Iconic Prosody is Rooted in Sensori-Motor Properties: Fundamental Frequency and the Vertical Space

Aleksandra Ćwiek (cwiek@leibniz-zas.de)
Leibniz-Centre General Linguistics
Berlin, Germany

Susanne Fuchs (fuchs@leibniz-zas.de)
Leibniz-Centre General Linguistics
Berlin, Germany

Abstract

The iconic cross-modal correspondence between fundamental frequency and location in vertical space ("high is up") has long been described in the literature. However, an explanation for this relationship has not been proposed. We conducted an experiment in which participants shot at cans projected on the wall in different vertical positions. We found that mean fundamental frequency was significantly influenced by vertical head position. Moving the head upwards changes the position of the larynx, which pulls on the cricothyroid muscle and changes the fundamental frequency. We thus propose that the iconic relationship between fundamental frequency and vertical space is grounded in the body.

Keywords: iconicity; prosody; fundamental frequency; vertical space; sensori-motor properties; embodied cognition

Introduction

Iconicity refers to the resemblance between the linguistic *form* and the intended *meaning* (Blasi, Wichmann, Hammarstrm, Stadler, & Christiansen, 2016; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015). The evidence brought forward in the last decade suggests that iconicity is an essential part of language (Perniss, Thompson, & Vigliocco, 2010; Perniss & Vigliocco, 2014). The origin of iconicity, however, is far from clear (cf. e.g., Imai & Kita, 2014, p. 9). This article brings together earlier work in phonetics and cognitive sciences by discussing the relationship between iconicity and prosody in speech.

The term prosody has often been used interchangeably with intonation, though these terms are used in different ways by different authors (for a discussion, cf. Hirst & Di Cristo, 1998). Here, we use prosody as an umbrella term for the suprasegmental stress, rhythm, and intonation properties in speech (Bussman, 1996; Trask, 2004). We investigate a single basic aspect of prosody, fundamental frequency, which is an acoustic correlate of intonation. Fundamental frequency expresses the rate at which the vocal folds vibrate during speech. It is most frequently measured in hertz. Pitch expresses how fundamental frequency is perceived. It is quantified by listeners' judgments. In this paper we use f0 in reference to production studies and pitch in reference to perception studies.

Previous research has shown that there is a relationship between location in vertical space and pitch (in speech perception) or f0 (in speech production), though little in the way of explanation has been proposed. In the literature, iconicity is mainly considered to be both a finding and the explanation for the finding: there is a form–meaning mapping, because the relationship is iconic. However, no explanation is given for why iconicity is present in natural language. Therefore, the aim of this study is to investigate a potential origin for iconicity. Using an experimental approach, we demonstrate that there is a link between vertical head movement and f0. We propose that the iconic correspondence between location in vertical space and f0 is rooted in head movement required to look at objects that are higher up. In other words, the origin for the relationship between f0 and location in vertical space lies in bodily constraints, namely vertical head movement.

Evidence for Iconic Pitch in Speech Perception

Researchers have long been interested in how people localize sounds in vertical space and what kind of cognitive processes are involved (e.g., Seashore, 1899). Pioneering work by Pratt (1930) and Trimble (1934) and later experiments by Mudd (1963) and Roffler and Butler (1968) show that, regardless of the actual vertical position of the sound source, participants tend to locate high-pitched sounds higher in the vertical plane and low-pitched sounds lower in the vertical plane (i.e., the Pratt effect, cf. above). Recent work by Parise, Knorre, and Ernst (2014) provides an insight into the possible source of this "frequency–elevation mapping." The authors recorded sounds in different natural environments with directional microphones mounted at various heights. They found a strong correlation between the frequencies of the noises in the environment, especially in the 1–6 kHz range, and the sound location in the vertical space. After accounting for the filtering properties of the outer ear, Parise et al. conclude that the ear is fine-tuned to the statistics of the noises in the environment. Thus, the relationship between a given sound's frequency and its vertical position is not language-specific. Apart from that, their results suggests that the listeners' expectations of an object's location in vertical space may be grounded in the statistical probabilities in the natural environment. It has to be noted that the noises in the environment come from both animate, like birds, and inanimate sources, like wind in the trees.

In 1994, Ohala proposed the *frequency code* as a possible cause of prosodic iconicity (Ohala, 1994). He argued that in various species, low-pitched vocalizations are associated with a large-sized animal, since the mass of the vocal folds correlates with body mass and, thus, size. Both body mass

and the size of an animal are crucial to estimate a potential threat, which in the animal kingdom is a matter of life and death. The lower the perceived pitch of a given animal, the more threatening, dominant, or aggressive the animal is assumed to be. And conversely, the higher the perceived pitch emitted by the sound source, the smaller its size is estimated to be, thus the source itself is interpreted as less threatening, dominant, and aggressive.

Keeping in mind the studies mentioned above and Ohala's frequency code, when we consider humans, it is apparent that there are two main, possibly contradictory, factors that affect pitch estimation. The first is vertical position – higher pitch is located higher in vertical space. The second factor is body size – higher pitch is associated with smaller body size. If two people of different size (height) stand beside each other, the larger person's mouth will always be higher in the vertical plane. On the one hand, according to Ohala's frequency code, we would expect the larger person to emit lower-pitched sounds. On the other hand, according to, e.g., Parise et al. (2014), the larger person, because their mouth is higher up in the vertical plane, should sound higher-pitched. This contradictory predictions have, to our knowledge, only been addressed by one study (Pisanski, Isenstein, Montano, O'Connor, & Feinberg, 2017).

Pisanski et al. (2017) explicitly investigated body size estimation based on pitch vs. vertical location. Their results show that low pitch was associated with a large body size even when it was played from a low vertical position. This suggests that pitch cues override spatial cues in the body size estimation. However, this finding may be due to the task, which was to estimate the body size of an animate being. Hence, animacy and the experimental task may also have influence on the perception of different frequencies.

Evidence for Iconic F0 in Speech Production

There is considerably less work on iconic f0 in speech production. Although the effects found in the studies mentioned below are mostly subtle, nevertheless they provide evidence for the use of iconic pitch with regard to location in vertical space. Items that are located higher up in vertical space (whether they are actual objects or mental concepts located in a metaphorical plane) are marked with higher fundamental frequency than item that are located lower in space.

In a series of experiments, Clark, Perlman, and Johansson Falck (2014) asked the participants to read stories related to vertical motion (up vs. down), emotions (positive vs. negative), and perceived sound (high-pitched vs. low-pitched). The authors expected that the participants would produce higher fundamental frequency in stories with higher elevation in the physical space, positive emotions, and high auditory pitch in contrast to stories reporting lower vertical space, negative emotions, and lower auditory pitch. A significant effect was found only for stories describing a vertical motion in which the f0 was on average 5 Hz higher in the "up" condition than in the "down" condition.

Nygaard, Herold, and Namy (2009) investigated prosody of adjective antonym pairs (e.g., *happy/sad*, *big/small*, *tall/short*). The adjectives were embedded in carrier sentences next to novel non-words and the participants were asked to use infant-directed speech. Their analyses suggest that the fundamental frequency was higher in the adjectives *happy*, *big*, *hot*, *tall*, *yummy* and *strong*, compared to their antonym counterparts. Depending on the item, f0 differences were between 20 and 90 Hz. However they might have been affected by how engaged the participant was in the infant-directed speech task.

Another study that tackles the problem of iconic f0 with regard to vertical space is that of Shintel, Nusbaum, and Okrent (2006). The authors analyzed the fundamental frequency of participants saying if an animated dot was moving up or down. The data revealed a significantly higher f0 for the "up" condition. The f0 differences were, however, relatively small, similarly to those reported by Clark et al. (2014).

The reviewed previous literature, both in perception and production, documents the relationship between f0 and location in vertical space that is proposed to be iconic. But the only explanation given for this relationship is iconicity itself. The purpose of this study is to test one potential origin of iconicity and thus also the origin of the relationship between f0 and location in vertical space.

Potential Anatomical Explanations for the Relationship between F0 and Vertical Space

The control of fundamental frequency is anatomically complex. It involves a difference in the subglottal pressure between the lungs and the oral cavity, in addition to the tensing of the vocalis muscles (the vocal folds). Moreover, different extrinsic muscles can indirectly influence the vocal folds. For example, activation of the cricothyroid muscle (CT) leads to the rotation of the crycoid cartilage, which in turn tenses the vocal folds, and thus increases f0 (Honda, 1996). Apart from that, fundamental frequency can be lowered by the actions of the external strap muscles (Erickson, Baer, & Harris, 1982).

But additional factors may come into play in f0 control by causing a change within the other parameters affecting f0, e.g., by varying the muscle tension around the larynx. Anatomically, muscle tension around the larynx can be changed by head movement. To the best of our knowledge, only one empirical study exists showing the influence of head motion on f0 in speech production (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Their sentence-by-sentence multiple regression analysis revealed that 63% of the variation in the fundamental frequency could be explained by speaker's head movement during speech production. The upward head motion raises the larynx, thereby pulling on the CT muscle, which elongates the vocal folds and thereby increases f0. Still, it has to be noted that the analyses were carried out on recordings of only one speaker.

The aim of this study is to investigate the potential anatomical origin of the iconic relationship between f0 and location in vertical space. Our first hypothesis is that the fundamental frequency is affected by the position of the head. Changes in

the position of the head influence the placement of the larynx, and simultaneously affect the fundamental frequency. We assume that people look at an object, the head follows the gaze. When an object is located higher up in space, people move their head upwards and when it is located lower, they move their head down. Thus, if an object is located at the higher position in space, we expect a higher f0 within the utterance produced at this head position. Our second hypothesis is that the size of an object has an impact on the fundamental frequency when referring to that object. We expect a higher f0 for smaller objects and lower f0 for larger objects.

Methods

Experimental Design

In the experimental task, participants were asked to "shoot" cans, which were projected onto the wall in front of them, with a laser pointer. Additionally, the participants had to say the word written on the can. One of two words was written on each can: *piff* [pɪf] or *paff* [paf], both of which are German onomatopoeic words imitating the sound of shooting, like English 'bang' or 'pow'. To measure the hypothesized effect of size of the object referred to, we used two sizes of cans in the experiment – a small and a large one, which was approximately twice as large. The cans appeared on five equidistant positions on the vertical and horizontal axes, resulting in 25 possible positions. The varying vertical position of the can enabled us to elicit the head movement, expected to have an influence on fundamental frequency – according to the first hypothesis. All conditions sum up to a total of 100 tokens per participant: two words x two can sizes x five vertical positions x five horizontal positions.

During the task, the participants stood at a landmark on the ground, at approximately 1 m distance from the wall. The projection surface measured 130 x 130 cm, with the lowest edge starting at 145 cm above the ground. When a can appeared, the participants were instructed to (1) point at the can with a laser pointer, and (2) say the word written on the can. After the participant successfully pointed at the can and uttered the word written on it, an animation of the can falling down was played and, after a short blank screen, a new can in a different position appeared. The presentation of cans and their animation was manually controlled to prevent the participants from predicting when the next can would appear.

Five datasets with prerandomized order of presentation were created to avoid order effects. For technical reasons, the presentation of the stimuli was divided in two sections, both of which consisted of 50 items. The whole experiment took 15–20 minutes, though the experimental task itself lasted no longer than 5–6 minutes. This was highly relevant in order to avoid boredom and its potential effect on fundamental frequency. The main task was preceded by a short familiarization trial, which consisted of five items – cans with different words written on them, than those that were used in the main task. The participants were given no specific instructions regarding the alignment of their movements. If one of them asked about it, they were told to act naturally, in a way similar to pointing a laser pointer at a particular word or image while giving an oral presentation.

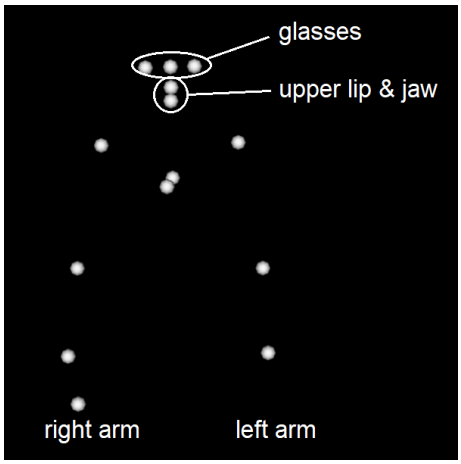

Figure 1: The placement of markers for the motion capture.

Acoustic and motion data were recorded simultaneously using a Sennheiser ME 64 cardioid microphone and an Optitrack motion capture system (Motive, version 1.9.0) with 12 cameras (Prime 13). Motion data was captured at 120 Hz sampling frequency and acoustic data at 44.1 kHz. In total, 15 markers were placed on different body parts of the participants: three on the glasses (center, left, and right); one each on the upper lip and the lower lip (jaw); one at the position of the sternum; one approximately at the location of the fourth thoracic spine vertebra; one on the laser pointer; and symmetrically two on the shoulders, elbows, and wrists. The placement of all markers is illustrated in Figure 1.

Due to technical problems with the recording equipment, the data of one participant had to be excluded.

Participants

Since males have an on average lower fundamental frequency than females, including different sexes would have yielded an additional factor in the experimental design. To avoid this, we focused on females as a participant group. A total of 31 German native speakers took part in the study (mean age = 27.84; mean height = 167.7 cm, with min = 152 cm and max = 183 cm). Twenty-six participants were monolingual and five reported being bilingual; all apart from one were right-handed. The participants were all recruited using a participant database. At the beginning of the session, they were given information about the experiment and signed a consent form. They received monetary compensation after the completion of the task. The project was approved by the ethical board of the DFGS and preregistered at Open Science Framework[1].

[1]The OSF repository can be visited under the following address: https://osf.io/ysr75/

Data Preprocessing and Annotation

The acoustic data were automatically labeled at the phoneme level using WebMAUS (Kisler, Reichel, & Schiel, 2017) and subsequently manually corrected using Praat (Boersma & Weenink, 2018). The on- and offset of the vowel were defined as the onset and offset of vocal fold oscillations, respectively. So far, we have annotated and corrected the data for 20 participants (mean age = 29.45; mean height = 165.5 cm, with min = 152 cm and max = 177 cm). Mean fundamental frequency was calculated for the whole vowel interval. The fundamental frequency range was set to 150–400 Hz to avoid octave jumps of the pitch tracker due to creaky voice.

The motion capture data were first extracted and processed with Mokka version 0.6.2 (Barré & Armand, 2014), and subsequently converted to be further processed with MATLAB (version R2017b). The maximal vertical position of the pointing wrist and of the center of the glasses were calculated within the vowel interval, provided by the annotated acoustic data.

Statistical Analyses

All statistical analyses were carried out within the R environment, version 3.5.1 (R Core Team, 2018), using the following packages: `plyr` (Wickham, 2011) for data wrangling, `car` (Fox & Weisberg, 2011) and `lme4` (Bates, Mächler, Bolker, & Walker, 2015) for statistical modelling, `RePsychLing` (Baayen, Bates, Kliegl, & Vasishth, 2015) for model evaluation, and `stargazer` (Hlavac, 2015) for the output table.

After the initial data exploration, we were forced to exclude the data of one participant (id7) from subsequent analysis. Her behavior during the experiment was atypical; she shrieked and laughed a lot during the experiment. This later had a negative effect on the reliable parameter extraction. Thus, all results refer to the group of 19 female participants.

Results

General Remarks

Though we did not explicitly investigate the coordination between the head movement (gaze), articulation (lip and jaw movement), and the pointing gesture, their temporal organization is shown in Figure 2 for reference. The speaker first visually locates the target and elevates the head (in the example shown in Figure 2, the target is situated higher in the vertical space). Then she starts the pointing gesture by visibly lifting the wrist. Finally, the speech itself begins. It can be observed in the acoustic signal itself, but also in the larger maximal distance of the upper lip and jaw.

Statistical Hypotheses Testing

The parsimonious mixed model approach was used to more reliably analyze data that are highly variable between subjects (Bates, Kliegl, Vasishth, & Baayen, 2015, p. 2). This approach starts by calculating a maximal model, as suggested by Barr (2013). Subsequently, a principal component analysis of the random effects' structure is run using the `RePsychLing`

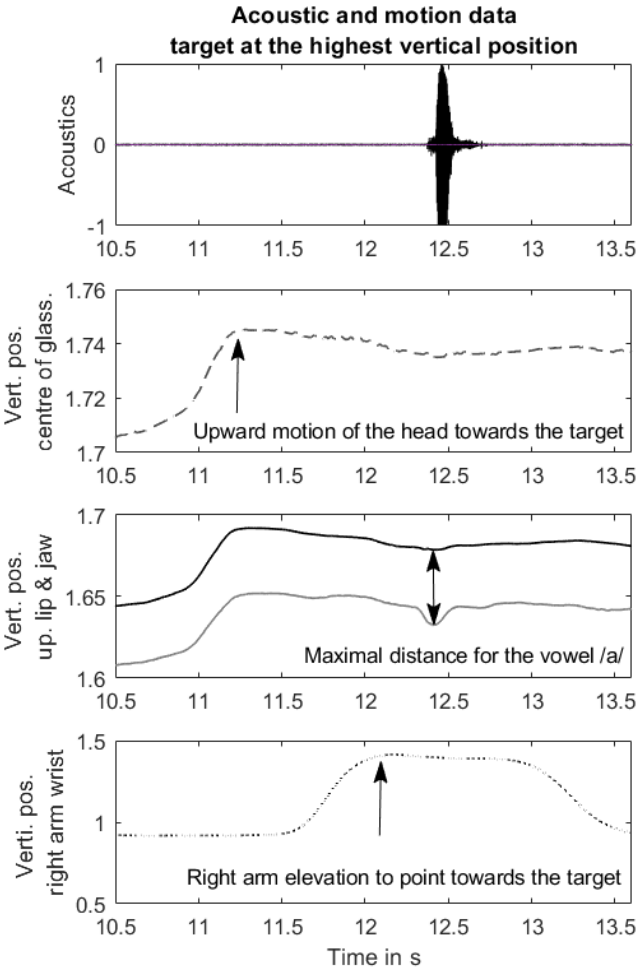

Figure 2: An example of acoustic and motion data for a trial with the stimulus located high in the vertical plane. Plots from top to bottom: (1) acoustic signal; (2) vertical head position (dashed line), determined by the marker in the middle of the glasses; (3) upper lip (black line) and jaw (gray line) markers; (4) marker on the pointing arm wrist (dashed line). All motion data are in meters. Note that the arrows in charts 2–4 only illustrate the coordination among different body parts, but all positions are calculated within the vowel interval.

package (Bates, Kliegl, et al., 2015). This procedure allows to reduce the random effects structure to the necessary components, according to the variance in the data. It is done by an iterative reduction of the model's structure and step-by-step comparison of subsequent models. In the current analysis, the parsimonious best-fit model was in fact not significantly different from the maximal model ($p = 0.54$).

All computed models consisted of the same set of fixed factors: vowel in the uttered word, subject's height, subject's head position, can's vertical position, can size, and the interaction of vowel and can size. Random intercept for subject was included in the random effects structure to account for intersubject variability. The random slopes after the iterative model reduction consisted of vowel in the uttered word, and the interaction of vowel and can size.

The results of the the linear mixed effects model presented in Table 1 reveal four main effects on the mean fundamental frequency, namely that of: the vowel segment, participant's body height, participant's head position, and the size of the can. The first two effects have been described in the past. No effect on the mean f0 has been found for either the can's vertical position or the interaction of vowel segment and can size.

The effect of the vowel on the mean f0 is in line with previous reports on the intrinsic f0 in vowels (Whalen & Levitt, 1995; Whalen, Gick, Kumada, & Honda, 1999). This is the most robust effect found in our data and it shows that higher mean f0 values were found for the high vowel /ɪ/ and lower values for the low vowel /a/. Furthermore, the data show that the taller the participant is, the lower her fundamental frequency is. This negative effect found for participant's height corroborates with the frequency code (Ohala, 1994) and the work by Pisanski and Rendall (2011).

Most importantly, both hypotheses put forward earlier gain support from our analysis. The model shows that the mean f0 increases with the elevation of the head, which is consistent with the first hypothesis. In addition, the size of the can had a significant impact on the mean f0 – participants produced lower mean f0 when referring to larger cans. This result is in line with the second hypothesis. However, it has to be pointed out that the mean f0 differences found between smaller and larger cans are rather small at only 2–3 Hz. Figure 3 illustrates that the effect found for size is the strongest in the highest can positions (1 and 2) and it diminishes or disappears completely at the lower positions (3–5).

Discussion

The analysis presented above demonstrates that in our data the mean fundamental frequency is influenced by the vertical head position rather than the vertical position of the object referred to. In the computed best-fit model with mean fundamental frequency as a dependent variable, we found that the tested anatomical factor – head position – plays a crucial role for the mean f0. In contrast, a factor depicting a purely iconic relationship in the location on the vertical plane – can's ver-

Table 1: The results of the linear mixed model analysis. The table shows the effect of fixed factors, listed on the left, on the dependent variable: mean f0. The estimated effect size is given for each factor, with the standard error given in brackets in the line below.

	Dependent variable:
	Mean f0
Intercept	225.455***
	(5.358)
Vowel	15.978***
	(2.347)
Participant's height	−13.481**
	(5.986)
Head position	9.371***
	(3.041)
Can's vertical position	−0.441
	(0.572)
Can size	2.225**
	(1.005)
Vowel * can size	1.326
	(1.894)
Observations	1,872
Log Likelihood	−7,841.494
Akaike Inf. Crit.	15,704.990
Bayesian Inf. Crit.	15,765.870
Note:	*p<0.1; **p<0.05; ***p<0.01

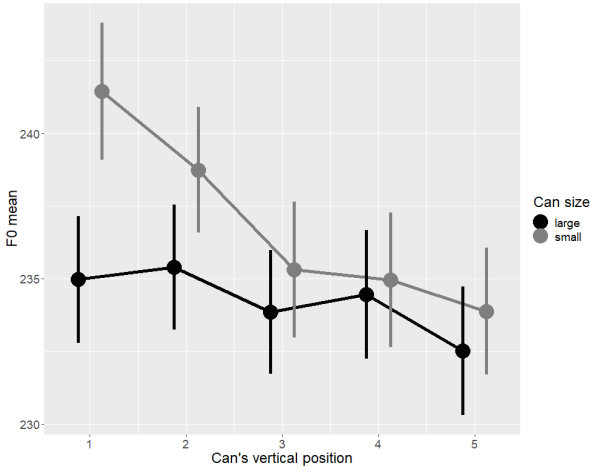

Figure 3: Mean f0 values for the can in different sizes and vertical positions. On the horizontal axis, 1 means high and 5 means low position. The darker line depicts a large-sized can and the lighter one a small-sized can.

tical position – did not significantly account for the variance in the mean f0. Thus, we propose that the origin for the relationship between f0 and location in vertical space in speech production is rooted in the body and its sensori-motor properties. The cross-modal relationship between f0 and vertical space, as discussed in the framework of iconicity, is not simply a result of an iconic form and meaning mapping ("high is up"). It is a result of embodiment, driven by the sensorimotor properties, which are influenced by the changes in the head position.

Furthermore, we found that the size of an object had a significant impact on mean f0 values in the current experiment (cf. Figure 3). In speech production, a correspondence between fundamental frequency and size of an object has been previously found in a story reading task by Perlman, Clark, and Falck (2015), apart from other studies mentioned in the introduction. In this study, participants were asked to read stories with concepts of *fast/slow* and *big/small*. Reading pace and f0 were measured and it was found that (1) the stories with "fast" concepts were read faster than those with "slow" ones, and (2) the stories with "big" concepts were read with lower pitch than those with "small" ones. Our study supplements previous findings on f0–size symbolism by showing that participants reliably signal a difference in size of an inanimate object by adapting their fundamental frequency. There is a large body of work on the iconic f0–size relationship on the segmental level (cf. e.g., Shinohara & Kawahara, 2010; Tsur, 2006; Ultan, 1978). It has been shown in various cross-linguistic analyses that high vowels are more frequently used in words depicting smaller objects, and low vowels in words depicting larger objects. Therefore it was highly relevant to control for the interaction of vowel and can size in the current analysis, though no significant effect was found.

We would like to point out that even though vertical head movement does affect the f0 in our data, it is not the only thing that affects f0 in speech production. Speakers have a high degree of control over f0 and it is often employed to signal prominence in speech (Terken, 1991), such as word stress and sentence accent. Speakers can manipulate their pitch according to the needs of the communicative situation. We found that speakers varied greatly in how they completed the task. Some participants barely moved their head, while others did, even if it did not seem necessary. Even though a small number of participants showed very little head movement, we still found that the head position was one of the strongest predictors of mean f0 variance in our data.

The degree of the vertical head movement varied not only between the participants, but also between the vertical positions of the cans. It can be seen in Figure 3 – lower vertical positions (3–5) yield smaller or no differences in mean f0 between one another. This can be a side effect of a sufficient body size to visually process lower targets. A thorough analysis of speaker behavior is yet to be conducted.

The present study proposes a potential origin for the iconic relationship between f0 and object location in vertical space. Our data show that head movement influences f0: upward head movement leads to higher f0, which is most likely a result of the larynx pulling on the cricothyroid muscle. We thus propose that the iconic relationship between f0 and vertical space is rooted in the body: when looking at an object located higher on the vertical plane, head movement generally reflects the location of the object. In this case, the head movement itself could be considered iconic, because the form is aligned with the meaning – the head moves upwards toward an upward target. Previous literature has already established an iconic relationship between f0 and object location. Our study supplements previous interpretations of this iconic relationship with evidence that the correspondence stems from bodily constraints. The upward movement of the head causes physiological changes that influence f0. In this way, we argue that the correspondence between f0 and vertical space is both embodied – because it is rooted in the body – and iconic – because the form and the meaning correspond. "High is up", because the head moves upwards and the effect is thus an embodied form–meaning correspondence.

Acknowledgments

This work was funded by the German Research Council as a part of the XPrag.de project PSIMS: The Pragmatic Status of Iconic Meaning in Spoken Communication (FU 791/6-1).

References

Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). Repsychling: Data sets from Psychology and Linguistics experiments [Computer software manual]. (R package version 0.0.4)

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

Barré, A., & Armand, S. (2014). Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine*, *114*, 80-87.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 2018-03-20. `http://www. praat. org/`.

Bussman, H. (1996). *Routledge Dictionary of Language and Linguistics*. New York: Routledge. (Translated and edited by Gregory P. Trauth and Kerstin Kazzasi)

Clark, N., Perlman, M., & Johansson Falck, M. (2014). Iconic pitch expresses vertical space. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), (pp. 393–410). Stanford: SCLI Publications.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015, October). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Erickson, D., Baer, T., & Harris, K. S. (1982). The role of the strap muscles in pitch lowering. *Haskins Laboratories: Status Report on Speech Research*, *70*, 275–284.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second ed.). Thousand Oaks CA: Sage.

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages*. Cambridge University Press.

Hlavac, M. (2015). stargazer: Well-formatted regression and summary statistics tables [Computer software manual]. Cambridge, USA. (R package version 5.2)

Honda, K. (1996). Biological Mechanisms for Tuning Voice Fundamental Frequency. *Koutou (THE LARYNX JAPAN)*, *8*(2), 109–115.

Imai, M., & Kita, S. (2014, August). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326 - 347.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, *66*(4), 347–352.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*(2), 133–137.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009, January). The Semantics of Prosody: Acoustic and Perceptual Evidence of Prosodic Correlates to Word Meaning. *Cognitive Science*, *33*(1), 127–146.

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

Parise, C. V., Knorre, K., & Ernst, M. O. (2014, April). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111*(16), 6104–6108.

Perlman, M., Clark, N., & Falck, M. J. (2015, October). Iconic Prosody in Story Reading. *Cognitive Science*, *39*(6), 1348–1368. doi: 10.1111/cogs.12190

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*.

Perniss, P., & Vigliocco, G. (2014, August). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 1–14.

Pisanski, K., Isenstein, S. G. E., Montano, K. J., O'Connor, J. J. M., & Feinberg, D. R. (2017, feb). Low is large: spatial location and pitch interact in voice-based body size estimation. *Attention, Perception, & Psychophysics*, *79*(4), 1239–1251.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201-2212.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*(3), 278–285.

R Core Team. (2018). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Roffler, S. K., & Butler, R. A. (1968, June). Localization of Tonal Stimuli in the Vertical Plane. *The Journal of the Acoustical Society of America*, *43*(6), 1260–1266.

Seashore, C. (1899). Localization of sound in the median plane. *Univ. Iowa Stud. Psychol*, *2*, 46–54.

Shinohara, K., & Kawahara, S. (2010, August). A Cross-linguistic Study of Sound Symbolism: The Images of Size. *Annual Meeting of the Berkeley Linguistics Society*, *36*(1), 396–410.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*, 167-177.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America*, *89*(4), 1768–1776.

Trask, R. L. (2004). *A Dictionary of Phonetics and Phonology*. New York: Routledge.

Trimble, O. C. (1934, January). Localization of Sound in the Anterior, Posterior and Vertical Dimensions of Auditory Space. *British Journal of Psychology*, *24*(3), 320–334.

Tsur, R. (2006, June). Sizesound symbolism revisited. *Journal of Pragmatics*, *38*(6), 905–924.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 525–568). Stanford University Press Stanford, CA.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999, April). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27*(2), 125–142. doi: 10.1006/jpho.1999.0091

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic f0 of vowels. *Journal of Phonetics*, *23*(3), 349–366.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from `http://www.jstatsoft.org/v40/i01/`