Congenitally Deaf Children Generate Iconic Vocalizations to Communicate Magnitude

Marcus Perlman (mperlman@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
University of Wisconsin-Madison
Madison, WI 53706 USA

Jing Z. Paul (jingzhangpaul@ufl.edu)
Department of Languages, Literatures and Cultures,
301 Pugh Hall
University of Florida
Gainesville, FL 32611 USA

Gary Lupyan (glupyan@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
University of Wisconsin-Madison
Madison, WI 53706 USA

Abstract

From an early age, people exhibit strong links between certain visual (e.g. size) and acoustic (e.g. duration) dimensions. Do people instinctively extend these crossmodal correspondences to vocalization? We examine the ability of congenitally deaf Chinese children and young adults (age M = 12.4 years, SD = 3.7 years) to generate iconic vocalizations to distinguish items with contrasting magnitude (e.g., big vs. small ball). Both deaf and hearing (M = 10.1 years, SD = 0.83 years) participants produced longer, louder vocalizations for greater magnitude items. However, only hearing participants used pitch—higher pitch for greater magnitude – which counters the hypothesized, innate size "frequency code", but fits with Mandarin language and culture. Thus our results show that the translation of visible magnitude into the duration and intensity of vocalization transcends auditory experience, whereas the use of pitch appears more malleable to linguistic and cultural influence.

Keywords: crossmodal correspondence; deafness; iconicity; language evolution; magnitude; vocalization

Introduction

People tend to link certain auditory dimensions to certain visual dimensions (Spence, 2011). For example, they associate loudness with size and brightness (a loud sound is big and bright), pitch with size and elevation (a high pitched sound is small and high), and the temporal duration of a sound with length (a temporally extended sound is long)[1]. A large body of evidence indicates that some of these cross-modal correspondences are highly robust, especially those involving prothetic dimensions that can be characterized in terms of more or less magnitude, such as loudness, quantity size, and duration (Walsh, 2003). These correspondences are detectable early in development, and can influence low-level perceptual processes, as well as high-level processes like the use of linguistic metaphor (Winter, Marghetis, & Matlock, 2014). In this study, we examine whether certain cross-modal correspondences between sight and sound also extend to the production of iconic vocalizations. Do people have a similarly instinctive sense of how to map visual dimensions of magnitude to qualities of their voice? To find out, we test whether people who are congenitally deaf are able to generate iconic vocalizations that reflect visual dimensions of magnitude. Such a result would indicate that size-vocalization correspondences can originate even in the absence of auditory experience.

[1] Dimensions like *size* and *elevation* may be primarily sensed through vision and are typically presented visually in experiments, but they can obviously be experienced through non-visual senses too.

Origins of Cross-Modal Correspondences

To understand the origins of cross-modal correspondences, scholars have focused on when these mappings arise in development. One possibility is that people learn through experience to associate auditory dimensions like pitch and loudness with visual dimensions like size because of their tight correlation in the environment. The physical laws of sound dictate that bigger objects tend to produce lower pitched and louder sounds, and research shows that listeners are sensitive to these properties when judging the size of falling objects (Grassi, 2005). Thus children might internalize the statistical correlations between size, pitch and loudness through their experience with colliding objects and other sound producing events (Spence, 2011).

A second potential source of some crossmodal mappings is language (Marks, Hammeal, & Bornstein, 1987; Smith & Sera, 1992). For example, a child learning English will learn to use the words "long" and "short" to describe extension in both space and time. Or a child learning Mandarin will learn that the word "gāo", meaning 'high' or 'tall', also occurs in the word "gāoyīn", which means 'high pitch', and in the word "gāoda" which can refer to someone or something that is big and tall. This association is also reinforced by many Chinese folk songs, in which the use of high pitch (i.e. gāoyīn) is used to express strength and power. Thus a Chinese child might learn to associate the concepts of tall, big, and high-pitched and their corresponding opposites.

Finally, it is also possible that certain cross-modal mappings—particularly those relating to magnitude—are innate and arise from evolved sensory and neural physiology. For example, humans may be equipped with a generalized mental magnitude system that represents prothetic dimensions like loudness, size, and brightness

according to a common, amodal or multimodal magnitude representation (Walsh, 2003). Bigger objects, louder sounds, and brighter lights may correspond because they are instinctively at the "more" end.

In support of claims for innateness, some crossmodal links have been observed in prelinguistic infants who have had just limited opportunity to learn associations between acoustic and visual events. For example, Srinivasan and Carey (2010) found that, like adults, nine-month-old infants are more likely to remember pairings of lines and tones when length and duration are positively correlated. Infants aged 3-4 months showed an association between pitch and both visual-spatial height and visual sharpness (Walker, et al., 2010). The earliest age for which there is evidence of crossmodal correspondence comes from a study by de Hevia et al. (2014), which tested neonates 7 to 94 hours old. Within just a few hours of birth, newborns showed sensitivity to cross-modal mappings between the prothetic domains of numerosity (number of spoken syllables), temporal sequences (duration of syllables), and spatial extent (visible line length).

Cross-Modal Correspondence in Vocalization

Substantial evidence indicates that, from an early age, people have a strong sense of correspondence between certain visual and auditory dimensions. Do they similarly possess a deeply ingrained sense of how these correspondences extend to vocalization? How readily can people generate *iconic* vocalizations that bear acoustic properties corresponding to visual dimensions of magnitude?

Some scholars have proposed that at least one mapping – that between size and the pitch of vocalization – has an ancient evolutionary history, evolving in adaptation to the physiology of tetrapod vertebrate vocal tracts (Morton, 1994). Large, threatening animals produce low-pitched vocalizations, and small, non-threatening animals produce high-pitched ones. Ohala (1994) suggests that this hypothetically innate size "frequency code" pervades spoken communication and underlies a number of important functions of intonation in speech, including the marking of questions and the expression of many affective qualities (e.g. deference, authority, submission, confidence).

According to the frequency code proposal, humans are born with an instinctive sense of how to express magnitude through the pitch of their voice. In addition, given evidence of infants' early sensitivity to correspondences between visual magnitude and auditory dimensions like duration and loudness, humans may also possess a strong sense of how to express magnitude through the duration and intensity of their voice. To assess these predictions, we examine whether congenitally deaf Chinese children and young adults, lacking auditory experience entirely, are nevertheless able to generate iconic vocalizations to communicate different visual dimensions of magnitude. We also test a comparison group of hearing Chinese children to further investigate the influence of language and culture in forming a sense of correspondence between magnitude and voice.

Methods

Participants

The first group of participants included 19 Chinese children and young adults with congenital deafness resulting in severe to complete hearing loss. Their mean age was 12.4 years (SD = 3.7 years). The second group consisted of 16 Chinese, Mandarin speaking children with normal hearing. Their mean age was 10.1 years (SD = 0.83 years).

Materials

Participants communicated a set of 8 items contrasting along four dimensions of magnitude: a short vs. a long string (length), a small vs. a big ball (size), a little vs. a lot of rice (amount), and a few (2) vs. many (5) marbles (quantity).

Procedure

Deaf participants were tested at the special education boarding school they attended. The experiment was conducted by a native speaker of Mandarin Chinese, who was assisted by a bilingual teacher at the school who spoke Mandarin and Standard Chinese Sign Language (CSL). The teacher provided instructions in CSL.

Participants were first introduced to the experiment as a group in their home classrooms. The assisting teacher placed the stimuli – the four contrasting pairs of items – on a desk in front of the class, with the two contrasting items of each pair placed next to each other. She noted that the objects in each pair were different, and asked the children to sign the difference. The children were generally able to identify each contrasting feature (e.g., "big" for the big ball and "small" for the small ball). After going through all the items, the teacher explained the basic procedure of the experiment.

The children were told they would play a "guessing game" with their teacher and the experimenter. The experimenter would point to one of the two items of each pair, and the children would make a vocal sound to communicate the selected item to their teacher, whose back would be turned so that she could not see. They were told that they should not try to make a corresponding Mandarin Chinese word nor a random sound. Instead they should try to make a meaningful sound that they thought would help their teacher choose the right item.

Participants were tested individually in a quiet office at the school immediately following the classroom introduction. The child was seated at a table beside the experimenter, and the teacher sat with her back to the table. All of the items were placed in a row in front of them, with paired items placed next to each other and extra space between pairs. The instructions were repeated by a signing assistant as necessary.

On each trial, the experimenter first announced in Mandarin the superordinate name of the pair of items that would be tested (e.g. ball) so that the guessing teacher knew which pair to select from. Then the participant produced a sound to communicate the selected item to the teacher. The teacher then turned toward the pair of items and pointed to indicate her guess of which one had been selected. No other feedback was provided.

Each item was tested once during a session. One item of each pair was selected in the first block, and the remaining item was selected in the second block. The order of items was randomized between participants. The session was audio-recorded for analysis.

This same basic procedure was also used with the hearing children who participated while attending their day school. The primary difference was that the instructions and experiment were conducted in Mandarin by the experimenter.

Analysis

Acoustic measurements Acoustic measurements were made with Praat phonetic analysis software (Boersma, 2001). The onset and offset of each vocalization was marked in a textgrid without knowledge of its associated, and afterwards the intervals were labeled for analysis. Duration, intensity, and pitch were measured automatically.

Statistical analyses Statistical analyses with mixed effects models were conducted using the lme4 package in R. Significance tests were calculated using chi-square tests that compared the improvement in fit of mixed-effect models with and without the factor of interest. Dimension (e.g. size) was included as a random effect in models collapsing across all items, and participant was included as a random effect in all models.

Results

Figure 1 shows the complete results for deaf and hearing participants.[2] As can be seen, for many of the domains, both groups reliably used the duration and intensity of their voice to communicate greater magnitude. In contrast, only hearing participants reliably used pitch – specifically, higher pitch for greater magnitude. Below we first report the results with each of the items collapsed together into greater versus lesser magnitude, and then we report each domain of magnitude separately.

Magnitude: Greater vs. Lesser (All Items)

All participants together produced vocalizations with a mean duration of 690 ms for greater items and 590 ms for lesser items. Magnitude was a reliable predictor of duration, $\chi^2(1) = 17.6$, $b = 0.10$, 95% CI = [0.06, 0.15], $p < .001$. There was no interaction between magnitude and hearing ability, $\chi^2(1) = 1.2$, $p = .27$. Separately, hearing participants produced a mean duration of 650 ms for greater items and 510 ms for lesser items, which was a reliable difference, $\chi^2(1) = 10.4$, $p = .001$, $b = 0.14$, 95% CI = [0.05, 0.22]. Deaf participants produced a mean duration of 730 ms for greater items and 650 ms for lesser items, which was also reliable, $\chi^2(1) = 8.8$, $p = .003$, $b = 0.08$, 95% CI = [0.03, 0.13].

Overall, participants produced a mean intensity of 62.8 dB for greater items and 58.9 dB for lesser items, which was a reliable difference, $\chi2(1) = 82.9$, $p < .001$, $b = 3.9$, 95% CI = [3.1, 4.7]. There was a reliable interaction between magnitude and hearing ability, $\chi2(1) = 6.1$, $b = 1.9$, 95% CI = [0.4, 3.5], $p = .01$. Hearing participants produced a mean intensity of 64.7 dB for greater items and 59.8 dB for lesser items, which was a reliable difference, $\chi2(1) = 63.9$, $p < .001$, $b = 4.99$, 95% CI = [3.9, 6.0]. Deaf participants produced a mean intensity of 61.3 dB for greater items and 58.2 dB for lesser items, which was also reliable $\chi2(1) = 26.9$, $p < .001$, $b = 3.0$, 95% CI = [1.9, 4.2].

Overall, participants produced a mean pitch of 297 Hz for greater items and 285 Hz. for lesser items. Magnitude was a reliable predictor of pitch, $\chi2(1) = 4.5$, $p = .034$, $b = 11.6$, 95% CI = [0.9, 22.4]. There was a reliable interaction between magnitude and hearing ability, $\chi2(1) = 11.4$, $p < .001$, $b = 37.0$, 95% CI = [15.7, 51.2]. Hearing participants produced a mean pitch of 300 Hz for greater items and 269 Hz for lesser items, which was a reliable difference, $\chi2(1) = 11.0$, $p < .001$, $b = 32.8$, 95% CI = [13.8, 51.7]. Deaf participants produced a mean pitch of 295 Hz for greater items and 296 Hz for lesser items, which was not reliable, $\chi2(1) = 0.39$, $p = .53$.

In summary, both groups showed a strong inclination to produce longer and louder vocalizations to communicate the greater items compared to shorter, quieter vocalizations for the lesser items. Hearing participants, but not deaf participants, produced higher pitched vocalizations for greater magnitude items and lower pitched vocalizations for lesser items.

Length: Long vs. Short

Overall, participants produced a mean duration of 760 ms for the long string and 590 ms for the short string. Length was a reliable predictor of duration,

$\chi2(1) = 11.1$, $p < .001$, $b = 0.18$, 95% CI = [0.08, 0.28]. There was a marginal interaction between length and hearing ability, $\chi2(1) = 3.1$, $p = .08$, $b = 0.17$, 95% CI = [-0.02, 0.36]. Hearing participants produced a mean duration of 840 ms for the long string and 570 ms for the short string, which was a reliable difference, $\chi2(1) = 6.6$, $p = .01$, $b = 0.27$, 95% CI = [0.07, 0.47]. Deaf participants produced a mean duration of 700 ms for the long string and 600 ms for the short string, which was also reliable, $\chi2(1) = 6.9$, $p = .009$, $b = 0.10$, CI = [0.03, 0.17].

Overall, participants produced a mean intensity of 63.7 dB for the long string and 58.9 dB for the short string.

[2] In a few cases, our analyses revealed a reliable interaction between age and magnitude for deaf participants. Because of the limited space available here, we save report of these interactions for a future article.

Length was a reliable predictor of intensity, $\chi2(1) = 26.0$, $p < .001$, $b = 4.8$, CI = [3.2, 6.3]. There was no interaction between hearing and length, $\chi2(1) = 0.81$, $p = .37$. Hearing participants produced a mean intensity of 65.6 dB for the long string and 60.1 dB for the short string, which was a reliable difference, $\chi2(1) = 13.3$, $p < .001$, $b = 5.5$, 95% CI = [3.0, 8.1]. Deaf participants produced a mean intensity of 62.0 dB for the long string and 57.9 dB for the short string, which was also reliable, $\chi2(1) = 13.1$, $p < .001$, $b = 4.1$, 95% CI = [2.2, 6.1].

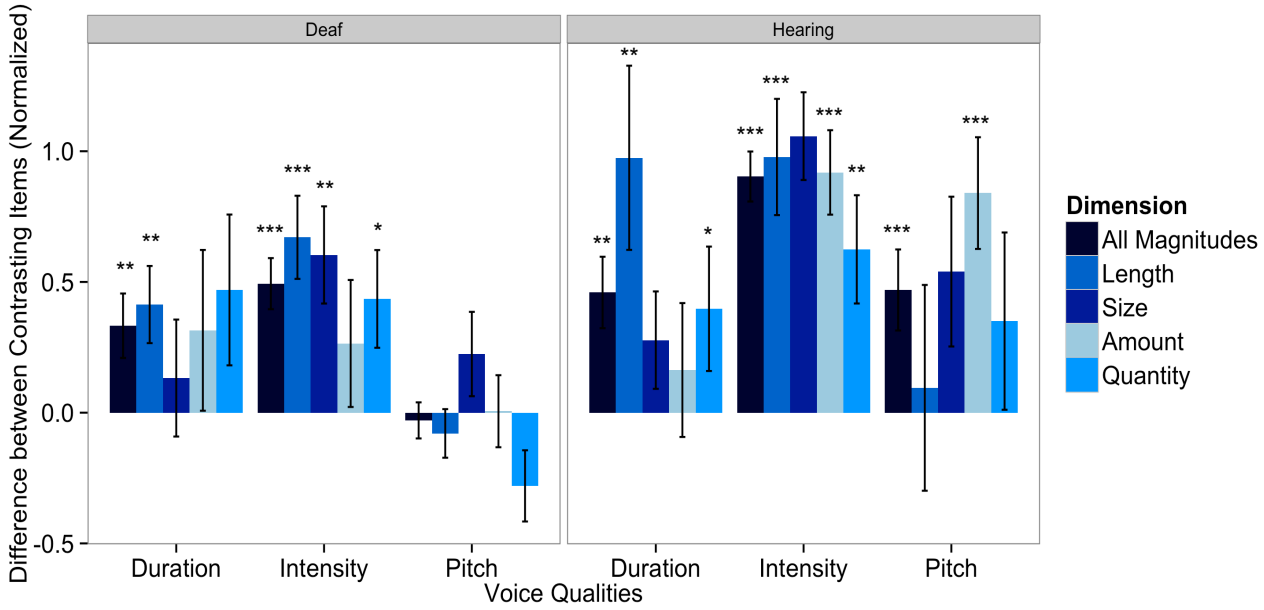

Figure 1. Average differences in acoustic properties between contrasting items. Deaf participants are displayed on the left, hearing participants on the right. The x-axis shows the three acoustic properties, and the y-axis shows normalized values for comparison between the properties. Error bars represent the standard errors of the mean differences. Stars indicate level of significance: * $p < .05$, ** $p < .01$, *** $p < .001$. For example, deaf participants produced vocalizations that were ~ 0.4 SDs longer in duration for the long string compared to the short string.

Overall, participants produced a mean pitch of 287 Hz for the long string and 290 Hz for the short string, which was not a reliable difference, $\chi2(1) = 0.04$, $p = .85$. There was no interaction between length and hearing ability, $\chi2(1) = 0.28$, $p = 0.60$. In summary, both groups produced longer and louder vocalizations to refer to the long string, and shorter, softer vocalizations to refer to the short string. Neither group used pitch to distinguish between the lengths of string.

Size: Big vs. Small

Overall, participants produced a mean duration of 620 ms for the big ball and 570 ms for the small ball, which was not a reliable difference, $\chi^2(1) = 1.98$, $p = .15$. There was no reliable interaction between size and hearing ability, $\chi^2(1) = 0.37$, $p = .54$.

Overall, participants produced a mean intensity of 63.5 dB for the big ball and 58.7 dB for the small ball. Size was a reliable predictor of intensity, $\chi^2(1) = 26.2$, $p < .001$, $b = 4.8$, 95% CI = [3.2, 6.3]. There was no reliable interaction between size and hearing ability, $\chi^2(1) = 2.23$, $p = .13$. Hearing participants produced a mean intensity 65.0 dB for the big ball and 59.1 dB for the small ball, which was a reliable difference, $\chi^2(1) = 20.7$, $p < .001$, $b = 6.0$, 95% CI = [4.1, 7.9]. Deaf participants produced a mean intensity of 62.2 dB for the big ball and 58.5 dB for the small ball, which was also reliable, $\chi^2(1) = 8.76$, $p = .003$, $b = 3.7$, 95% CI = [1.4, 6.0].

Overall, participants produced a mean pitch of 308 Hz for the big ball and 282 Hz for the small ball. Size was a reliable predictor of pitch, $\chi^2(1) = 5.11$, $p = .024$, $b = 27.0$, 95% CI = [3.8, 50.1]. There was no reliable interaction between size and hearing ability, $\chi^2(1) = 0.61$, $p = 0.43$. Hearing participants produced a mean pitch of 299 Hz for the big ball and 263 Hz for the small ball, which was a marginally reliable difference, $\chi^2(1) = 3.26$, $p = 0.071$, $b = 36.9$, 95% CI = [15.4, 76.9]. Deaf participants produced a mean pitch of 315 Hz for the big ball and 295 Hz for the small ball, which was not reliable, $\chi^2(1) = 1.94$, $p = .16$.

In summary, both groups produced louder, but not longer, vocalizations to distinguish between the big and small ball. Both showed a trend of using higher pitch for the big ball; however, this tendency was only marginally reliable for hearing participants and not reliable for deaf participants.

Amount: A Lot vs. A Little

Overall, participants produced a mean duration of 690 ms for a lot of rice and 630 ms for a little rice. Amount was not a reliable predictor of duration, $\chi^2(1) = 1.66$, $p = .20$. There was no reliable interaction between amount and hearing ability, $\chi^2(1) = 0.05$, $p = 0.83$.

Overall, participants produced a mean intensity of 62.1 dB for a lot of rice and 59.1 dB for a little rice. Amount was

a reliable predictor of intensity, $\chi^2(1) = 9.19$, $p = .002$, $b = 3.11$, 95% CI = [1.18, 5.03]. There was a marginally reliable interaction between amount and hearing ability, $\chi^2(1) = 3.09$, $p = .079$, $b = 3.3$, 95% CI = [-0.40, 6.96]. Hearing participants produced a mean intensity of 64.2 dB for a lot of rice and 59.6 dB for a little rice, which was a reliable difference, $\chi^2(1) = 16.21$, $p < .001$, $b = 5.06$, 95% CI = [3.16, 6.88]. Deaf participants produced a mean intensity of 60.4 dB for a lot of rice and 58.7 dB for a little rice, which was not reliable, $\chi^2(1) = 1.21$, $p = 0.27$.

Overall, participants produced a mean pitch of 294 Hz for a lot of rice and 272 Hz for a little rice. Amount was a reliable predictor of pitch, $\chi^2(1) = 5.04$, $p = .024$, $b =$, 95% CI = [3.3, 45.3]. There was a reliable interaction between amount and hearing ability, $\chi^2(1) = 8.62$, $p = .003$, $b = 57.90$, 95% CI = [20.8, 94.8]. Hearing participants produced a mean pitch of 303 Hz for a lot of rice and 249 Hz for a little rice, which was a reliable difference, $\chi^2(1) = 10.1$, $p = 0.001$, $b = 57.84$, 95% CI = [26.8, 88.1]. Deaf participants produced a mean pitch of 288 Hz for a lot of rice and 288 Hz for a little rice, which was not reliable, $\chi^2(1) = 0$, $p = 0.98$.

In summary, neither group distinguished a lot from a little with the duration of their vocalizations, although hearing participants reliably made this distinction by intensity. Deaf participants showed the same pattern, although it was not reliable. Hearing, but not deaf participants, distinguished a lot from a little with higher pitch.

Quantity: Many vs. Few

Overall, participants produced a mean duration of 690 ms for many marbles and 560 ms for a few marbles. Quantity was a reliable predictor of duration, $\chi^2(1) = 6.07$, $p = 0.014$, $b = 0.12$, 95% CI = [0.03, 0.22]. There was no reliable interaction between hearing and quantity, $\chi^2(1) = 0.04$, $p = 0.85$. Hearing participants produced a mean duration of 610 ms for many marbles and 470 ms for a few marbles, which was a reliable difference, $\chi^2(1) = 3.92$, $p = 0.048$, $b = 0.13$, 95% CI = [0.10, 0.27]. Deaf participants produced a mean duration of 750 ms for many marbles and 640 ms for a few marbles, which was not reliable, $\chi^2(1) = 2.61$, $p = 0.11$.

Overall, participants produced a mean intensity of 61.9 dB for many marbles and 58.9 dB for a few marbles. Quantity was a reliable predictor of intensity, $\chi^2(1) = 11.92$, $p < .001$, $b = 3.03$, 95% CI = [1.41, 4.65]. There was no reliable interaction between quantity and hearing ability, $\chi^2(1) = 0.24$, $p = 0.62$. Hearing participants produced a mean intensity of 63.8 dB for many marbles and 60.4 dB for a few marbles, which was a reliable difference, $\chi^2(1) = 7.40$, $p = 0.007$, $b = 3.48$, 95% CI = [1.13, 5.81]. Deaf participants produced a mean intensity of 60.5 dB for many marbles and 57.8 dB for a few marbles, which was also reliable, $\chi^2(1) = 5.00$, $p = 0.025$, $b = 2.69$, 95% CI = [0.37, 5.01].

Overall, participants produced a mean pitch of 299 Hz for many marbles and 296 Hz for a few marbles. Quantity was not a reliable predictor of pitch, $\chi^2(1) = 0.03$, $p = 0.85$, $b = -2.19$, 95% CI = [-25.86, 22.17]. There was a reliable interaction between quantity and hearing ability, $\chi^2(1) = 4.32$, $p = 0.038$, $b = 48.6$, 95% CI = 2.94, 94.15]. Hearing participants produced a mean pitch of 310 Hz for many marbles and 279 Hz for a few marbles, which was not a reliable difference, $\chi^2(1) = 1.93$, $p = 0.16$. Deaf participants produced a mean pitch of 292 Hz for many marbles and 309 Hz for a few marbles, which was marginally reliable, $\chi^2(1) = 3.56$, $p = 0.059$, $b = -23.03$, 95% CI = [-46.53, 0.99].

In summary, both groups produced louder vocalizations for many compared to few marbles. Only hearing participants reliably produced longer vocalizations for many, although deaf participants showed the same numeric pattern. Hearing participants did not use pitch to distinguish amount, and deaf participants showed only a marginal trend to produce lower pitched vocalizations for many.

Discussion

From an early age, people show evidence of strong associations between certain visual and auditory dimensions, such as size, duration, and loudness. Do people possess a similarly robust sense of how to extend these cross-modal correspondences to the production of iconic vocalizations? We examined the ability of congenitally deaf Chinese children and young adults—who literally lack any auditory experience to speak of—to generate iconic vocalizations to communicate different visual dimensions of magnitude. Thus we investigated whether mappings between visual magnitude and different vocal qualities can originate even in the absence of auditory experience.

Both deaf participants and a comparison group of hearing Chinese children reliably produced longer and louder vocalizations for greater magnitude items, compared to shorter, quieter vocalizations for items with lesser magnitude. Separate analysis of each domain suggests that both deaf and hearing participants also made more nuanced vocal distinctions between the different dimensions. For instance, both more consistently used the duration of their voice to distinguish length compared to other dimensions. Altogether, these results show that people share a strong sense of how to translate dimensions of visible magnitude into the duration and intensity of their vocalizations, and that this sense transcends auditory experience.

Some scholars have postulated an innate size frequency code that humans have inherited in adaptive response to the physiology of tetrapod vertebrate vocal tracts (Ohala, 1994). However, we found that only hearing, and not deaf participants reliably used pitch to distinguish magnitude. One likely reason for this is the especially fine motor control required to modulate pitch (Fitch, 2010), at which our deaf participants are relatively unpracticed and disadvantaged given their lack of auditory feedback. The result also suggests that the association between pitch and size may depend on a functioning auditory system and learning.

Further evidence in favor of learning comes from our results with hearing participants. Counter to the size frequency code, hearing Chinese children produced *higher*

pitched vocalizations for greater compared to lesser magnitude items. Notably, this pattern also differs from two previous studies using a vocal charades task with American undergraduates, who tended to produce high-pitched vocalizations for small and low-pitched vocalizations for big (Perlman & Cain, in press; Perlman, Dale, & Lupyan, under review). The hypothesis that associations between size and pitch are subject to learning is also supported by previous developmental studies. For example, whereas adults matched higher pitched tones with smaller lights, children did not make this association until 11 years of age (Marks, et al., 1987).

There is good reason to consider that hearing participants' use of high pitch for greater magnitude is shaped by their experience with Mandarin and Chinese culture. Previous work has found that the conventional metaphorical expressions a language uses to describe pitch—for example "high" and "low" in English or the equivalents of "thin" and "thick" in Farsi—influence the spatial dimensions by which pitch is conceptualized by speakers (Dolscheid, Shayan, Majid, & Casasanto, 2013). Similarly, we described above the use of the Mandarin root "gāo," which, can be used to refer to a person who is big and tall in size, and also to a high pitched sound. This association is also displayed by other aspects of Chinese culture, such as the use of high pitch to express strength and power in folk songs. Thus it is likely that the hearing children in our study were influenced by these linguistic and cultural conventions.

An additional explanation for the pitch-size correspondence produced by Chinese children may relate to the physiology of vocalization. According to the "effort code," vocalizations involving higher effort and intensity tend to occur with a rise in pitch (Gussenhoven, 2002). Thus the production of higher pitch may have been a physical consequence of producing more intense vocalizations. However, while worth consideration, the disassociation between size and pitch with deaf participants weighs against this possibility.

Conclusion

Our findings highlight the human potential to generate novel vocalizations that are grounded in our conceptions of magnitude and space. While the association of size and pitch may be subject to linguistic and cultural influence, the association of size with vocal qualities of duration and intensity is quite robust. Even when people entirely lack auditory experience, they nevertheless share a strong sense of how to translate visual dimensions of magnitude into the duration and intensity of vocalization. The results show that cross-modal correspondences between dimensions of magnitude extend to the motor system and vocal tract, and they are instinctively incorporated into the production of iconic vocalizations.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341-345.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Science, 11,* 4809-4813.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science, 24,* 613-621.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*, 6:1–12.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics, 67,* 274-284.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference.* Aix-en-Proence: ProSig and Université de Provence Laboratoire Parole et Language.

Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development, 52,* 1-102.

Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nicholls, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge: Cambridge University Press.

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

Perlman, M., & Cain, A. A. (in press). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*.

Perlman, M., Dale, R., & Lupyan, G. (under review). Iconicity can ground the creation of a vocal symbol system.

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

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

Srinivasan, M., & Carey, S. (2010). The long and the short of it: on the nature and origin of functional overlap between representations of space and time. *Cognition*, *116*, 217-241.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synesthetic cross-modality correspondences. *Psychological Science*, *21*, 21–25.

Walsh, V. (2003). A theory of magnitude: Common cortical metrices of time, space, and quality. *Trends in Cognitive Sciences, 7,* 483-488.

Winter, B., Marghetis, T., & Matlock, T. (2014). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex, 64,* 209-224.