shapes for the

Figure 1. *The visual shapes representing a round shape correlated with bouba and an angular shape correlated with kiki. Adapted from Ćwiek et al. (2022), under CC BY 4.0 license.*



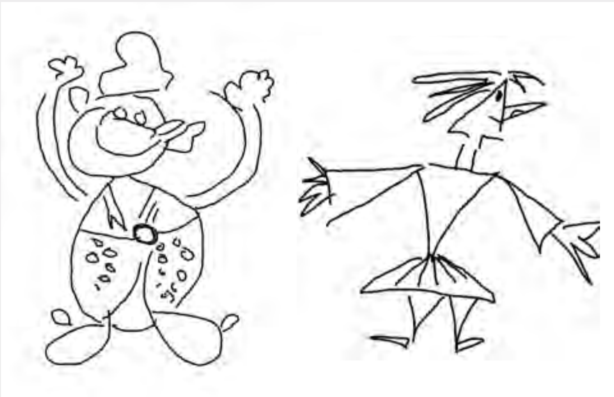


Figure 2. *Drawings of fantasy characters called bouba (**left**) and kiki (**right**). Courtesy of painter Mary Wünsch.*

bouba character and spikier shapes for the kiki character. Interestingly, even the sound of drawing a round versus a spiky shape on paper can be reliably characterized and assigned to the correct bouba and kiki figure, respectively (Margiotoudi and Pulvermüller, 2020).

The effect is a robust cross-linguistic phenomenon (Ćwiek et al., 2022), which has also been shown in infants and children. There are different opinions as to why the sound-to-shape matching takes place. Some say the effect originates from articulation. In bouba, the /uː/ is a rounded sound involving lip protrusion and the /b/ is produced with the lips. In kiki, the /k/ causes the tongue to move up and down, releasing into the vowel. So, the articulatory motions themselves seem rounder or smoother in bouba and more abrupt in kiki.

Others have argued that the acoustic characteristics of the pseudowords match the visual shapes (Ćwiek et al., 2022). During /k/, the fundamental frequency is absent, leading to a period of silence followed by high-intensity, noisy spectral energy. The fundamental frequency and formants only begin at /i/. In contrast, bouba is produced with a continuous fundamental frequency and overall lower spectral energy, so that the changes are less extreme.

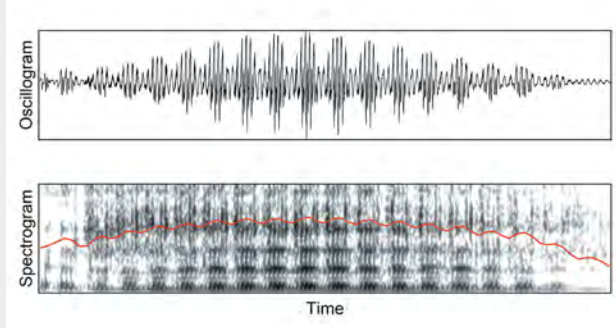
A potential bias with orthography (a system of writing conventions) has also been discussed in the literature. Some researchers have wondered whether the sound-symbolic mapping is not a cross-modal correspondence between sound and shape but rather a shape-to-shape matching between the orthographic shape of letters to the depicted images. The Roman letters <b>, <o>, and <a> are obviously round, whereas <k> and <i> are spiky. However, people around the globe, using different alphabets, match shape and sound in a similar way relatively independent of their orthography (Ćwiek et al., 2022).

For example, in Georgian, the orthographic representations of bouba = ბუბა and kiki = კიკი both look round. Nevertheless, Georgian speakers match the sound of bouba and kiki to the respective round and angular shapes as reliably as speakers of other languages and orthographic systems. If orthography can evoke a bias, this bias is rather weak, whereas sound-to-shape mapping is a robust phenomenon.

Sounds Can Map to Texture

Cross-modal correspondences between sound and touch are less known, but there is reason to believe that these correspondences are deeply connected in evolution. Touch is crucial in sucking, swallowing, mastication, and speech production. The major articulator, the tongue, does not move in free space but is rather in close contact with different vocal tract boundaries. Putting hands on the front of the neck, one can feel the resonances caused by the vocal fold vibrations in sounds that include phonation. There has even been a method called "Tadoma" for deaf-blind individuals that allows them to perceive speech via touching the cheeks and neck of their interlocutor (Rosenblum, 2019). Resonances can also appear on the skin of the neck while vibrating the uvula in a uvular-/ʀ/, similar to when a person is gargling. **Figure 3** shows an acoustic signal

Figure 3. *Acoustics of a tongue tip trilled r-sound. **Top:** an oscillogram. **Bottom:** a spectrogram with the amplitude envelope superimposed (**red line**).*



of a tongue tip /r/-trill, one variety of an r-sound that is common in many languages.

A recent comprehensive study by Winter et al. (2022) has demonstrated compelling evidence for a close link between trilled /r/ and rough surface textures. A rough texture can correspond to an oscillating amplitude envelope (**Figure 3,** *bottom, red line*), but to our knowledge, this idea has not yet been tested explicitly.

Touch and sound are not only linked during speech production but are also connected during object manipulation. Moving the hands along plain paper or the bark of an old tree causes natural sounds that depend on the surface and structure of these two objects. Hence, sound and touch are related, reflecting the texture of the physical world around us.

Winter et al. (2022) found out that words describing rough surfaces (e.g., coarse, barbed, jagged), in comparison to smooth ones (e.g., smooth, oily, slick), have an overrepresentation of r-sounds. To further substantiate their findings on the cross-modal correspondence in sound and touch, they looked at Hungarian, a language with different family roots from English. In Hungarian, similar to English, r-sounds occurred much more frequently in words describing rough textures.

In addition, Winter et al. (2022) compared the antonym pairs *rough* versus *smooth* across 179 languages with a trilled r-sound in their sound inventory and across further 153 languages with a non-trilled r-sound (for different variations of r-sounds, see youtu.be/K9eN2B7Wj68). Only languages with a trilled r-sound, such as Finnish and Indonesian, show a higher probability of /r/ in a word for rough compared with smooth. This shows the resemblance of touch being mapped onto the articulation and acoustics. These findings are an initial attempt to unravel correspondences between touch and sound. If we think about cross-modal correspondences during ingestion as in our initial example, it would not be surprising to find links between sound, touch, and taste as well. We are looking forward to future discoveries along such lines.

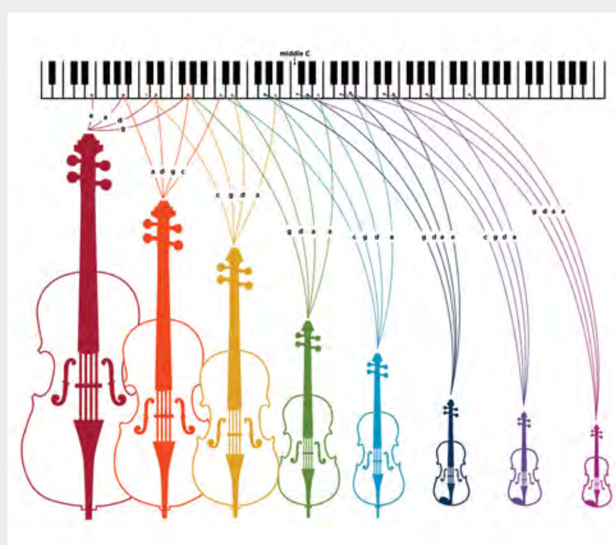
Sounds Can Map to Visual Size

The relationship between certain sounds and size is perhaps the most well-known example of sound symbolism. It turns out that specific speech sounds express the notion of size. The most famous example might be the one between the two vowels /a/ and /i/. In many languages, the physical size dimension is conveyed by those opposing vowels (e.g., Winter and Perlman, 2021). Whereas /a/ is an open vowel, /i/ is a closed one. The radical difference in their articulation evokes the contrast. When we say "teeny tiny," we might even squinch our eyes and lips to make the sensation even more closed and smaller. The opposite is the openness of something "large" or "huge," with a dropped jaw and low voice.

And truly, the reason behind this difference may be as simple as the fundamental frequency of the voice and certain spectral characteristics. Imagine two dogs, a 50-kg (110-lb) German shepherd and a 2.5-kg (5.5-lb) chihuahua. How do you expect their barking to sound? Certainly, even without expertise in acoustics but simply some basic life experience, you know that a 50-kg dog would have a lower pitch bark than a 2.5-kg dog. This example from the animal kingdom extends to other realms.

Let us look at bowed string instruments of various sizes in **Figure 4**. This image stems from the Spring 2020 issue of *Acoustics Today*, where Carleen Hutchins' creation of a violin octet with different resonances, but tonally matched instruments, was featured (Whitney, 2020). You

Figure 4. *Bowed string instruments of various sizes and tuned across a piano range. See text for details. From Whitney (2020), with permission of the New Violin Family Association.*



can see the tuning of those instruments across a piano range. Those instruments have different physical properties, from the resonant body, through the length of the neck, to the thickness of the strings. All of them have four strings, but those strings are tuned very differently from one another. For example, we would not expect any of the strings of a large instrument, like a double bass, to be tuned to E5, which is usually the tuning of the highest string on the regular-sized violin. The larger the instrument, the lower it might sound and vice versa.

The correspondence between sound and visual size might stem from such physical properties. The theory explaining this connection in humans and other animals is called the "frequency code" and was proposed by Ohala (1994). Among other examples, Ohala stated that lower frequencies typically originate from larger sources and higher frequencies from smaller sources. This drives our expectations and can, in turn, be mapped onto more abstract relationships that we can create with the tone of voice, such as the fundamental frequency. Thus, he says that a lower fundamental frequency expresses dominance and a higher fundamental frequency submissiveness. In the end, being larger in the animal kingdom might make an animal threatening and more attractive for mating.

These basic correspondences can move into broader sociocultural uses. Using a high fundamental frequency with exaggerated peaks is typical for child-directed speech. And the reason for that might be that it seems less threatening and more playful to children. As an opposite example, we might look at the so-called "creaky voice," as an extreme case of vocal fry (sounding like an aperiodic low voice; see youtu.be/4L7-9N1xQZA) caused by shortening the vocal folds and lowering the fundamental frequency, occurring mostly in young women. It started gaining popularity in the 1960s and was later covered in pop culture (see Frank and Moon Zappa's song "Valley Girl," at youtu.be/Qb21lsCQ3EM). Originally, the strategy of lowering the fundamental frequency should evoke the image of competence; however, the actual perception of this phenomenon is mixed. Therefore, we have to bear in mind that some abstract social meanings such as competence or politeness seem more complicated and may interfere with the use of the frequency code (Winter et al., 2021). Nevertheless, we might still find support for it in our own voice use in different contexts.

Cross-Modal Correspondence and Language Evolution

Meaningful cross-modal correspondence between sound and size is not only something that humans perceive and use. It is a crucial mechanism in the animal kingdom and its relevance has at least the two following reasons. First, animals sense danger using a variety of perceptual signals, with hearing being a particularly powerful sense.

Second, some animals use the sound-to-size correspondence as a deceptive strategy to their evolutionary advantage (e.g., Bee et al., 2000). A signal can be deceptive if it does not correspond to the actual body size of the vocalizing animal but implies a bigger animal. In that case, it can be a potential threat to the receiver and may increase the chance of survival and mating success. To what extent such a signal is intentional or not is not the main concern of this paper. The use of voice for size estimation is perhaps best visible in the behavior of various deer species (e.g., Reby et al., 2005). The question here is about perception and adaptation. When confronted with roars that suggest a large-sized caller, male red deer respond more frequently and extend their vocal tract due to laryngeal lowering that changes their resonant frequencies (Reby et al., 2005). Studies also show that larger deer have lower frequencies and through that, indirectly, a higher mating success (Vannoni and McElligott, 2008).

The use of sound to deceive and communicate an exaggerated size has been shown in a variety of species so far (e.g., Bee et al., 2000). Some examples include squirrels, birds, and frogs. The deception can function across and within species. For example, juvenile squirrels imitate the voice of their parents to drive away predators (Matrosova et al., 2007). Birds may strategically use vocalizations of their predators to fool fellow birds and enjoy an uninterrupted feast on flies themselves (Munn, 1986). Some frog species lower their fundamental frequency to seem larger in the eyes of other frogs and protect their territory (Bee et al., 2000).

Research on Primates: Empirical Research Is Mostly Unimodal

Although discoveries about sound-sized linkages in animal communication promised to be very fruitful, most studies on primates have been unimodal. In the early stages of research in this field, many findings evolved around the gesture-first theory (Hewes, 1973).

Later, alternative approaches gained attention. From those, we learn that western lowland gorillas, *Gorilla gorilla gorilla*, have impressive control over their breathing behavior (Perlman, 2017). Perlman reported that Koko, a female gorilla (for information and videos, see bit.ly/3HKY47C), could drink through a straw and used different types of breath signals to communicate her attitude. She blew gently onto the face of a person she was fond of and harshly onto someone she did not wish to talk to. Koko not only employed breathing signals but also vocal signals; both were frequently accompanied by gestures (Perlman and Clark, 2015). Works like these might introduce us to the onset of volitional control over vocal behavior. However, it does not stop there. We have known for a few decades that primates, and here we specifically refer to free-ranging East African vervet monkeys, *Chlorocebus pygerythrus*, have different alarm calls for different predators. A call for an eagle is distinct from a call for a leopard, among other threats and predators (Seyfarth et al., 1980). In the end, hiding from an eagle is different than hiding from a leopard.

Even more compelling, however, is the evidence from human-fostered individuals or groups. Kanzi might be the best-known among his species (Savage-Rumbaugh et al., 1986). This male bonobo, *Pan paniscus*, was taught to communicate with so-called lexigrams, a set of symbols conveying a certain meaning (Savage-Rumbaugh et al., 1986). Less known, although not less impressive, was his ability to communicate using vocalizations. Kanzi not only used specific vocalizations in different semantic contexts, but he also modified those vocalizations (Hopkins and Savage-Rumbaugh, 1991). His behavior was similar to modulations such as talking with a higher fundamental frequency to express child-directed speech. Finally, it has previously been shown that both orangutans (*Pongo*) and chimpanzees (*Pan troglodytes*) are able to develop novel vocalizations to capture the attention of their caretakers (Lameira et al., 2016). Such vocalizations are structurally more like human speech than typical primate vocalizations. All in all, this evidence shows that great apes do not only rely on gestures. They very much strategically use vocal signals, too.

According to Slocombe et al. (2011), theoretical approaches in favor of gestures or vocalizations might be biased by empirical work. In a meta-analysis of empirical work carried out between 1960 and 2008, only 5% of the 553 studies on primates analyzed vocalizations and gestures together. In the rest of the studies, researchers focused on one modality only, supporting the respective theoretical view. That means researchers following the gesture-first theory supported their claims with empirical data from gestures but did not challenge it with vocalizations. The same is true for theories focusing on vocalizations, which used empirical data on vocalizations without confronting it directly with gesture data.

Another methodological issue is that approximately 90% of the empirical evidence comes from animals in captive environments with all its consequences on, for example, social behavior, motion, and diet. Furthermore, the living environment of the animals is tightly coupled to the objective of the study. For animals recorded in the wild, researchers focus on vocalizations rather than gestures, which is very likely due to methodological challenges such as the unpredictability of gestures, moving animals, animals interacting within a larger social group, and changes in the camera perspectives. On the other hand, gesture research has often been carried out in captive environments such as zoos. These methodological differences led Slocombe et al. (2011) to a call for more multimodal research to substantiate the theoretical approaches.

Ten years later, Liebal et al. (2022) conducted a similar meta-analysis across 294 studies published between 2011 and 2020. Although certain research gaps have been closed, Liebal and colleagues reported a significant decrease; only 2% of the studies were multimodal.

Pros and Cons for the Role of Vocalizations in Language Evolution

The unimodal perspective on the emergence of communication has spread across the discipline and affected the view on human communication as well. With hand gestures, one can, for example, indexically refer to spatial locations, iconically imitate actions, and depict shapes and sizes of objects. One argument for gestures as the onset of communication is that it was possible to teach some captive apes American Sign Language but not spoken communication.

Another frequently noted fact is that gestures facilitate communication with infants and babies who yet cannot speak. Babies who were taught to use a simplified visual communication system can communicate their

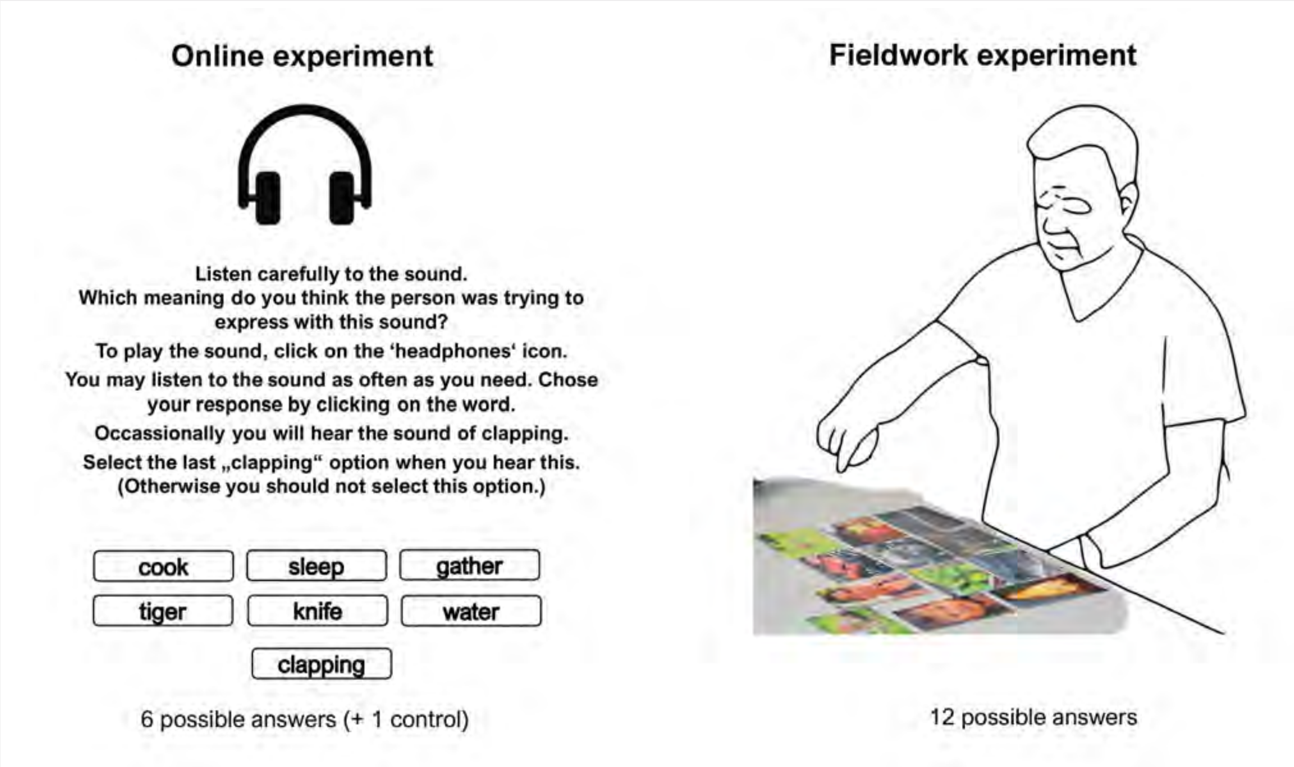


Figure 5. *An overview of the experimental setup.* ***Left:*** *online version of the study.* ***Right:*** *setup in the field experiment. Taken from Ćwiek et al. (2021), under CC BY 4.0 license.*

needs much earlier than would be possible with speech (Barnes, 2010).

A major problem with the assumption that human communication has its origin exclusively in gestures, however, is that it does not explain at *what stage* and *why* a switch from the visual modality toward the auditory modality should have taken place. Furthermore, it has long been assumed that vocalizations might be less depictive than gestures for the creation of novel form-meaning relations.

However, a recent investigation has shown that vocalizations have a much larger iconic potential than previously assumed and can, thus, ground meaning (Ćwiek et al., 2021). In two experiments, an online study and a field experiment, listeners from all over the world heard acoustic signals that were created without using conventional language. These signals expressed a variety of basic concepts like *fire, water, man, woman, snake, hunt, eat, big,* or *many* that might have played a role in the communication of our ancestors (see the Open Science Framework repository for examples at osf.io/4na58). Almost 1,000 participants from 28 languages and 12 language families listened to the vocalizations and selected what they felt was the intended meaning from among different options.

Figure 5 presents how the procedures looked. In the online study, listeners chose 1 from among 6 potential concepts, whereas in the field experiment, they chose the meaning from among 12 pictures. This was done to make the setup accessible to people from different educational backgrounds.

Against previous assumptions, the results of the two experiments showed that participants around the globe were able to comprehend the meaning of these concepts far above the chance level. Thus, the acoustic signal alone has the potential for humans to infer meaning without using language. Still, this does not imply that communication necessarily started only with vocalizations.

The emerging conclusion is that the interplay of gestures and vocalizations might have been crucial at the dawn of communication. It is an advantage to use both because they have different affordances. On one hand, those affordances are situational. Vocal calls are useful to reach a distant receiver, and visual communication might be preferred in close communication or even demanded when we do not wish to attract the attention of others. On the other hand, different affordances relate to the expression of different sensory dimensions. Something visual, such as the shape of an object, is easier to convey using gestures, whereas something auditory, like the tick tock of a clock, may be easier expressed with vocalizations. The connection between the modalities, exposed by cross-modal correspondences, only proves that they are both vital and intertwined. Although it has been pointed out that for human communication it is multimodal at the core, this may also be true for other primates when communicating.

Concluding Remarks

Sounds can become meaningful when they are fused with other sensory information. This property was important to get language off the ground and goes against the traditional assumption in phonology that sounds are only described as meaning-distinguishing units. We provide examples for robust cross-modal correspondences of sound-to-vision and sound-to-touch mappings. Cross-modal correspondences are part of our daily life. They can occur in child-directed speech, in product names, and in fictional characters of cartoons or movies. Sound-to-size correspondence is also meaningful in the animal kingdom. Perceiving the size of a predator in the vocal call of its voice (sound-to-size mapping) can become a matter of survival. However, interdisciplinary work might be necessary to move comparative studies on humans and primates forward because most empirical work on animals focuses on either auditory vocalization or visual gestures. Studies in the wild may face all kinds of methodological challenges. Joint effort by scientists working on open databases for acoustics of animal communication, tools for signal processing, machine learning, optical flow, and video analyses will be necessary for future discoveries.

References

Barnes, S. K. (2010). Sign language with babies: What difference does it make? *Dimensions of Early Childhood* 38(1), 21-30.

Bee, M. A., Perrill, S. A., and Owen, P. C. (2000). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology* 11(2), 169-177.

Blasi, D.E. Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America* 113(39), 10818-10823.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., and Lupyan, G. (2021). Novel vocalizations are understood across cultures. *Scientific Reports* 11(1), 10108.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., and Lupyan, G. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377(1841), 20200390.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology* 14(1/2), 5-24.

Hopkins, W. D., and Savage-Rumbaugh, E. S. (1991). Vocal communication as a function of differential rearing experiences in *Pan paniscus*: A preliminary report. *International Journal of Primatology* 12(6), 559-583.

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters* 11(1), 5-20.

Klink, R. R. (2009). Gender differences in new brand name response. *Marketing Letters* 20(3), 313-326.

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright, New York, NY.

Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., and Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports* 6(1), 30315.

Liebal, K., Slocombe, K. E., and Waller, B. M. (2022). The language void 10 years on: Multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution* 1-14.

Margiotoudi, K., and Pulvermüller, F. (2020). Action sound-shape congruencies explain sound symbolism. *Scientific Reports* 10(1), 12706.

Matrosova, V. A., Volodin, I. A., Volodina, E. V., and Babitsky, A. F. (2007). Pups crying bass: Vocal adaptation for avoidance of age-dependent predation risk in ground squirrels? *Behavioral Ecology and Sociobiology* 62(2), 181-191.

Munn, C. A. (1986). Birds that 'cry wolf.' *Nature* 319(6049), 143-145.

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

Perlman, M. (2017). Debunking two myths against vocal origins of language. *Interaction Studies* 18(3), 376-401.

Perlman, M., and Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition* 18(5), 1165-1179.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences* 272(1566), 941-947.

Rosenblum, L. D. (2019). Does sensory modality matter? Not for speech perception. *Acoustics Today* 15(3), 47-54. https://doi.org/10.1121/AT.2019.15.3.47.

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., and Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology General* 115(3), 211-235.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour* 28(4), 1070-1094.

Slocombe, K. E., Waller, B. M., and Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour* 81(5), 919-924.
Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics* 73(4), 971-995.
Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2), 121-137.
Vannoni, E., and McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PLOS ONE* 3(9), e3113.
Whitney, Q. (2020). American luthier: The art and science of Carleen Hutchins. *Acoustics Today* 16(1), 10-19. https://doi.org/10.1121/AT.2020.16.1.10.
Winter, B., and Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics* 6(1), 1-13.
Winter, B., Oh, G. E., Hübscher, I., Idemaru, K., Brown, L., Prieto, P., and Grawunder, S. (2021). Rethinking the frequency code: A meta-analytic review of the role of acoustic body size in communicative phenomena. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376(1840), 20200400.
Winter, B., Sóskuthy, M., Perlman, M., and Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports* 12(1), 1035.

About the Authors

Susanne Fuchs
fuchs@leibniz-zas.de
Leibniz Zentrum Allgemeine Sprachwissenschaft (Leibniz Center for General Linguistics)
Berlin, Germany

Susanne Fuchs is a senior research scientist and group leader of the Laboratory Phonology Group. She studied social psychology, drama, and linguistics; received a Diploma in Speech Science at the Humboldt University of Berlin, Germany, and a PhD at Queen Margaret University in Edinburgh, United Kingdom. Susanne is a Fellow of the Acoustical Society of America and an associate editor of *The Journal of the Acoustical Society of America* (*JASA*) in the Speech Communication Section. She works on speech motor control, respiration, body motions, cross-linguistic comparisons, and iconicity.

Aleksandra Ćwiek
cwiekaleksandra@gmail.com
Leibniz Zentrum Allgemeine Sprachwissenschaft (Leibniz Center for General Linguistics)
Berlin, Germany
and Humboldt Universität zu Berlin (Humboldt University of Berlin)
Berlin, Germany

Aleksandra Ćwiek earned her BA and MA from Bielefeld University, Bielefeld, Germany, where her area of specialization was acoustic phonetics and her PhD from the Humboldt University of Berlin, Germany. During her PhD, she worked at the Leibniz-Centre General Linguistics in Berlin on iconicity in language. In late 2022, she will start a postdoc fellowship at Radboud University, Nijmegen, The Netherlands, where she will continue her research on iconicity and multimodality. Aleksandra's research interests are iconicity, sound symbolism, multimodality, cognition, and language evolution.