ISCI, Volume *4*

Supplemental Information

Human Listeners Can Accurately Judge Strength and Height Relative to Self from Aggressive Roars and Speech

Jordan Raine, Katarzyna Pisanski, Anna Oleszkiewicz, Julia Simner, and David Reby

Supplemental Information

Transparent Methods

All experiments were approved by the University of Sussex's Life Sciences & Psychology Cluster-based Research Ethics Committee (C-REC) (Certificates of approval: ER/JR307/8, ER/JR307/9) and comply with the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

Participants

Vocal stimuli were recorded from 30 male and 31 female (*M* age = 22.79 ± 1.12) drama or acting students from the Royal Central School of Speech and Drama and the University of Sussex, United Kingdom, who received monetary compensation in exchange for their participation.

We recruited separate samples of participants to provide voice-based assessments of the relative strength and height of vocalizers. The sample that rated strength (hereafter Experiment 1) consisted of 19 females and 26 males (age = 31.44 ± 8.33) recruited from Tromso and surrounding rural towns in Norway (N = 11, all fluent English speakers), and from the University of Sussex, UK (N = 34), in return for prize draw monetary compensations (5 x £20). The sample that rated height (hereafter Experiment 2) consisted of 31 females and 25 males (age = 34.27 ± 10.39), recruited from the USA using Amazon Mechanical Turk, and compensated with $1.75 USD. Participants from both experiments provided informed consent and completed the experiment online using a custom computer interface. Data from one female and male participant in Experiment 1, and from two female and two male participants in Experiment 2, who did not complete the experiment but rated more than half of the stimuli, were included in our analysis.

Materials

Vocal stimuli.

Vocalizers were audio recorded producing an aggressive roar and aggressive speech in a quiet, anechoic room, standing 150 cm from a Zoom H4n microphone. A chair was placed at this distance to restrict participants from moving closer to the microphone. Vocalizers were instructed to produce the speech sentence, ‘That’s enough, I’m coming for you!’, followed by a nonverbal vocalisation expressing the same motivation, while imagining themselves in a battle or war scenario, about to charge and attack. This resulted in a total of 122 vocal stimuli (see Electronic Supplementary Materials for examples of aggressive roars and aggressive speech).

To obtain realistic vocal stimuli, participants were encouraged to take as much time as they needed to immerse themselves in each imagined context, and to ‘let go of their inhibitions’. Participants were also given the option not to vocalise if they felt that they could not naturally produce the sentence or nonverbal vocalisation, and to repeat each sentence or vocalisation until they were satisfied with their portrayal. Recordings were saved as WAV files at 44.1 kHz sampling frequency and 16 bits amplitude resolution.

Physical formidability measures.

We measured the height of vocalizers using metric tape. The average height of our sample of vocalizers (male *M* = 182.03 ± 0.97 cm; female *M* = 167.10 ± 1.19 cm) compares well with that of the general UK population (male *M* = 175.3cm, female *M* = 161.9 cm, Moody, 2013). Flexed bicep circumference and handgrip strength were also measured, and these measurements were aggregated to produce a single, equally weighted, z-scored strength value for each subject (following Sell et al. 2009; Puts et al. 2011, and others). These measures explain approximately 55% and 24% of the variance in strength as measured by weight-lifting machines in male college students, respectively (Sell et al., 2009).

To measure flexed bicep circumference (male *M* = 32.09 ± 0.60 cm; female *M* = 28.96 ± 0.70 cm), participants were instructed to rest the elbow of their dominant arm on a table while seated, to clench their fist, and to curl their forearm perpendicular to the table. The experimenter measured the circumference of the bicep at its highest point. A Baseline hydraulic hand dynamometer in its standard use was used to measure the handgrip strength of participants' dominant arm (male *M* = 41.57 ± 1.36 kg; female *M* = 26.98 ± 1.06 kg). Each strength measure was recorded twice per subject and the highest achievable score, representing greatest strength, was used in analyses.

Procedure

All playback experiments were completed online on Syntoolkit, a dedicated online testing platform for psychology studies (e.g., Hughes, Gruffydd, Simner & Ward, in press; see Simner & Alvarez, in prep) that is particularly suited to running studies with sensory or multisensory stimuli. Listeners were instructed to use headphones and complete the experiment in a quiet place. To allow listeners to complete the experiment at a comfortable but audible volume, they were instructed to first set their volume to its lowest level. Listeners then heard a demo sound file (amalgamating a loud and quiet stimulus), and were instructed to raise their volume until they could clearly hear the quiet stimulus, while the louder stimulus did not cause discomfort. Following this, listeners were asked not to adjust the volume settings during the experiment unless it became too uncomfortable, and were asked at the end of the experiment if they had done so. Due to the agonistic nature of the stimuli, listeners were made aware that if they felt uncomfortable or distressed listening to the sounds, they could stop the experiment.

In playback experiments, vocal stimuli (n = 122) were blocked by sex and stimulus type (speech/roar). The order of blocks and stimuli within blocks was randomised. Before each block, participants were reminded to listen to each stimulus in full, and informed that they could take a break at any time. Listeners rated the physical strength (Experiment 1) or height (Experiment 2) of each voice stimulus ("Rate by how much this person is stronger/taller or weaker/shorter than you") on a 101-point scale from -50 (much weaker/shorter) to 50 (much stronger/taller). We set the slider's default position to 0 (described as 'same as you') and did not compel listeners to move the slider so as not to artificially force directional judgments.

Listeners were debriefed upon completion that the roars and screams were acted, and that the vocalizers were not really experiencing aggression or distress. We examined reaction times against stimulus durations to ensure that participants listened to the stimuli before entering their ratings. No participants responded before half of the stimulus had elapsed on more than five trials, thus no listeners were excluded.

To assess whether listeners could accurately judge the physical characteristics of vocalizers relative to their own, we measured listeners' own physical characteristics. In

Experiment 1, we used a tailor's tape measure to measure bicep circumference (male *M* = 33.89 ± 0.46 cm; female *M* = 28.12 ± 0.57 cm), and a Takei hand dynamometer to measure handgrip strength (male *M* = 46.11 ± 1.67 kg; female *M* = 33.03 ± 1.10 kg), in identical fashion to measurements taken from vocalizers. These measures were taken in person, prior to the listener completing the playback experiment online at a time of their choosing. Both vocalizer and listener strength z-scores were calculated based on a pooled sample of the listeners' and the vocalizers' measurements. Experiment 2 relied on a self-report measure of height given at the start of the playback experiment (male *M* = 176.38 ± 1.30 cm; female *M* = 169.36 ± 1.48 cm). The validity of self-report measures of height has been extensively studied, and despite slight overestimations, self-reported height closely reflects measured height within the age range of our sample of listeners (Krul, Daanen, & Choi, 2011; Lim, Seubsman, & Sleigh, 2009; Parker, Dillard, & Phillips, 1994; Wada et al., 2005).

Coding and Statistical Analysis

To examine strength/height estimation in functionally relevant terms, we divided the actual difference in strength/height into five categories. In Experiment 1, percentage differences between -10% and 10% were coded as 'similar strength', differences between ± 10% and ± 30% were coded as 'vocalizer is stronger (weaker) than listener', and differences greater than ± 30% were coded as 'vocalizer is much stronger (weaker) than listener'. In Experiment 2, we calculated by how many centimetres the vocalizer was taller than the listener. Values were coded into identical categories of 11 cm intervals. This interval was chosen as it produced a similar distribution to that observed for our actual strength difference categories.

In both experiments, we coded the rated difference in strength/height between listener and vocalizer into three categories. Ratings between 45 and 55 were categorised as ‘rated as similar strength’, and ratings above (below) this range were coded as ‘vocalizer rated as stronger (weaker)’. We computed a linear mixed multinomial logistic regression, testing the effects of the actual strength/height difference between listener and vocalizer, vocalizer sex, listener sex, and stimulus type on the rated difference between listener and vocalizer, excluding actual difference categories with sample sizes less than 15. In all models, we included listener identity as a subject variable, and vocalizer identity as a random factor, thus allowing the intercepts and slopes of the relationships between predictors and outcomes to vary between both vocalizers and listeners and testing null hypotheses based on the average of these intercepts and slopes.

List of Supplemental Audio Files

(F = female vocalizer; M = male vocalizer)

F01 Roar.wav

F01 Speech.wav

F02 Roar.wav

F02 Speech.wav

F03 Roar.wav

F03 Speech.wav

M01 Roar.wav

M01 Speech.wav

M02 Roar.wav

M02 Speech.wav

M03 Roar.wav

M03 Speech.wav

References

Hughes, J.E.A., Gruffydd, E., Simner, J., and Ward, J. (in press). Synaesthesia aids in savant-skill acquisition: training calendar calculation in sequence-space synaesthetes. Cortex.

Krul A.J., Daanen H.A.M. and Choi H., Self-reported and measured weight, height and body mass index (BMI) in Italy, The Netherlands and North America, *Eur. J. Public Health* **21**, 2011, 414–419.

Lim L.L., Seubsman S. and Sleigh A., Validity of self-reported weight, height, and body mass index among university students in Thailand: implications for population studies of obesity in developing countries, *Popul. Health Metr.* **7**, 2009, 15.

Moody, A. (2013). Adult anthropometric measures, overweight and obesity (Health Survey for England - 2012). NHS. Retrieved from http://digital.nhs.uk/catalogue/PUB13218.

Parker J.M., Dillard T.A. and Phillips Y.Y., Impact of using stated instead of measured height upon screening spirometry, *Am. J. Respir. Crit. Care Med.* **150** (6 Pt 1), 1994, 1705–1708.

Sell A., Cosmides L., Tooby J., Sznycer D., von Rueden C. and Gurven M., Human adaptations for the visual assessment of strength and fighting ability from the body and face, *Proc. R. Soc. Lond. B Biol. Sci.* **276**, 2009, 575–584.

Simner, J., & Alvarez, J. (in prep). The Syntoolkit: an online testing platform for behavioural research experiments.

Puts, D.A., Apicella, C.L., & Cárdenas, R.A., Masculine voices signal men's threat potential in forager and industrial societies, *Proc. R. Soc. Lond. B Biol. Sci.* rspb20110829.

Wada K., Tamakoshi K., Tsunekawa T., Otsuka R., Zhang H., Murata C. and Toyoshima H., Validity of self-reported height and weight in a Japanese workplace population, *Int. J. Obes. (Lond.)* **29**, 2005, 1093–1099.