The role of loudness in vocal intimidation.

Andrey Anikin, Daria Valente, Katarzyna Pisanski, Clement Cornec, Gregory A Bryant, David Reby

The role of loudness in vocal intimidation

Andrey Anikin[1,2], Daria Valente[2,3], Katarzyna Pisanski[2,4,5], Clement Cornec[2], Gregory A. Bryant[6], and David Reby[2,7]

[1] Division of Cognitive Science, Lund University, Lund, Sweden
[2] ENES Bioacoustics Research Lab / Lyon Neuroscience Research Centre (CRNL), University of Saint-Etienne, CNRS UMR5292, INSERM UMR_S 1028, Saint-Etienne, France
[3] Department of Life Sciences and Systems Biology, University of Torino, Turin, Italy
[4] CNRS French National Centre for Scientific Research, DDL Dynamics of Language Lab, University of Lyon 2, 69007, Lyon, Fance
[5] Institute of Psychology, University of Wrocław, Poland
[6] Department of Communication, University of California, Los Angeles
[7] Institut universitaire de France

Address correspondence to andrey.anikin@lucs.lu.se

Abstract
Across many species, a major function of vocal communication is to convey formidability, with low voice frequencies traditionally considered the main vehicle for projecting large size and aggression. Vocal loudness is often ignored, yet it might explain some puzzling exceptions to this frequency code. Here we demonstrate, through acoustic analyses of over 3000 human vocalizations and four perceptual experiments, that vocalizers produce low frequencies when attempting to sound large, but loudness is prioritized for displays of strength and aggression. Our results show that, although being loud is effective for signaling strength and aggression, it poses a physiological tradeoff with low frequencies because a loud voice is achieved by elevating pitch and opening the mouth wide into a-like vowels. This may explain why aggressive vocalizations are often high-pitched and why open vowels are considered “large” in sound symbolism despite their high first formant. Callers often compensate by adding vocal harshness (nonlinear vocal phenomena) to undesirably high-pitched loud vocalizations, but a combination of low and loud remains an honest predictor of both perceived and actual physical formidability. The proposed notion of a loudness-frequency tradeoff thus adds a new dimension to the widely accepted frequency code and requires a fundamental rethinking of the evolutionary forces shaping the form of acoustic signals.
Keywords: vocal communication, voice, loudness, body size, strength.

Public significance statement

Speakers can intimidate by being loud or by lowering voice frequency: low frequency appears to work best for size exaggeration, whereas loudness is more effective for displaying strength and aggression. However, it is physiologically difficult to achieve both simultaneously, making the low-and-loud combination an honest index of physical formidability and the elusive key to vocal intimidation.

The role of loudness in vocal intimidation

Since Darwin, researchers have attempted to find order in the extraordinary diversity of communicative displays produced by different species, including human nonverbal signals such as smiles or screams, by studying their evolutionary origins and functions. One of the most successful and generally accepted principles is that displays intended to demonstrate physical rigor for purposes of intimidation are often either relatively honest indices or exaggerations of the signaler's body size (Charlton & Reby, 2016). This is a powerful observation that makes sense of numerous signals because so much of nonverbal communication is about winning dominance contests and attracting mates (Bradbury & Vehrencamp, 1998). In the case of vocal communication, the core principle of acoustic size exaggeration is that low auditory frequency is robustly associated with large body size, and therefore low-frequency vocalizations are expected to express dominance and aggression, whereas high-frequency vocalizations should express submission and fear (Morton, 1977; Ohala, 1984). Here we argue that this principle, the frequency code, is correct but incomplete: there is an equally general loudness code, and the two clash because of physiological constraints on voice production that give rise to a tradeoff between being loud and being low. We begin with a review of previous work that supports this hypothesis and then fill in several important gaps with our own empirical evidence.

Frequency code

Nearly 50 years ago, Morton observed that "[b]irds and mammals use harsh, relatively low-frequency sounds when hostile and higher-frequency, more pure tonelike sounds when frightened, appeasing, or approaching in a friendly manner" (Morton, 1977, p. 855). He explained this principle by reference to the physical link between an animal's size and the rate at which sound-producing anatomical structures vibrate. Given this physical constraint, Morton concluded that low frequency sounds can only be produced by large animals and therefore sound intimidating, whereas high frequencies generally convey smallness and diminutive features. Morton's basic insight has stood the test of time (August & Anderson, 1987; Briefer, 2012; Taylor & Reby, 2010), although he subsumed several acoustic phenomena under the umbrella term "frequency" and failed to incorporate vocal harshness, beyond a vague reference to harshness being a side effect of having large and "flaccid" vocal membranes (p. 864).

Ohala (1984) picked up the frequency aspect of Morton's biological argument (ignoring harshness) and broadened its scope to size-related connotations of different vowels, melodic contours of tonal languages, prosody patterns in questions, and other cases of linguistic iconicity in human speech. In particular, Ohala argued that the frequency code used in speech prosody and human non-linguistic vocalizations had its origins in animal communication and sexual selection. Being a phonetician, he also unpacked Morton's notion of "frequency." As described by the myoelastic-aerodynamic theory of phonation (Behrman, 2021), the vocal folds are set in motion by the flow of air from the lungs, and the rate of their vibration (*fundamental frequency*, or f_0 – see Table 1 for a glossary of key terms, italicized at first mention) depends on the subglottal pressure and the tension of laryngeal muscles. This source of acoustic excitation, consisting of f_0 and its harmonics, is filtered by *resonances* of the supraglottal vocal tract, which preserves certain energy bands and dampens others. The frequencies and bandwidths of these resonances, or *formants*, are lower in longer vocal tracts, which is why lowering the larynx or otherwise elongating the vocal tract shifts all formants down in frequency (Behrman, 2021). Ohala claimed that both f_0 and formants could be lowered independently to convey large size and thus, secondarily, ritualized to communicate dominance or aggression.

Much research has since validated and refined the predictions of the frequency code. Looking across species, f_0 is typically lower in larger animals (Charlton & Reby, 2016), and there is now strong evidence from numerous studies that human listeners associate low *pitch* – the perceptual correlate of f_0 – with largeness and dominance (reviewed in Aung & Puts, 2020; Pisanski & Bryant, 2019). Yet, within species, the correlation between f_0 and size is much weaker within sex and age classes (Charlton

& Reby, 2016; Fitch, 1997), and there is ongoing debate about whether low frequency in human males has evolved as an *honest* signal of formidability or via sensory exploitation (Aung & Puts, 2020; Feinberg et al., 2018). Nevertheless, there is clearly selection pressure on males in many mammalian species, including humans, to produce lowered voice pitch, as evidenced by the dramatic sexual dimorphism in vocal fold development during puberty (Puts et al., 2016). Furthermore, formant frequencies are more reliable indices of body size compared to f_o due to stronger physical constraints on increasing the length of the vocal tract (Reby & McComb, 2003; Fitch & Giedd, 1999). As a result, there is an impressive body of research showing the importance of *vocal tract length* (VTL) (Reby et al., 2005) and its dynamic adjustments in communicating body size and aggression, including in human vocal signals (Barreda, 2016; Pisanski & Bryant, 2019; Pisanski & Reby, 2021).

In addition to lowering f_o and elongating the vocal tract, the latter of which lowers all formants, a third way to sound "low" is to use small articulatory movements that lower only one or two of the lower formants, changing *vowel quality*. In particular, the second formant (F2) appears to be important for size judgments, as when people say *hoot* or *haw* instead of *hey*: back vowels with a low F2 are often perceived as larger than the front vowels /i/ and /e/ (Ekström, 2022; Fitch, 2016; Sidhu & Pexman, 2018). This phonetic pattern sometimes manifests itself in relative frequencies of different vowels in words for small and large objects in world languages, particularly as an over-representation of /i/ in words for "small" (Blasi et al., 2016; Winter & Perlman, 2021), and even in female versus male first names (Pitcher et al., 2013). The effect of manipulating individual formants on perceived size has also been shown experimentally (Barreda, 2016; Pisanski et al., 2022). Most nonverbal vocalizations also contain one or more vowels: for example, an angry roar might sound something like /oɑə/. Thus, it is possible to strategically modulate vowel quality not only in speech, but also in nonverbal vocalizations.

More recently, it has been proposed that *nonlinear vocal phenomena*, broadly understood as deviations from regular phonation, offer a fourth mechanism for acoustic size exaggeration (Anikin et al., 2021; Fitch et al., 2002). Subharmonics, sidebands, and deterministic chaos lower the apparent pitch and make the voice sound harsh or rough (Anikin et al., 2021; Arnal et al., 2015), harkening back to Morton's original proposal that low frequency and harshness work together in agonistic calls.

Overall, the frequency code has provided a coherent and extremely powerful unifying framework for the study of vocal behaviors ranging from speech prosody (Ohala, 1984) to the roars of deer stags (Reby et al., 2005). Yet, there are many puzzling observations that defy predictions, such as the extremely high f_o in aggressive screams (Anikin & Persson, 2017; Green et al., 2011; Leinonen et al., 1991; Owings & Morton, 1998; Slocombe & Zuberbühler, 2005), low f_o in quiet polite speech (Winter & Perlman, 2021), or surprisingly high F1 in "large" vowels like /ɑ/ (Newman, 1933; Pisanski et al., 2022; Sapir, 1929; Sidhu & Pexman, 2018). Naturally, any general constraint on the physical form of communicative signals is probabilistic, so exceptions are expected, but we suggest that they follow a systematic pattern indicating another major force at play: *loudness*.

Why loudness matters

The loudness of vocal signals has traditionally been sidetracked in research on vocal intimidation and dominance in favor of spectral characteristics like f_o and formant frequencies (or apparent VTL), which are both easier to measure in field recordings and more robust to degradation due to background noise and propagation over (often unknown) distances. While research on these voice frequencies has led to major advances in bioacoustics and the voice sciences (Briefer, 2012; Taylor & Reby, 2010), we see several theoretical reasons why loudness should also play a role in the communication of size and aggression, as outlined below.

1. Loud is large. The positive association of low auditory frequency with large size may plausibly be a learned, statistical correspondence: because larger objects vibrate at lower frequencies, an animal learning about its environment from observation could gradually associate low frequency with large size (Sidhu & Pexman, 2018; Spence, 2011). Likewise, larger sources of sound in the environment are typically louder. Within our hearing range of frequencies, subjectively perceived

loudness is primarily determined by sound intensity, which is proportionate to the pressure gradient created by a moving object. In the case of a vibrating sound source, such as the vocal folds, fluctuations in air pressure are determined by the maximum acceleration rather than size of the vibrating object (Titze, 2000). However, increasing the moving area is conceptually similar to multiplying the number of sound sources and therefore their cumulative acoustic power, which translates into higher intensity as multiple sound sources converge upon a receiver as long as the wavelength is much greater than the source, so that there is no destructive interference. Furthermore, larger objects can both absorb and release more energy as they move around and therefore make louder impact sounds (Grassi, 2005). The proportion of kinetic energy that can be converted into sound, or radiation efficiency, also increases with size, which is particularly critical at low frequencies (Jakobsen et al., 2021). In the animal world, surprisingly, the maximum observed loudness of a vocal signal in the air is independent of body size and peaks at about 120 dB at a distance of 1 m in various clades (Jakobsen et al., 2021), but the *average* loudness does scale with body size (Gillooly & Ophir, 2010).

A second source of the loudness-size association is purely perceptual: louder sounds are mapped onto larger stimuli, and vice versa, based on matching the magnitudes of perceptual inputs. Such magnitude-based or prothetic cross-modal correspondences are pervasive, robust, and reliably developing, and loudness has been repeatedly shown to be associated with other quantitative dimensions across modalities – for example, with visual brightness or contrast (Anikin & Johansson, 2019; Spence, 2011). There is also some experimental evidence that loudness is perceptually coupled with size. Both preschoolers and adults explicitly matched louder pure tones to longer objects (Teghtsoonian, 1980) and to larger animals (Smith & Sera, 1992), although Knoeferle et al. (2017) found that loudness did not affect the way consonant-vowel syllables were matched with visual shapes of different sizes. In a recent cross-cultural study, both deaf and hearing children and adolescents increased the duration and loudness of their nonverbal vocalizations to refer to the larger of two novel objects; this magnitude matching was more consistent than the use of frequency cues, and listeners relied on it to guess the intended reference (Perlman et al., 2022).

2. Loud is fit. Agonistic vocalizations can be understood as a means to negotiate conflicts through mutual assessment to avoid a physical confrontation. Thus, it is of central importance for receivers to focus on honest indices of formidability (Bryant, 2020; Maynard Smith et al., 2003; Sell et al., 2010). Increasing loudness offers one way to demonstrate physical fitness and stamina because loud vocalizations are metabolically costly (Demartsev et al., 2019) and can typically only be produced by healthy individuals with strong abdominal muscles (Tsai et al., 2010). Empirically, vocalizations produced in dominance contests are typically louder than other calls in deer (Reby et al., 2005), baboons (Fischer et al., 2004), and humans (Šebesta et al., 2019). Loudness also predicts perceptual ratings of formidability in the roars produced by contestants in mixed martial arts (Šebesta et al., 2019), but acoustic predictors of actual physical strength in the human voice remain elusive (Kleisner et al., 2021). The most commonly investigated variable is f_0, but it is at best a weak predictor of human physical strength (Aung & Puts, 2020; Valentova et al., 2019), height, and weight (González, 2006; Pisanski et al., 2014) when sex and age are controlled for. This motivated us to test whether loudness is a better predictor of physical formidability than is voice f_0.

3. Loud is aversive and alarming. The key objective of vocal displays of dominance and aggression is to intimidate receivers, and loud sounds can be frightening. Rapid increases in loudness are highly salient acoustic events that involuntarily attract attention as an evolutionarily conserved adaptation for detecting approaching or looming hazards (Ghazanfar et al., 2005; Neuhoff, 2018; Tajadura-Jiménez et al., 2010). Sudden loud sounds may trigger an unconditioned acoustic startle reflex (Lang et al., 1990) or more generalized defense mechanisms (LeDoux, 2012; Tajadura-Jiménez et al., 2010). Even without the surprise factor, loud sounds can be unpleasant, and even painful beyond some threshold; thus, vocalizations such as harsh, loud screams can directly repel receivers by virtue of being

physically aversive (Owren & Rendall, 1997). As such, the intrinsically threatening nature of loud vocalizations makes them suitable for vocal intimidation.

4. Loud is harsh. Morton (1977) highlighted vocal harshness as a crucial marker of aggression, and it has since been confirmed that harshness enhances the perception of both aggression and large size in mammal vocalizations, including those of humans (Anikin et al., 2021; Briefer, 2012). A voice is perceived as harsh when it deviates from regular periodic phonation (Anikin et al., 2021; Fitch et al., 2002), and the production of most *nonlinear vocal phenomena* requires high subglottal pressure (Fitch et al., 2002; Herzel et al., 1995). Assuming that nonlinearities do not cause a noticeable drop in voice loudness (Lagier et al., 2017), high vocal effort could thus be conducive to vocal intimidation not only as a direct effect of loudness per se, but also indirectly via the production of harsh-sounding nonlinear phenomena.

5. Loud is prepared for physical conflict. Voice loudness is determined primarily by subglottal pressure (Lagier et al., 2017), and increased subglottal pressure requires powerful activation of abdominal muscles (Johnstone & Scherer, 1999). Thus, loud vocalizations produced in agonistic contexts become honest signals of preparedness for physical conflict (Tsai et al., 2010; but cf. Zahavi, 1982). Indeed, vocalizations produced during intense physical effort tend to be loud and relatively high-pitched (Anikin, 2023; Emanuel & Ravreby, 2022). Elevated lung pressure may also function to increase the integrity of the shoulder girdle in preparation for vigorous activity (Pouw & Fuchs, 2022), which means that it would be physically awkward to produce quiet (or, in fact, tonal and low-frequency) vocalizations while engaged in, or preparing for, a physical confrontation. By the same logic, submissive and affiliative vocalizations can be expected to be comparatively quiet and tonal, revealing that the animal is relaxed or adopting a submissive posture.

The loudness-frequency tradeoff

To summarize the argument so far, the frequency code predicts that a vocalizer wishing to intimidate by displaying or exaggerating their body size should lower their voice frequency. When we unpack what "frequency" means, the prediction includes lowering f_o, lowering all formants by elongating the vocal tract (increasing VTL), producing a vowel with the first few formants shifted as far down as possible – ideally, a closed, back, rounded /u/ – and perhaps creating additional low-frequency components such as subharmonics or amplitude modulation. Now consider the loudness code: a vocalizer trying to intimidate should also be as loud as possible (see Table 2 for a summary of hypotheses). The problem is that low frequency and loudness are difficult to achieve at the same time. In fact, of the four frequency components, loudness is synergistic with only one (nonlinear vocal phenomena) and directly conflicts with at least two (f_o and closed vowel quality), and possibly three (vocal tract elongation with rounded lips).

The most obvious clash between frequency and loudness is that loud vocalizations tend to be high-pitched because increasing subglottal pressure and vocal effort, which is necessary for being loud, simultaneously raises f_o (Behrman, 2021; Titze, 2000). Of particular relevance to vocal intimidation, voice range profiles demonstrate that the *maximum* achievable loudness tends to increase as f_o rises, at least up to some threshold (Gramming & Sundberg, 1988; Hunter & Titze, 2005; Lamesch et al., 2012; Titze, 1992). Another reason to increase f_o if loudness is a priority is that radiation efficiency (i.e., the proportion of generated acoustic power that leaves the mouth) improves at higher frequencies. Much less acoustic energy is reflected back to the mouth, and more energy escapes as radiated sound when the wavelength is comparable to the diameter of the mouth or beak (Titze & Palaparthi, 2018). According to different mathematical models, the optimal frequency range for sound radiation in humans is about 450 Hz (Fletcher, 2004) or even higher, above 1 kHz (Titze & Palaparthi, 2018), although direct measurements that could validate these calculations are lacking. Finally, increasing vocal effort not only raises f_o, but also amplifies its harmonics (Mittal & Yegnanarayana, 2013; Stout, 1938; Traunmüller & Eriksson, 2000), potentially raising the perceived frequency even higher than expected from the changes in f_o alone. Interestingly, while the covariance between pitch and loudness is

well established in voice science, not all empirical evidence is consistent with the theory. Call types classified as high-amplitude did not have a higher average f_o in a meta-analysis of mammal vocalizations (Gustison & Townsend, 2015), and professional singers are trained to maintain stable loudness across a wide pitch range, just as they are taught to control the larynx position in specific ways depending on the musical tradition (Shipp, 1987; Sundberg & others, 1977). More evidence from real-world vocalizations is needed. Unfortunately, voice loudness is seldom reported, especially in recordings made outside of the lab, because it is more technically challenging to measure than are spectral characteristics, requiring both a calibrated microphone and an accurate, precise estimate of distance from the vocalizer.

The second reason for a loudness-frequency tradeoff is that the efficiency of sound radiation depends not only on its frequency, but also on the size of the aperture (Titze & Palaparthi, 2018). Indeed, loud vocalizations are often produced with the mouth (Mercer & Lowell, 2020) or beak (Ohms et al., 2012) open wide. This can potentially affect both the overall VTL (and thus all formant frequencies) and the vowel quality (i.e., specifically the relative positions of the first few formants). The physical length of the vocal tract is usually described as decreasing with mouth opening, especially if the corners of the mouth are retracted (Titze, 2000), which is further exacerbated by the tendency to raise the larynx on high-pitched notes (Shipp, 1987). As long as the aperture formed by the lips and teeth is narrower than the mouth cavity, the acoustically effective VTL also decreases in proportion to the square root of mouth opening. Furthermore, human vocalizers have not only to spread the lips, but also to lower the jaw in order to open the mouth really wide, which raises F1 and produces an a-like vowel sound (Sundberg, 1977; Titze, 2000). This suggests a possible explanation for the fact that /ɑ/ is consistently listed among the “large” vowels – in fact, the original study by Sapir (Sapir, 1929) contrasted *mil-mal* rather than, for instance, *mil-mool*, although the formants in /u/ are much lower than in /ɑ/. In other words, F2 may be more directly involved in *sound symbolism* of size (the /ie/ vs. /ɑou/ contrast) compared to F1 (the /iu/ vs. /ɑ/ contrast) because lowering F1 sacrifices loudness, which counteracts the intimidating effect of lowering voice frequency. In fact, the “darkest” *voice quality* can be achieved by closing the mouth altogether and saying *mmm*, but the resulting sound pressure level will obviously be very low. Among open-mouth vowels, /a/ should theoretically be the loudest, and this is borne out by the few published reports of voice range profiles in singers (Behrman, 2021; Gramming & Sundberg, 1988; Hunter & Titze, 2005; Lamesch et al., 2012). The evidence is incomplete, however, as the ranking of other vowels depends on f_o and singing mode (falsetto vs. chest voice), and voice range profiles are different from ecologically relevant aggressive vocalizations.

The crux of this built-in tradeoff between loudness and frequency (f_o and the first formant) is that a vocalizer may not be able to maximize loudness while simultaneously maintaining low frequency or closed vowel quality, necessitating a strategic choice of the mechanism of vocal intimidation in a particular situation. This is the central question that we address empirically in this paper. After asking participants to sound small or large, weak or strong, and submissive or aggressive, we measure both voice loudness and frequency components of their nonverbal vocalizations. The main findings are shown alongside our predictions in Table 2. The effectiveness of various strategies of vocal intimidation is then compared in a series of perceptual experiments, in which we play back the recordings and ask listeners to evaluate the physical characteristics and motivations of the speakers.

Table 1

Glossary of key terms as they are used in this paper

Formant	A vocal tract *resonance* or a prominent spectral peak resulting from a resonance. The relative positions of the first three formants (F1 to F3) determine the *vowel quality*, whereas formant spacing can be used to estimate apparent *vocal tract length (VTL)* (Behrman, 2021) and thus body size (Fitch & Giedd, 1999).

Fundamental frequency (f_o)	The lowest frequency at which a periodic signal is repeated, measured in Hertz (Hz). For the vast majority of voiced signals, f_o corresponds to the rate at which the vocal folds are vibrating and is the perceptual correlate of ***pitch***. Longer and/or looser vocal folds vibrate more slowly and produce a relatively lower f_o than do shorter and tenser vocal folds (Titze, 2000).
Honesty	***Honest*** or reliable signals provide true information about the signaler. Honesty may be ensured because a signal is a hard-to-fake index of the signaler's condition (e.g., physical size) or because it constitutes a costly handicap (Maynard Smith et al., 2003). For example, formant dispersion in deer stags is considered an index of the animal's size because it is determined by the animal's maximum ***vocal tract length***, which is achieved when the movable larynx is lowered all the way to the sternum (Reby et al., 2005).
Loudness, dB (SPL)	"Objective" loudness in decibels of sound pressure level, a physical measure of the amplitude of pressure fluctuations in a sound wave, here calculated as root mean square of an audio recording calibrated to correspond to sound pressure level at 1 m from the sound source.
Loudness, sone	"Subjective" loudness in sones, a perceptual measure of how loud a vocalization sounds when played back at a particular sound pressure level, which is estimated with respect to equal loudness curves with summation across critical frequency bands (Fastl & Zwicker, 2006).
Nonlinear vocal phenomena	Deviations from regular phonation such as frequency jumps (sudden changes in f_o), sidebands (amplitude modulation of the glottal source by a low-frequency oscillator higher up the vocal tract), subharmonics (irregular vibration of the vocal folds, which produces a weaker secondary frequency at an integer fraction of f_o), and deterministic chaos (non-periodic vibration of the vocal folds) (Anikin et al., 2021; Fitch et al., 2002). In this study we also analyzed vocal fry, also known as pulse or strohbass register – a mode of phonation in which the vocal folds mostly remain closed and produce only occasional, irregular pulses (Roubeau et al., 2009).
Pitch	The perceived highness or musical tone of a sound. In voiced signals, pitch is mainly determined by the ***fundamental frequency*** (f_o), wherein perceived pitch increases roughly logarithmically with f_o.
Resonance	In physics, the natural frequency of an oscillator at which an external force produces maximum response. In voice science, the frequency band that is preferentially transmitted by the vocal tract (see ***formant***).
Sound symbolism	Cases of iconic, non-arbitrary sound-meaning associations, as when *mal* is judged to be a better name for large objects, and *mil* for small objects (Sapir, 1929; Spence, 2011).
Vocal tract length (VTL)	The length of the airway from the vocal folds to the aperture through which the sound is radiated into the environment (e.g., the mouth, nostrils, or beak). Formant frequencies scale linearly and inversely with VTL: elongating the vocal tract by 10% lowers formants by 10% (Behrman, 2021). A common way to estimate the apparent VTL from formant frequencies, also used here, is to use the regression method (Reby et al., 2005).
Voice quality	Any perceptual characteristic that distinguishes two voices at the same loudness and pitch, by analogy with *timbre* in musicology. Two aspects of voice quality relevant to this study are ***vowel quality*** and the amount of energy in harmonics. A voice with strong harmonics of f_o, and thus a lot of high-frequency energy in the spectrum, has a bright or brassy quality, whereas a voice with weak harmonics may sound dark, breathy, or fluty.
Vowel quality	Formants are equidistant in a cylindrical vocal tract, which corresponds to the neutral *schwa* vowel /ə/ (all phonetic symbols are taken from the International Phonetic Alphabet). When articulatory movements change the shape of the vocal tract, the lower formants shift around, and different vowels are produced. Because absolute formant frequencies depend on ***vocal tract length***, and thus vary across speakers, vowel quality is here operationalized as speaker-normalized F1 and F2 relative to *schwa*. Open or low vowels have a higher F1 than do closed or high vowels (/aɑ/ vs. /iu/); front vowels have a higher F2 than do back vowels (/ie/ vs. /uo/).

Table 2

Summary of predicted and observed voice production strategies

predicted / observed		Size		Strength		Aggression	
		Small	Large	Weak	Strong	Submissive	Aggressive
Loudness code	Voice loudness	↓	↑	↓	↑	↓	↑
		–	↑ ✔	↓ ✔	↑ ✔	–	↑ ✔
Frequency code	Voice pitch	↑	↓	↑	↓	↑	↓
		↑ ✔	–	↑ ✔	↑ ✖	↑ ✔	↑ ✖
	Vocal tract length (lower formants = greater apparent VTL)	↓	↑	↓	↑	↓	↑
		↓ ✔	↑ ✔	–	↑ ✔	–	–
	Vowel quality	[e] [i]	[u] [o]	[e] [i]	[u] [o]	[e] [i]	[u] [o]
		[i] ✔	–	–	–	–	–
	Nonlinear vocal phenomena	↓	↑	↓	↑	↓	↑
		–	Chaos, ✔ sidebands	Vocal fry ✖	Chaos, ✔ sidebands	–	Chaos, ✔ sidebands

↑ increased ↓ decreased ✔ evidence for ✖ evidence against – no evidence either for or against

Materials and Methods

Experiment 1: Voice Production

Vocalizers

Vocalizers (N = 72, 40 F + 32 M, age 18 to 60, mean age 30) were recruited and recorded at the ENES Bioacoustics Research Lab in Saint-Etienne, France (n = 38) and at Lund University in Lund, Sweden (n = 34). They were students or junior staff members with no professional training in singing or drama, who participated in the study voluntarily and received no formal compensation. Samples sizes of vocalizers per condition were as follows: body size n = 20 (10 M + 10 F), strength n = 27 (12 M + 15 F), aggression n = 25 (10 M + 15 F). In addition to obtaining vocalizer sex and age, we measured grip strength three times in each hand using a dynamometer (model Baseline 12-0241 LITE). All participants provided informed consent and were debriefed after the recording. Ethical approval for performing perceptual experiments with human subjects was provided by the Comité d'Ethique du CHU de Saint-Etienne (IRBN692019/CHUSTE).

Recording Procedure

Following baseline recordings of neutral speech, speakers were asked to produce a number of vocalizations following on-screen instructions, as summarized in Table 3. The instructions were provided in English or French. Prior to the experiment, participants were asked to always produce voiced sounds rather than, for example, whistles, whispers, or hand claps. Verbal information, including personal details in the baseline recording, was not coded as only nonverbal measures were extracted from speech.

Vocalizers were alone in a room (anechoic chamber at the ENES lab, quiet office space at Lund University) and followed the instructions on a laptop screen, with a dB meter on the side of the screen providing feedback in real time for the *loud* condition, in which participants were asked to drive the meter into red in order to encourage them to overcome inhibitions and be really loud. To ensure adequate quality and avoid clipping over a large range of sound pressure levels, we used multiple recording channels and a two-stage calibration scheme, as follows. At both settings, participants wore a

tie clip microphone (55 dB gain), and we also used a stationary audio recorder placed at a fixed distance of approximately 1.5 m from the speaker (ENES: Tascam DR-44WL, set at 44.1 kHz and 16-bit amplitude resolution, using two built-in channels at 20 dB and 40 dB gain; Lund: Tascam DR-05 with a single channel at 0 dB gain). The stationary recorder was calibrated once by recording a steady tone, whose intensity was measured with a standard sound level meter (Rion NL-52; dB-C) at a distance of 1 m from the sound source. This provided a global anchor for converting RMS amplitude in the calibrated channel to dB-C at 1 m. We set the tie clip microphone at about 20 cm from the participants' mouth, but because its exact placement varied between participants, a calibration signal (a single syllable without clipping in any channel) was also extracted for each participant, and the difference between the channels on this calibration signal was used to adjust dB values across channels relative to the "anchor" channel. Because small variation in the physical setup existed between participants and especially between the recording sites (ENES vs Lund), the absolute values of loudness in dB are only an approximation, but, crucially, their change between conditions within a single participant is much more consistent.

Table 3

Recording Protocol in the Body Size Condition

1. Baseline neutral voice recording. **Baseline 1** "Please read aloud this sentence in your natural voice: Where were you a year ago?" **Baseline 2** "Please introduce yourself in your natural voice: Tell us your name, age, employment"	Acoustic benchmark when analyzing the change in loudness, f_0, etc.
2. Sound as small as possible (three takes) **Small** "Make any vocal sound to make yourself seem as small as possible"	Spontaneous strategies for conveying size, strength, and aggression before being cued to be loud
3. Sound as large as possible (three takes) **Large** "Make any vocal sound to make yourself seem as large as possible"	
4. Sound as small as possible NONVERBAL (three takes) **Small nonverbal** "Again! This time use a non-verbal vocal sound to make yourself seem as small as possible"	Like #2-3 above, but ensuring the production of purely nonverbal vocalizations needed for formant analysis
5. Sound as large as possible NONVERBAL (three takes) **Large nonverbal** "Again! This time use a non-verbal vocal sound to make yourself seem as large as possible"	
6. Vocal warm-up **Ramp 1** "Make a series of 5 voiced sounds, from the quietest to the loudest that you can comfortably produce (please don't hurt your throat!)" **Loud** "Please drive the sound level meter into the red - you have to get loud to do that!" **Ramp 2** "Now, once again, please make a series of 5 voiced sounds, from the quietest to the loudest"	The effect of forced changes in loudness on pitch and vowel quality
7. **Fight/free** (three takes) "You are in a fight. Without using words, yell as loudly as you can to make yourself seem as large as possible" [+visual aid] 	Cued to be loud in a ecologically relevant scenario of vocal intimidation; used for predicting physical strength from acoustics and for testing the effect of voice acoustics on apparent formidability (Experiments 4 & 5)
8. **Fight/vowel** u/a/o/e/i/ɯ (three takes per vowel, vowels in random order) "You are in a fight. Yell this vowel as loudly as you can to make yourself seem as large as possible" [+visual aid as in Task 7]. Vowels explained to speakers as follows: [u] as in f**oo**d, sh**oe** [+ audio recordings of each vowel]	Used for testing whether vowels have intrinsic loudness and pitch and for testing the effect of voice acoustics on apparent formidability (Experiments 4 & 5)

[a] as in m**a**mma, p**a**ppa
[o] as in fl**aw**, p**aw**
[e] as in p**e**n, B**e**n
[i] as in m**e**, degr**ee**
[ʉ] as in French T**u** as v**u** **u**ne...?

Note. *Strength* and *aggression* conditions only differed in the key words: *aggressive* / *submissive* or *strong* / *weak*, respectively, instead of *large* / *small.*

Acoustic Analysis

The recordings were manually segmented into vocalizations (N = 3279) by saving each take from the loudest input channel without clipping. The boundaries between takes were not always obvious because participants often vocalized several times before or while pressing the button to move on to the next page of instructions. When a participant repeated qualitatively the same vocalization several times in a single take, this was saved as a single recording; qualitatively different vocalizations were saved and analyzed separately, occasionally resulting in more than three takes per condition. Likewise, sometimes participants skipped a question or provided invalid output (e.g., sounds without phonation such as hisses or clicks). Speech produced in the "nonverbal" conditions was saved, but tagged as such so it could be excluded from the relevant data analyses (e.g., of vowel quality). Each of about five syllables in vocal ramps was saved as a separate file.

All vocalizations were analyzed acoustically with the function *analyze* from R package *soundgen* 2.5.1 (Anikin, 2019) using pre-extracted, manually corrected (in the *pitch_app* interactive environment) contours of the fundamental frequency. The RMS amplitude was converted to approximate sound intensity at 1 m using the calibration signals (see Procedure), which in turn was fed into the *getLoudness* function in order to estimate the subjective (perceived) loudness of each vocalization in sone units (Fastl & Zwicker, 2006). Formants F1-F4 were extracted manually from 2425 out of 3527 vocalizations by two independent expert coders (AA and DV); any identified cases of disagreement were discussed and resolved by the two coders. The formants could not be measured reliably when f_o was too high, and we further excluded from the formant analysis any short verbal or emblem-like exclamations or interjections, such as *Hey*, because their phonemic content dictates the choice of vowel and also skews estimates of VTL. Closed-mouth vocalizations were also excluded because their vowel quality (*mmm*) is not directly comparable to that of open-mouth vowels (Titze, 2000). For all these reasons, formants could not be reliably measured in 1102 recordings. In contrast, longer speech fragments (e.g., in the baseline conditions) were included in the formant analysis by taking the average value of F1-F4 over the entire utterance. The measured formant frequencies in Hz were converted to speaker-normalized values by estimating the VTL with the regression method (Reby et al., 2005) and then calculating the difference between the observed frequencies and their predicted values in a schwa vowel sound – that is, in a cylindrical vocal tract of the same length. To be independent of VTL and formant number, these differences were expressed as proportions of formant spacing. Finally, nonlinear phenomena were annotated by one coder (AA), who manually labeled frequency jumps and segments with subharmonics, amplitude modulation, chaos, or vocal fry.

Experiments 2-5: Voice Perception

We conducted four perceptual experiments. An implicit association test (Experiment 2) was used to confirm the existence of a cross-modal correspondence between loudness and physical size. Three perceptual rating studies (Experiments 3-5) were then conducted to investigate the effectiveness of different voice production strategies in Experiment 1. First, we played back the recorded stimuli and asked listeners to rate the unseen speakers on size, strength, and aggression (Experiment 3). In two additional experiments, the loud aggressive vocalizations from the *fight/free* and *fight/vowel* conditions

were rated on size, strength, aggression, and formidability either at the original loudness (Experiment 4) or after being normalized to have the same peak amplitude (Experiment 5).

Each experiment was performed by a new, independent sample of participants. Playback experiments 2-5 included data from a total of 498 listeners (study-specific sample sizes and demographics given below). All participants were recruited on http://prolific.co and paid for their time. Demographic information was collected at the beginning of each experiment as self-reported age and sex (chosen from three options: *Male*, *Female*, and *Unspecified*). Sample sizes were chosen to ensure sufficient precision of estimated effects at the relevant level in Bayesian multilevel models. For the implicit association test, 20 to 30 participants provided good precision in previous research (Pisanski et al., 2022). For perceptual ratings tasks (Experiments 3-5), the choice of sample size is more complicated as precision depends on the number of speakers, the number of stimuli per speaker, the number of times each stimulus is rated on each scale, and the consistency of ratings across participants and stimuli. In every case, the precision of population-level effects was adequate for accurately describing all substantively non-trivial effects of interest.

Experiment 2: Implicit Association Test

We implemented a web-based version of the implicit association test as described in (Parise & Spence, 2012), which we successfully used with vocal stimuli in earlier studies (Anikin & Johansson, 2019; Pisanski et al., 2022). Listeners were required to learn a rule associating the left arrow on a keyboard or touchscreen with one image and sound, and the right arrow with another image and sound. For example, in one block of trials the small image and the quiet sound might be assigned to the left arrow key, and the large image and the loud sound to the right arrow key. In the next block, the rule would change, and all four possible combinations would recur in random order in multiple blocks throughout the experiment. The goal was to compare response times and accuracy in blocks with congruent vs. incongruent combinations of stimuli assigned to the same button.

To ensure that the results were replicable, two pairs of acoustic stimuli were tested on two independent samples of participants: an aggressive human nonverbal vocalization from the voice production study (n = 30 listeners, 14 F + 16 M, age 18 to 59, mean age 27) and a synthetic animal call from an earlier study (Pisanski et al., 2022) (n = 21 listeners, sex and age not recorded). One stimulus in each pair was normalized to maximum amplitude, and the other to -12 dB. The visual stimuli were line drawings that differed only in size: the large icon was twice the size of the small one, and thus four times larger in surface area. Once the participant had understood the procedure and achieved an accuracy of 75% or better in two practice blocks, they proceeded to complete 16 test blocks of 16 trials each. As each trial began, a fixation cross was shown in the middle of the browser screen for a random period of 500-600 ms. After a delay of 300-400 ms the stimuli were presented. Visual stimuli were shown for 400 ms in the same location as the fixation cross against a uniform white background; the sounds were about 500 ms in duration. If the response of the listener was correct, the next trial began immediately. If the response was incorrect, a red warning cross was flashed for 500 ms before proceeding to the next trial.

Experiment 3: Playback of All Conditions

To test how well speakers had managed to convey the intended meanings (body size, strength, aggression, or vocal intimidation), we tested a sample of nonverbal (speech excluded) vocalizations under 5 s in duration from all conditions except for vocal ramps, presented at the same relative loudness as they were produced, but compressed into approximately half the original dynamic range (40 dB instead of ~80 dB). Thus, the loudest vocalization was normalized to have a peak amplitude of 0 dB (as loud as possible), and the quietest one -40 dB. Normalization was necessary because the recordings were obtained from different microphones with different gain levels intended to capture both very quiet and very loud vocalizations without clipping or loss of quality. However, the full original dynamic range of over 80 dB proved impossible to play back through the headphones in the

same experiment: either the quieter recordings were inaudible, or the louder ones became painful; therefore, we compressed the dynamic rage to 40 dB. To reduce the number of stimuli to test, only a single take per person was included from the *fight/free* and *fight/vowel* conditions, for a total of 1055 stimuli.

Three representative sounds were presented before the first test trial as examples to familiarize the listeners with the kind of vocalizations they would rate. To ensure that the sound playback was working and that the volume setting was not too low, participants also had to enter a spoken three-digit code, which had the same amplitude as the quietest stimulus. Participants were instructed not to change their volume setting throughout the remainder of the experiment. Ratings were provided on a horizontal visual analog scale from 0 to 100 (transformed to 0 to 1 for the analysis), and each sound could be replayed an unlimited number of times. Every participant rated a random subset of 100 stimuli in two blocks of 50 trials. Each block corresponded to one of three rating scales: size (labeled *small* to *large*), strength (*weak* to *strong*), and aggression (*submissive* to *aggressive*). Thus, each participant rated 50 randomly chosen sounds on one randomly chosen scale, and then another 50 on a different scale. 125 participants completed at least 20 trials each and were included in the analysis (62 F + 63 M, aged 19 to 65, mean age 29), rating each sound ~3.8 times on each of three scales.

Experiment 4: Playback of Fight Conditions at the Original Loudness

While in Experiment 3 we solved the problem of excessive dynamic range by compressing it, in Experiment 4 we preserved the original dynamic range, but only tested the relatively loud vocalizations in the *fight/free* and *fight/vowel* conditions, again randomly selecting a single take per condition and per subject (n = 599 stimuli). This made it possible to test the effects of loudness and other voice properties on perceived speaker characteristics. One sample of listeners provided ratings on the formidability scale (*Not at all intimidating* to *Extremely intimidating*), presented and illustrated in the same way as in the original elicitation, namely a physical confrontation. A second sample of listeners rated the same vocal stimuli on size, strength, and aggression, as in Experiment 3. In total, 162 listeners (66 F + 95 M + 1 unreported, age 19 to 59, mean age 29) rated each sound 6.7 times on each of the four scales.

Experiment 5: Playback of Fight Conditions after Normalization

As a more powerful alternative to statistically controlling for loudness when analyzing the effect of pitch and vowel quality on perceived formidability, in this experiment we simply replicated Experiment 4 after removing all variation in loudness. Namely, we tested the same 599 stimuli from the *fight/free* and *fight/vowel* conditions, but this time after normalizing them to have the same peak amplitude. A total of 160 listeners (one sample for formidability, another for size, strength, and aggression; 77 F + 83 M, age 14 to 69, mean age 30) rated each vocalization 6.7 times on each of the four rating scales.

Data Analysis

Data were analyzed with Bayesian multilevel models using the R package *brms* 2.17.0 (Bürkner, 2017) with mildly informative conservative priors. All estimates shown in the text are medians of posterior distributions and 95% credible intervals. Unbounded continuous variables, such as pitch or VTL, were analyzed with Gaussian models, and bounded ratings from perceptual experiments with beta models (rescaled to range from 0.01 to 0.99). Details on the structure of each model are given in the text and figure legends.

Transparency and Openness

Vocal stimuli (except for background speech with non-anonymizable personal information), datasets, and R scripts for reproducing all analyses can be downloaded from https://osf.io/ngwcp/.

Results

Strategies for Conveying Size, Strength, and Aggression

We began by analyzing how adult human speakers (N = 72) modified their voices when asked to exaggerate or minimize their apparent physical size, strength, and aggression, recording a separate sample of participants for each of these three questions. Three key voice characteristics (loudness, fundamental frequency f_o, and vocal tract length VTL) were analyzed in separate mixed models as a function of experimental condition, allowing the effect of condition to vary across participants. Because of marked sexual dimorphisms in these voice characteristics, we included vocalizer sex as a covariate.

Participants consistently increased both loudness and f_o to maximize their apparent size (conditions *large* vs. *small*), strength (*strong* vs. *weak*), and aggression (*aggressive* vs. *submissive;* Fig. 1C). *Large*, *strong*, and *aggressive* conditions were all associated with markedly greater loudness compared to their baseline neutral, relaxed voice (+11.7 dB [7.5, 15.8], +13.7 dB [10.2, 17.1], and +15.5 dB [13.4, 17.7], respectively). When speakers were explicitly instructed to be as loud as possible in the *fight/free*, *fight/vowel*, and *loud* conditions, voice loudness increased further, by ~25-30 dB compared to baseline (Fig. 1B). In contrast, loudness did not change noticeably when speakers tried to sound *small* (-0.5 dB, 95% CI [-3.8, 3]) or *submissive* (-1.1 dB [-3.8, 1.6]) and decreased in the *weak* condition (-5.8 dB [-9.7, -2.1]).

F_o increased compared to baseline in all conditions. As expected, speakers greatly raised f_o to sound *small* (+16 semitones [12.6, 19.6]), and to some extent, also to sound *weak* (+3.4 semitones [1.3, 5.3]) and *submissive* (+5.2 semitones [3.4, 7.1]), but there was notable individual variability in the magnitude and even direction of these f_o changes (violin plots in Fig. 1B). Contrary to the predictions of the frequency code, f_o was also raised compared to baseline in the *strong* (+7.2 semitones [5.2, 9.2]), *aggressive* (+7.0 semitones [5.6, 8.4]), and even *large* conditions (+2.6 semitones [0.3, 4.9]). Looking at large-small, strong-weak, and aggressive-submissive contrasts (Fig. 1C), only size exaggeration involved the pitch contrast predicted by the frequency code: f_o was more than an octave (13.4 semitones [9.5, 17.3]) higher for *small* vs. *large*, although in both cases it was above baseline. Likewise, participants raised their voice pitch by over an octave compared to baseline when instructed to be *loud* (+14.6 semitones [13.0, 16.3]) or to intimidate an opponent in the *fight/free* (+15.0 semitones [13.6, 16.5]) and *fight/vowel* (+13.3 semitones [12.0, 14.6]) conditions.

We also observed some of the predicted changes in apparent VTL, which was estimated from resonance frequencies (formant spacing) and thus does not necessarily correspond to the true anatomical length of the vocal tract. The vocal changes were particularly clear when participants attempted to project a *small* (-1.0 cm [-2.1, 0.1]) or *large* (+1.8 cm [1.1, 2.5]) body size (Fig. 1B), creating an impressive difference of up to 2.9 cm in apparent VTL between the large and small conditions (95% CI [1.7, 4.0]; Fig. 1C). Participants thus raised their formants and shortened their vocal tract to sound small, and did the opposite to sound large, corroborating prior research (Pisanski & Bryant, 2019). There was also a weak and uncertain tendency to extend the vocal tract when trying to sound *strong* rather than *weak* (+0.6 cm [0.0, 1.3]) and *aggressive* rather than *submissive* (+0.6 cm [-0.1, 1.3]). The change from baseline was highly uncertain in the *fight/free* condition (+0.4 cm [-0.1, 0.8]), and any changes of VTL in the *fight/vowel* condition are hard to interpret because the apparent VTL depends strongly on the vowel.

Relating the observed vocal strategies to our predictions (Table 2), the increase in loudness for vocal intimidation was clearly confirmed for all three dimensions: size, strength, and aggression. In contrast, f_o and VTL changes predicted by the frequency code mostly occurred when the task was to convey size. If anything, f_o was e*levated* in the loud vocalizations conveying strength and aggression.

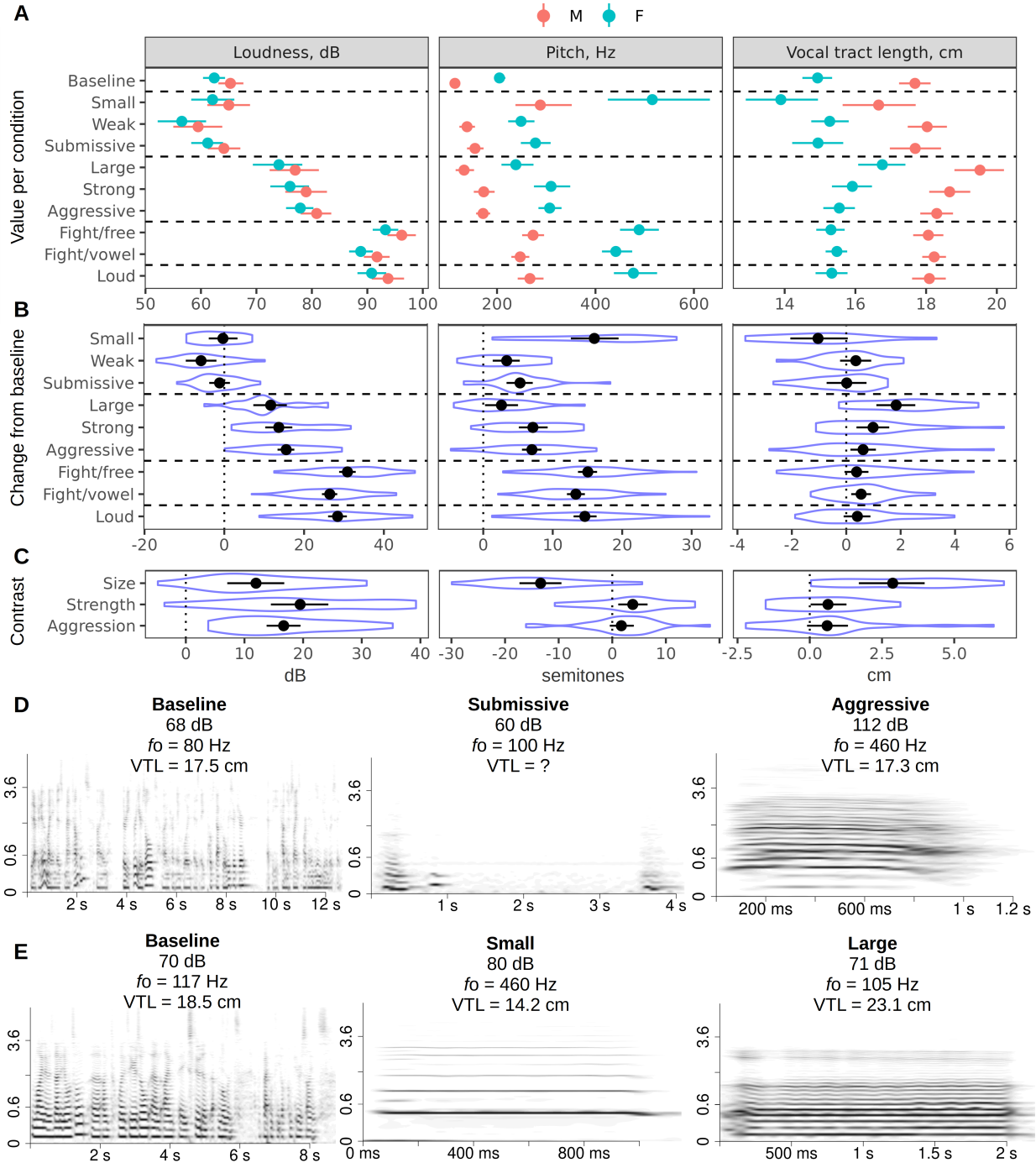

Figure 1 Production strategies for vocal size exaggeration and intimidation.
(**A-C**) Speakers consistently vocalize more loudly to convey large body size, strength, and aggression, whereas pitch and VTL are modulated by speakers primarily for size exaggeration, but not for conveying strength and aggression. Fitted values from mixed models for voice loudness, pitch, and estimated VTL per condition (A), change from baseline (B), and contrasts between pairs of conditions (C): size = *large* vs. *small*, strength = *strong* vs. *weak*, aggression = *aggressive* vs. *submissive*. The points are medians of posterior distributions and 95% CI. Violin plots show the distribution of strategies across speakers (N = 2541 recordings from 72 speakers).
(**D-E**) Examples of individual vocal strategies by two male speakers prioritizing either loudness (D) or frequency (E). Reassigned spectrograms with 30 ms Gaussian windows and frequency from 0 to 6 kHz on a quasi-logarithmic bark scale. Formants, and thus apparent VTL, could not be measured in the submissive example. Anonymized supplementary audio SA1 to SA6.

Pitch-Loudness Tradeoff

We found that participants raised their f_o not only to sound small, weak, and submissive, but also to sound large, strong, and aggressive. To test our prediction that these latter effects might be a side effect of maximizing loudness to sound formidable, we controlled for loudness when estimating the effect of condition on f_o. Indeed, the accompanying change in loudness completely explained the increase of f_o in the *large* condition and partly also in the *strong*, *aggressive*, *fight/free*, *fight/vowel*, and *loud* conditions (Fig. 2A).

A more rigorous demonstration of a pitch-loudness tradeoff is provided by the *ramp* condition, in which speakers were instructed to simply make a series of vocalizations from the quietest to the loudest possible, without trying to convey any particular meaning and with no constraints on pitch (f_o) or vowel use. Using mixed models that accounted for individual variation by including both a random intercept and a random slope per participant, we found a robust population-level association between voice loudness and f_o in these vocal ramps (Fig. 2C). F_o increased by 1.9 semitones (95% CI [1.5, 2.3]) in women and 1.7 [1.3, 2.2] semitones in men for every doubling of loudness (+6 dB). This change in f_o was much smaller than described by Titze (1994), who reported an increase of 8 to 9 dB per octave in the opposite setup (controlled stepwise increases of f_o rather than loudness), but otherwise in line with theoretical expectations and previous evidence. Interestingly, f_o rose by about an octave (12 semitones) over the typical range of loudness in vocal ramps (30 dB), and both of these values were very close to the changes in the *fight/free* and *fight/vowel* conditions relative to baseline, again suggesting that the rise of f_o in intimidating vocalizations may be a simple side effect of moving from relaxed to very loud phonation.

The apparent VTL slightly decreased in women during vocal ramps from quiet to loud (-1.0 mm [-2.0, -0.2] for a doubling of loudness) and remained unchanged in men (+0.3 mm [-0.7, 1.5]; Fig. 2D). Modeling the change from baseline loudness as a function of the change in f_o and VTL also showed a clear distinction between loud high-pitched and quiet low-pitched vocalizations (red and blue regions in Fig. 2B, respectively), with VTL modulated relatively independently of both loudness and pitch. In short, f_o and loudness changes were strongly coupled, while VTL adjustments provided an extra degree of freedom for vocal size exaggeration.

The pitch-loudness covariance poses a problem for vocalizers wishing to intimidate or assert dominance because elevated f_o potentially counteracts the intimidating effect of a loud voice. Is there some vocal strategy that could mitigate this problem? Testing the prediction that vocal harshness could be used to make high-pitched vocalizations more intimidating, we annotated nonlinear phenomena and found systematic and predictable differences in their prevalence across the vocal tasks (Fig. 2E). Frequency jumps were rare in general, whereas subharmonics were relatively common in most conditions. However, sidebands and chaos were associated with loud vocalizations in the *fight/free* and *fight/vowel* conditions. Chaos, in particular, was highly specific to aggressive vocalizations in the *fight/free* (4.7% [2.4, 8.4] of voiced frames) and *aggressive* conditions (3.4% [1.6, 6.7]). Thus, vocalizers trying to intimidate produced precisely those nonlinearities that are known to sound highly aggressive and to lower the subjectively perceived pitch (Anikin et al., 2021). This may be a way of making loud vocalizations sound both harsher and lower in pitch, in part compensating for the undesirably high f_o that occurs as a byproduct of loudness. In contrast, speakers trying to sound small and weak either maintain a tonal voice or lower their subglottal pressure to the point of descending into vocal fry, whose predominance in baseline (8.8% [6.3, 11.7]) and *weak* (5.6% [3.4, 8.9]) conditions suggests that vocal fry was a side effect of relaxed, quiet phonation.

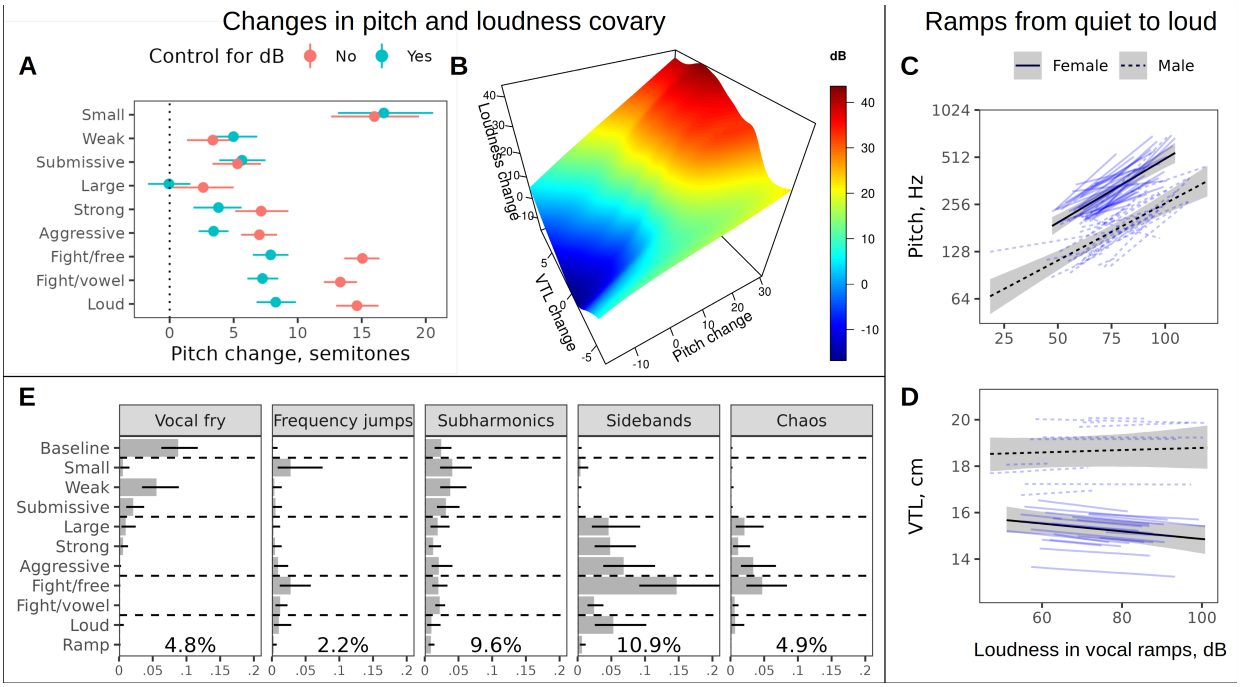

Figure 2 Pitch-loudness tradeoff.
(**A**) Loudness largely accounts for the elevated pitch in intimidating vocalizations. Change in pitch (f_o) vs. baseline with or without controlling for the accompanying increase in loudness (N = 2541 vocalizations).
(**B**) Pitch and loudness are elevated together, whereas changes in pitch and VTL from baseline are not correlated, confirming source-filter independence. Generalized additive mixed model predicting change of loudness from baseline as a smooth function of change in pitch and VTL, with a random intercept per speaker (N = 2013 vocalizations).
(**C-D**) Pitch rises with increasing voice loudness in vocal ramps from quiet to loud (C), while VTL barely changes (D). The black lines show population-level effects, and blue lines show subject-level effects from mixed models. N = 630 vocalizations from 63 speakers in (C) and 240 vocalizations with measurable formants from 31 speakers in (D).
(**E**) Loud and high-pitched intimidating vocalizations contain many harsh-sounding nonlinear phenomena, especially sidebands and chaos. Fitted values from mixed models for the number of frequency jumps (Poisson regression) or, for all the remaining nonlinearities, their cumulative duration as a fraction of voiced vocalizations (zero-one-inflated beta regression). For example, vocal fry is predicted to affect about 3% of the voiced frames in the "baseline" condition; the number "4.8%" corresponds to the overall observed proportion of recordings with some vocal fry in any condition (N = 3279 vocalizations).

Loudness Is a Perceptually Effective Strategy

In the first two sections we discussed vocal production strategies, but communication involves both a signaler and a receiver, and thus it was also crucial to verify the effectiveness of vocal intimidation in affecting listeners' perceptions of speakers. In a series of perception experiments (see Methods), we began by confirming the existence of a perceptual cross-modal correspondence between loudness and size in Experiment 2, using an implicit association test (Parise & Spence, 2012). This method is based on the observation that it is easier to remember which two stimuli are assigned to the same button if they are implicitly perceived as "matching" or congruent. In this case, we tested pairs of sounds that differed only in loudness and pairs of visual stimuli that differed only in size. Two different vocalizations were used as stimuli: a human nonverbal vocalization from Experiment 1 and, to check whether the result generalized to completely different sounds, a synthetic animal call. Accordingly, the visual stimuli were two drawings of human shapes for the human vocalizations and of deer profiles for the animal calls. In either case, listeners made fewer errors (by 2.8%, 95% [1.8, 4.6] and 3.8% [2.2, 7.1], respectively) and responded faster (by 84 ms [52, 124] and 138 ms [95, 193], respectively) when the larger figure was paired with the louder stimulus (Fig. 3A). This indicates that loudness was implicitly associated with large size by listeners, even without instructions to attend to these features or

match them. The extent to which speeded classification tests can isolate purely bottom-up cross-modal correspondences is debated (Getz & Kubovy, 2018), and the role of top-down attentional mechanisms in matching the stimuli remains to be investigated. However, taken together with earlier evidence (Perlman et al., 2022; Smith & Sera, 1992; Teghtsoonian, 1980), this robust perceptual association of loudness with size certainly suggests that a vocalizer can project a large size merely by being louder.

Next, we verified the effectiveness of different vocal strategies by asking listeners in Experiment 3 (see Methods) to judge how large, strong, aggressive, or formidable each speaker sounded. When speakers intended to exaggerate their perceived size, strength, or aggression, listeners indeed judged them as larger, stronger, more aggressive, and more formidable compared to the opposite conditions, in which they appeared to sound small, weak, or submissive (Fig. 3C). Male speakers were more successful than female speakers at exaggerating their size and strength, but both sexes were equally good at exaggerating their aggression. Interestingly, attempting to sound large affected not only the perceived size of speakers (+0.32 on a scale of 0 to 1 when contrasting *large* vs. *small* conditions averaging across male and female speakers, 95% CI [0.28, 0.35]), but also their strength (+0.32 [0.28, 0.36]) and aggression (+0.33 [0.29, 0.36]). In fact, size exaggeration had a greater effect on perceived strength (+0.23 [0.19, 0.27]) and aggression (+0.31 [0.28, 0.34]) than when speakers were specifically instructed to sound strong or aggressive (Fig. 3C).

However, the largest effects in exaggerating strength (+0.46 [0.42, 0.49] compared to the average of *small*/*weak*/*submissive* conditions) and aggression (+0.51 [0.48, 0.55]) were achieved in the *fight/free* conditions, when speakers were presented with a concrete ecologically valid scenario of a physical fight and asked to intimidate an imaginary opponent by vocalizing as loudly as possible (Fig. 3B). Perceptually, there are thus sizable carry-over effects between vocal displays of size, strength, and aggression, which is consistent with earlier observations that traits like perceived size and masculinity largely overlap in human nonverbal vocal communication (Pisanski et al., 2012). Yet, there is also a distinction between the predominantly frequency-mediated size exaggeration and loudness-mediated expression of aggression and strength.

Pooling together vocalizations from all conditions in the same perceptual experiment had the drawback that the natural variability in loudness was too broad and had to be compressed, otherwise some sounds would have been inaudible and others painfully loud (see Methods). Therefore, we performed two more playback experiments using only the relatively loud vocalizations from the *fight/free* and *fight/vowel* conditions, played back at either original or normalized loudness. Because aggression, strength, and formidability ratings were closely correlated (all pairwise Pearson's correlations of ratings per sound > 0.81 at the original loudness, > 0.66 after loudness normalization), we focus here on ratings of formidability and size (see Experiments 4 and 5 in Methods). When the audio was not normalized (Experiment 4), loudness was a major determinant of perceived size (+0.08 [0.06, 0.09] when doubling the loudness from 90 to 96 dB, in both male and female speakers) and formidability (+0.11 [0.10, 0.13] in men and +0.10 [0.09, 0.11] in women, Fig. 3D). In fact, VTL had surprisingly little effect on perceived size (+0.02 [0.01, 0.03] per cm in men and +0.01 [0.00, 0.02] in women) and essentially no effect on formidability (+0.01 [0.00, 0.01] in men and +0.00 [-0.01, 0.00] in women). Contrary to the predictions of the frequency code, f_o was *positively* associated with perceived formidability (+0.23 [0.18, 0.28] per octave in men and +0.15 [0.12, 0.18] in women) and even with perceived size, albeit with more uncertainty (+0.04 [-0.01, 0.09] in men and +0.07 [0.03, 0.11] in women). Once the audio was normalized for loudness (Experiment 5, Fig. 3E), however, the association of f_o with perceived size disappeared in women (-0.01 [-0.05, 0.02]) and became negative in men (-0.05 [-0.09, 0.00]).

Once again, this shows that the loudness-pitch correlation at the production level undermines or even reverses the association of low pitch with large size and formidability, which would otherwise be expected from the frequency code (Morton, 1977; Ohala, 1984). Furthermore, listeners may be able to estimate how loud the vocalization was originally, even if the audio is normalized, as our unpublished

analyses suggest. High-pitched vocalizations may thus be judged to be more aggressive not because of high f_0 *per se*, but because these sounds are interpreted as originally loud – a possibility to explore in future studies.

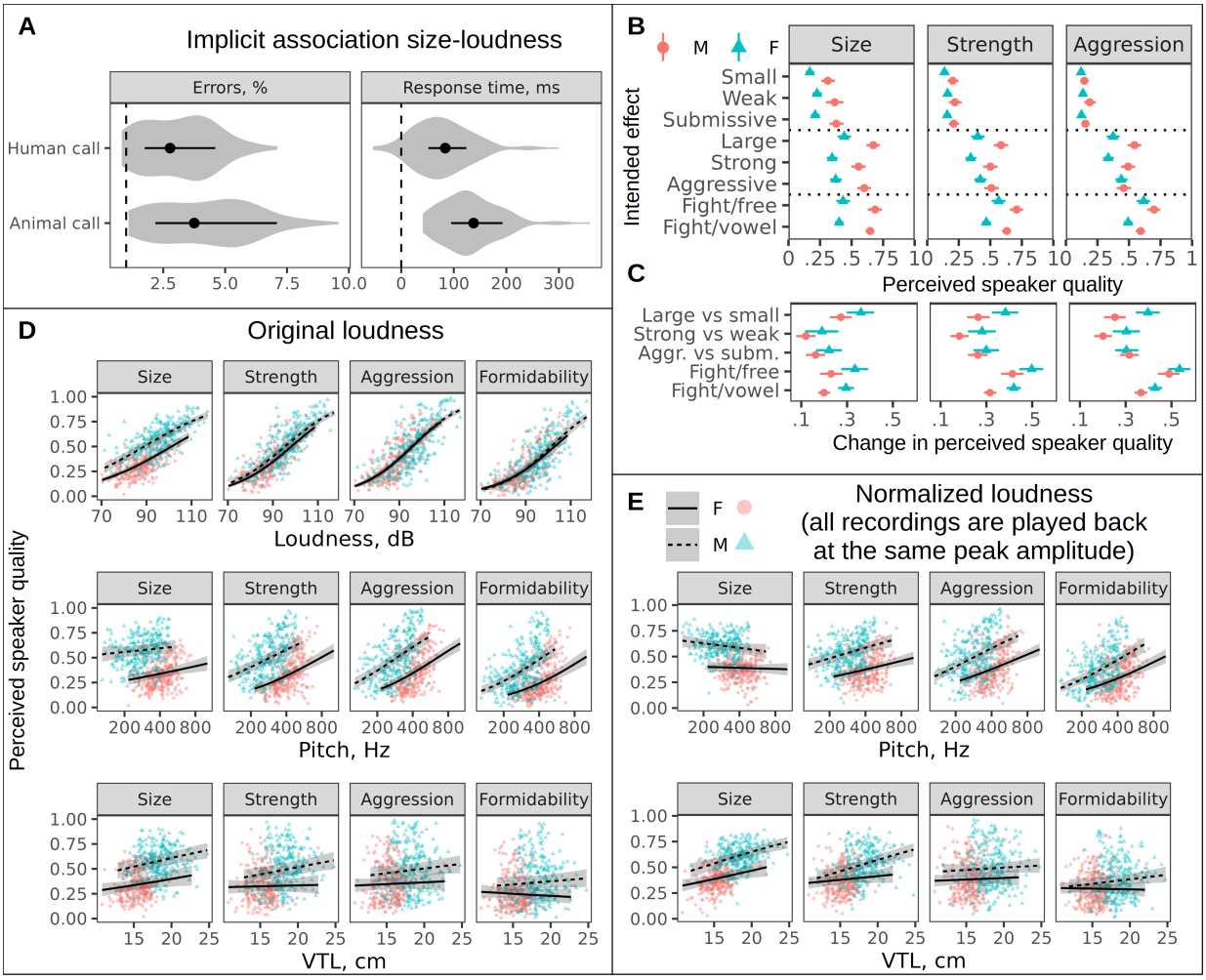

Figure 3 Perceptual effects of loudness, pitch, and VTL.
(**A**) Evidence of a cross-modal correspondence between physical size and loudness in an implicit association test: faster and more accurate responses are obtained when the larger picture is paired with the louder sound. The congruence effect is estimated with logistic and lognormal mixed models for accuracy and response time, respectively, assuming that the congruence effect can vary across participants. Violin plots show the distribution of effect sizes across participants. *N* = 12892 trials in two independent samples of 30 and 21 participants (Experiment 2).
(**B-C**) Perceived speaker characteristics vary as a function of what the speaker is trying to convey. Size, strength, and aggression are successfully exaggerated, but not clearly differentiated: a person trying to portray aggression also sounds larger and stronger, etc. (B) Fitted values per condition and speaker sex from mixed models with random intercepts per vocalization, speaker, and listener. (C) Contrasts between conditions (*fight/free* and *fight/vowel* conditions are compared to an average of *small*, *weak*, and *submissive*). *N* = 12146 ratings of 1055 recordings by 125 listeners (Experiment 3).
(**D-E**) Perceived speaker characteristics vary as a function of acoustic characteristics of recordings played back at the original loudness (D, Experiment 4) or at a standard, normalized loudness (E, Experiment 5). Loudness and, to some extent, VTL predict vocal size and formidability, but the positive association of pitch with size is reversed when the audio is loudness-normalized. Mixed models predicting the ratings on each of four perceptual scales, separately for loudness, pitch, and VTL, with random intercepts per sound, speaker, and listener. *N* = 16172 ratings of 599 sounds by 162 listeners in D (Experiment 4) and 16179 ratings of 599 sounds by 160 listeners in E (Experiment 5).

The Role of Vowels

Thus far we have focused on loudness, pitch, and overall VTL in vocal intimidation, but it may also be possible to convey large size or aggression by producing particular vowels. We began by modeling vowel quality (operationalized as speaker-normalized formants F1 and F2) as a function of experimental condition, focusing only on fully nonverbal vocalizations from all conditions because the choice of vowel is linguistically constrained in verbal utterances. Nonverbal vocalizations in most of the conditions clustered around the middle of the vowel space, surrounded by the six target vowels (i, e, a, o, u, ʉ), indicating that there were no consistent attempts by vocalizers to manipulate vowel quality in order to sound weak or strong, submissive or aggressive, or even large (Fig. 4A). However, the typical vowel quality in the *small* condition was shifted towards /i/ (Fig. 4A). Specifically, F1 was 0.15 [0.04, 0.26] units of formant dispersion lower than its average value in the main cluster of conditions (*large*, *weak*, *strong*, *submissive*, and *aggressive*), and F2 was 0.36 [0.06, 0.64] higher when speakers attempted to sound small.

Further, the two conditions in which participants were asked to be as loud as possible (the *loud* and *fight/free* conditions) displayed a shift of F1 upwards by 0.14 [0.10, 0.17] and 0.16 [0.13, 0.20], respectively, relative to the main cluster (Fig. 4A). In other words, the loudest vocalizations tended toward a very open a-like vowel sound. This is in line with our analysis of vocal ramps, which also showed a slight increase of F1 with loudness, probably due to greater mouth opening (Fig. 4B). Specifically, F1 in vocal ramps increased by 0.04 units of formant dispersion (95% CI [0.03, 0.06]) , and F2 increased about half as much, by 0.02 [0.00, 0.05], for every increase in loudness by one standard deviation (~12 dB). Thus, vowel quality in vocalizations became slightly more open as loudness increased, and there was a marked shift toward /i/ when the task was to appear as small as possible.

In the analysis of vocal ramps, the need to increase loudness drove changes in vowel quality. However, in many cases the vowel may be fixed instead. For example, a person might yell a particular word (e.g., *stop!*), round the lips, or lower the first two formants in order to sound larger (Pisanski et al., 2022), and the question is whether the choice of vowel would affect voice loudness. This is what we tested in the *fight/vowel* condition, in which speakers were asked to intimidate an opponent while being as loud as possible and using a particular target vowel.

We found that /i/ and /ʉ/ were about 6 dB quieter than /a/ and /o/, while /e/ and /u/ were intermediate in terms of loudness (Fig. 4C-D), controlling for a slight trial effect (+0.37 dB [0.12, 0.62] with each of the six vowels, recorded in random order). The analysis of loudness based on sound intensity measured in dB ignores the fact that the human frequency sensitivity curve is not flat. A more appropriate measure for comparing vowels is subjective loudness expressed in sones (Fastl & Zwicker, 2006), which accounts for the fact that open or front vowels, such as /a/ and /i/, are “brighter” in timbre and therefore sound louder at the same intensity than do closed or back vowels such as /u/. The main difference when using sones rather than decibels was that /u/ dropped to the level of /i/ and /ʉ/, while /e/ approached /a/ and /o/ in loudness (Fig. 4C-D). This is quite intuitive as /e/ is a “bright” vowel, whereas /u/ has very little energy in harmonics. Because participants did not always produce the intended vowel correctly, we also analyzed loudness as a function of the actually produced vowel, namely speaker-normalized values of F1 and F2, and controlling for f_o (because high-pitched vocalizations tend to be louder). The results confirmed the analysis based on intended vowels: regions of the vowel space around /a/, /o/, and /e/ were consistently the loudest (Fig. 4D).

Open vowel sounds with a high F1, especially /a/ and to some extent /o/ and /e/, were thus optimal for maximizing loudness. This confirms earlier evidence on intrinsic vowel loudness obtained in singers (Gramming & Sundberg, 1988; Hunter & Titze, 2005; Lamesch et al., 2012) and suggests that vowel quality has important implications in terms of the maximum achievable loudness. We also observed the expected differences in intrinsic vowel pitch: f_o was highest in /i/ and /u/ and lowest in /a/, with a difference between vowels of about 1 to 2 semitones (Fig. 4C-D). This is in line with previous

reports of higher f_0 in open vs. closed vowels in speech, possibly because of mechanical coupling, source-filter interaction with f_0 locking to F1 in high-pitched sounds, or even voluntary control by speakers (Ladd & Silverman, 1984; Ohala & Eukel, 1978). Finally, there were large differences between vowels in terms of the apparent VTL, which was longest for /u/ and shortest for /i/, due to large changes in the relative frequencies of the first few formants. We did not analyze formants above F4, which may be more stable within speaker across vowels, or measure the physical VTL, and listeners may be able to adjust speaker size estimates by taking into account the vowel that they hear (Barreda, 2016; Pisanski et al., 2022). However, the fact remains that vowels are not equal in terms of their intrinsic loudness, pitch, and formant frequencies, and all three of these acoustic characteristics may be involved in sound symbolism. How do these differences between vowels affect how large, strong or aggressive the speaker sounds?

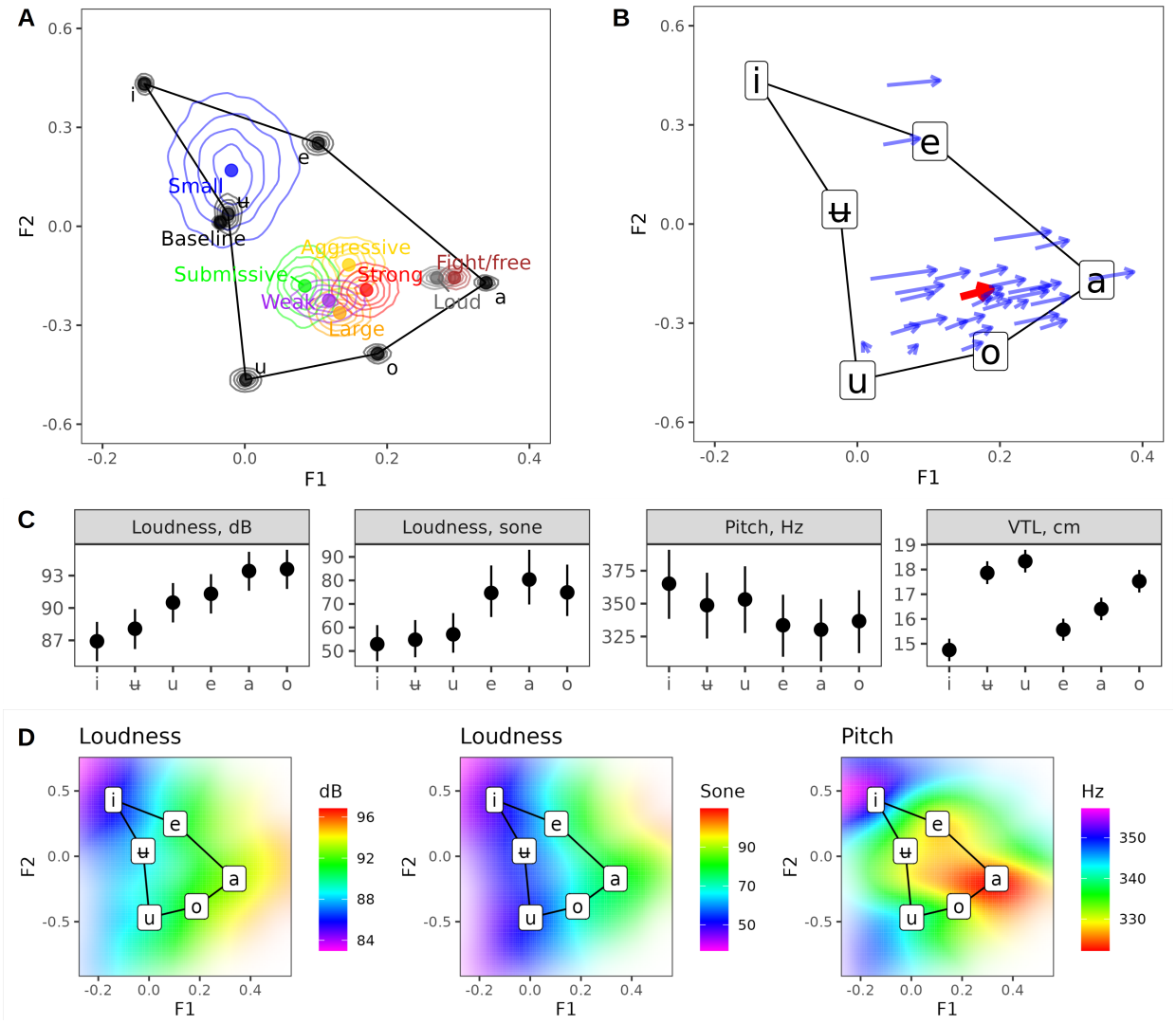

Figure 4 Acoustic differences between vowels.
(**A**) Speakers attempting to sound small are more likely to produce i-like vowels, whereas vowel quality in loud vocalizations approaches /a/. Text labels and points in color: centroids and density plots per condition. The estimates are fitted values from a multivariate mixed model predicting speaker-normalized F1 and F2 as a function of condition, the effect of which can vary per subject, in all nonverbal vocalizations except vocal ramps. *N* = 1990 vocalizations with measurable formants from 72 speakers.
(**B**) The vowel becomes more open, with a higher F1, in vocal ramps from quiet to loud, suggesting that speakers open their mouth wide to be loud. The arrows show predicted change in speaker-normalized F1 and F2 as the loudness varies by

one standard deviation around the average value (fine blue arrows = per individual, thick red arrow= population-level effect). For example, an arrow pointing right indicates that F1 rises (the vowel becomes more open) as the speaker increases voice loudness. *N* = 240 vocalizations by 31 speakers (only purely nonverbal ramps with measurable formants). (**C**) Voice characteristics vary with *intended* vowel in the *fight/vowel* condition. *N* = 1266 vocalizations from 70 speakers in the *fight/vowel* condition.
(**D**) Voice characteristics vary with the *actually produced* vowel. Gaussian additive mixed models predicting voice characteristic (pitch or loudness controlling for pitch) as a smooth function of speaker-normalized F1 and F2 with a random intercept per speaker. Hue shows fitted values, and transparency encodes kernel density of observations. Centroids of six target vowels are plotted to show the approximate borders of the vowel space. *N* = 1140 vocalizations with measurable formants from 69 speakers in the *fight/vowel* condition.

To test this, we asked listeners to rate the vocalizations from the *fight/free* and *fight/vowel* conditions, played back either at the original loudness level (Experiment 4) or normalized to the same peak amplitude (Experiment 5). At the original loudness, /a/ was associated with the greatest perceived size, closely followed by /o/ and /u/ (Fig. 5A-B). The difference between /a/ and /u/ was not statistically robust (0.02 [-0.02, 0.06]), but /i/ sounded considerably “smaller” than both /a/ (by 0.11 [0.07, 0.15]) and /u/ (by 0.09 [0.05, 0.13]). Strength, aggression, and formidability scales produced a similar ranking of vowels, with /i/ sounding the least formidable and /o/ the most. Notably, /u/ was perceived to be considerably less formidable than /a/ (by 0.03 [0.00, 0.07]) or /o/ (0.04 [0.01, 0.07]).

When the vowels were played back at the same standard loudness, the perceptual differences between them became attenuated (Fig. 5C-D). In particular, /a/ was no longer significantly “larger” than /i/ (0.02 [-0.02, 0.07]), and /u/ moved slightly ahead of /a/ in terms of perceived size (0.03 [-0.01, 0.07]). The differences in perceived aggression and formidability essentially disappeared, the greatest difference in perceived formidability being between /u/ and /i/ (0.03, 95% CI [0.00, 0.07]). Interestingly, vocalizations from the *fight/free* conditions with no constraints on the vowel were by far the most intimidating (Fig. 5A,C). They tended to be a few decibels louder than those in the *fight/vowel* condition (Fig. 1A), perhaps because trying to produce a specific vowel made speakers more self-conscious, but the main difference is probably in the vast amount of nonlinear phenomena in the *fight/free* condition (Fig. 2E). In sum, differences in intrinsic loudness have a clear effect on the perceived “size” of different vowels and their suitability for vocal intimidation. In particular, the need to balance frequency (low formants) with loudness might explain the otherwise mysterious appearance of open, a-like vowels alongside /u/ in *mil-mal* sound symbolism (Newman, 1933; Pisanski et al., 2022; Sapir, 1929).

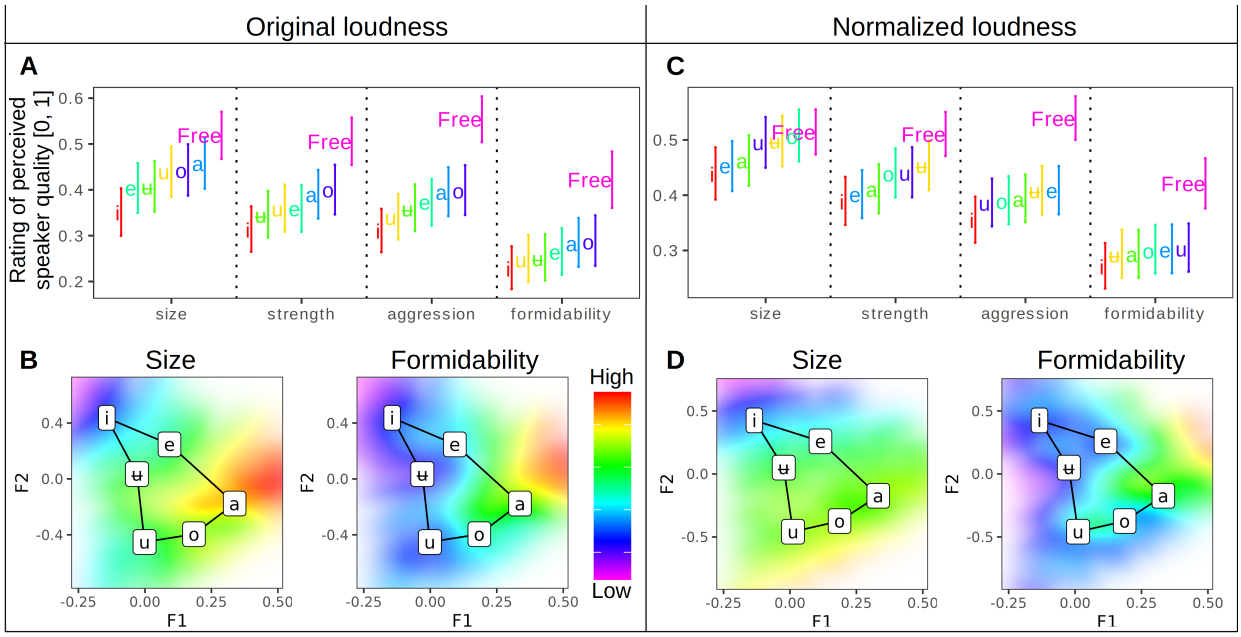

Figure 5 Perceptual effects of vowel quality in vocal intimidation.
(**A-B**) When vowels are played back at the original loudness, /i/ sounds small, while /a/ and /o/ sound large and formidable. *Free* = participants produced any sound while pretending to intimidate an opponent (*fight/free* condition); *all other vowels* = instructed to use the given specific vowel (*fight/vowel* condition). Mixed models predicting the perceptual ratings of vocalizations intended to intimidate an opponent in a fight as a function of vowel, with random intercepts per sound, speaker, and listener. N = 16172 ratings of 599 vocalizations by 162 listeners (Experiment 4).
(**C-D**) When vowels are played back at standard loudness (normalized for peak amplitude), /i/ still sounds small, but other distinctions between vowels become very minor. N = 16179 ratings of 599 vocalizations by 160 listeners (Experiment 5).

Physically Strong Vocalizers Are Simultaneously Low-Pitched and Loud

An important aspect of vocal communication in general, and of vocal formidability in particular, is its honesty – that is, the extent to which a signal is a reliable predictor of the size or strength of the vocalizer (Maynard Smith et al., 2003). We tested for reliable acoustic cues to height, body mass index, and physical strength (operationalized as average hand grip strength in both hands) in baseline speech and in aggressive vocalizations intended to intimidate an imaginary opponent (the *fight/free* condition, see Methods). Simple bivariate relationships between acoustic and physical variables may in this case be misleading because f_o is positively correlated with loudness within each sex, so that loudness masks the negative association of f_o with strength. It is more meaningful to estimate, within each sex, the conditional effects of f_o, VTL, and loudness using multiple regression. In such a model examining relaxed baseline speech (Fig. 6A), only loudness reliably predicted actual physical strength within each sex (+0.30 SD in strength for each extra SD of loudness, 95% CI [0.15, 0.45]), and only VTL predicted height (+0.47 SD [0.19, 0.76]). Looking at loud vocalizations intended to intimidate the opponent (Fig. 6B), loudness was positively associated with actual strength (+0.39 SD [0.19, 0.60]), while f_o was negatively associated with actual strength within each sex (-0.38 SD [-0.64, -0.13]). Likewise, low f_o controlling for loudness was a marginal predictor of actual height in aggressive vocalizations (-0.27 SD [-0.57, 0.03]). No acoustic predictors of body mass index were found, and the models explained only about 15% of its variance (Bayesian R^2), compared to 55-65% of variance in actual strength and height.

Thus, it appears that stronger, and perhaps also taller, people attempting to intimidate are louder (at a given pitch) or, equivalently, have a lower pitch (f_o) for a given loudness compared to people who are weaker and shorter. This finding implies that a combination of low voice pitch plus high loudness provides some information about the vocalizer's actual strength, serving as a hitherto elusive acoustic

index of physical formidability: a "low *and* loud" optimum. In future studies it will be important to verify how reliably a combination of low frequency and high loudness predicts strength and size in different types of vocalizations and species, and whether it can be achieved by weaker individuals – that is, whether it is indeed an honest index of physical formidability.

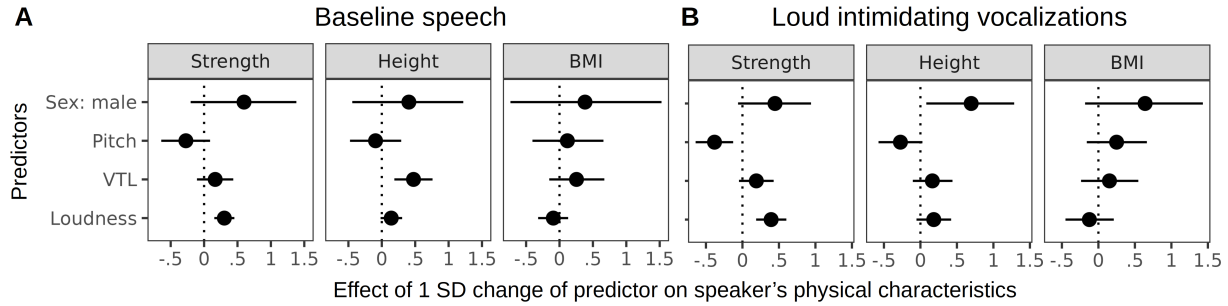

Figure 6 Loudness is an honest index of physical formidability.
Speaker sex and voice characteristics predict physical strength, height, and body mass index (BMI) in (**A**) baseline speech and (**B**) aggressive nonverbal vocalizations in the *fight/free* condition. Loudness is positively associated with strength within each sex after accounting for pitch and VTL. Multiple regression models on mean values per participant, separately for strength, height, and BMI. Both predictors and outcomes are z-transformed (pitch is also log-transformed prior to normalization), so the values correspond to the predicted change in the outcome (in standard deviations) when changing a predictor by 1 SD and keeping all other predictors constant. N = 69 speakers for baseline speech and 63 for aggressive vocalizations (6 speakers were omitted from the analysis of aggressive vocalizations because the formants, and thus VTL, were not measurable).

Discussion

We present theoretical arguments for why loudness could be a crucial, yet relatively neglected, dimension in the vocal communication of both human and non-human animals, particularly in confrontational contexts and dominance displays, where selection is expected to operate on signalers to exaggerate their apparent formidability. This is substantiated by converging evidence from a voice production study and four perceptual experiments, which together demonstrate that vocal loudness is indeed consistently and successfully exploited for vocal intimidation and size exaggeration by humans.

Since loudness covaries with other acoustic dimensions, the prominent role of loudness in vocal intimidation has important implications for our understanding of the relationship between the form and function of vocal signals. The easiest way to vocalize loudly is to increase vocal effort (and thus f_o) and/or to produce open vowels with a high F1. This brings loudness in direct conflict with the well-established principle that low auditory frequencies are employed to exaggerate body size and intimidate an opponent (Morton, 1977; Ohala, 1984). The strong coupling between loudness and pitch may explain why some correlational studies unexpectedly report that elevated pitch conveys dominance (Salais et al., 2022), whereas low pitch correlates with dominance when it is manipulated experimentally, such that loudness is normalized across pitches (e.g., Puts et al., 2007), or when partial effects of pitch and loudness are teased apart with multiple regression (Hodges-Simeon et al., 2010). And while low-pitched voices are typically attributed to larger individuals (Pisanski & Bryant, 2019), the pitch-loudness tradeoff may also help to explain the otherwise mysterious finding that high pitch is successfully used to refer to the larger, not smaller, of two objects in vocal charades (Perlman et al., 2022). Namely, experimentally shifting f_o preserves all other acoustic variables, including voice quality and loudness, but the mechanism of voice production is such that lowering f_o incurs a heavy cost in terms of sacrificing loudness.

Based on how listeners in our studies rated the original or amplitude-normalized intimidating vocalizations, this accompanying drop in loudness outweighs any benefit that speakers would otherwise accrue from lowering their voice pitch in vocal intimidation. As a result, vocalizing loudly at

a relatively high pitch is apparently a better strategy than growling quietly if the objective is to intimidate the opponent in a fight, although low frequency may be more effective for sounding as large as possible. An interesting issue for follow-up studies is whether there is an upper limit beyond which gains in loudness begin to plateau or become offset by increasingly shrill voice quality. Likewise, it will be important to investigate how the pitch-loudness tradeoff plays out in contexts that are not directly confrontational and involve a demonstration of confidence and social dominance rather than physical formidability, such as political debates. Indeed, prior research has shown evidence of a distinction between physical and social dominance, wherein these two related but distinct constructs can also be expressed differentially in human vocal signals (Aung & Puts, 2020; Puts et al., 2006).

In addition to being high-pitched, loud vocalizations are produced with a wide-open mouth, which improves sound radiation, but has an undesirable side effect from the point of view of vocal size exaggeration: it raises formant frequencies, especially F1. Despite this increase in F1, open a-like vowels do make a speaker sound larger and more formidable when the audio recording of their voice is not amplitude-normalized because a-like vowels offer a boost of about 6 dB relative to /i/ and /u/. If the audio is normalized, as in nearly all perceptual experiments in voice research, vowels with the lowest formant frequencies, especially /u/, no longer have to pay the penalty of being quiet and gain some advantage, as also reported in previous studies (Barreda, 2016; Pisanski et al., 2022). Thus, both vowel quality and its intrinsic loudness, or sonority (Behrman, 2021), are relevant in the context of vocal intimidation and size-related sound symbolism. In particular, the most consistent and robust vowel effect, in both production and perception, was the association of /i/ with small size. This is in line with previous research (Blasi et al., 2016; Pitcher et al., 2013; Winter & Perlman, 2021) and not surprising as /i/ has the lowest intrinsic loudness, highest pitch, and smallest apparent vocal tract length, making it particularly good for communicating small size and unsuitable for expressing aggressive intent.

Apart from being intrinsically louder, open vowels with a high F1 may be advantageous for high-pitched loud vocalizations intended to convey aggression because register transitions from chest voice to falsetto, in which only parts of the vocal folds vibrate, tend to occur as f_o approaches F1 (Tokuda et al., 2010). Assuming that falsetto sounds less intimidating at a given f_o, which seems intuitive but needs to be confirmed in future studies, vocalizers should attempt to maintain the ordinary chest register in aggressive yells, while raising f_o so as to maximize loudness (as in the example shown in Fig. 1D). Raising F1 can thus help to prevent switching to falsetto on high notes, but this means that the vowel will tend to be /a/ or /o/. Sonority may also be relevant to other types of sound symbolism apart from vowels. For instance, the association of voiced consonants with large size is usually explained in terms of their lower frequencies compared to unvoiced phonemes (Ekström, 2022), but the greater available dynamic range in voiced phonemes could be another factor.

Given a physiological tradeoff between voice frequencies (particularly f_o and F1) and loudness, various vocal strategies may be optimal depending on the species, their vocal anatomy, physical proximity between the signaler and the receiver, and other social or contextual factors. In fact, many animals possess a whole arsenal of aggressive vocalizations, including quiet low-frequency calls (growls or grunts), loud high-frequency calls (screams), or some compromise between these two extremes (roars, barks). Based on what we observed in humans, loudness trumps frequency in the vocal expression and perception of aggression and strength in fight-like contexts, but frequency is more important for size exaggeration. Interestingly, despite theoretical predictions that the vocal tract could be shortened in loud vocalizations due to increased mouth opening, we show that apparent vocal tract length is the only component of the frequency code that remains largely unaffected by vocal loudness, reinforcing its important and independent role in vocal size exaggeration (Barreda, 2016; Pisanski & Bryant, 2019; Pisanski & Reby, 2021).

An important question is whether the vocal strategies described above are based on pure sensory exploitation, or whether some vocal characteristics are honest indices of actual physical formidability (Aung & Puts, 2020; Feinberg et al., 2018). We found that high vocal loudness and low

pitch predicted people's actual physical strength when they were considered simultaneously, suggesting that a combination of "low and loud" may be difficult to achieve without being genuinely large and strong, and therefore the elusive holy grail of vocal intimidation. A compensatory strategy, common in both nonverbal communication (Anikin et al., 2021; Arnal et al., 2015) and rock singing (Borch et al., 2004), is to produce nonlinear vocal phenomena, which lower the perceived frequency in loud vocalizations with undesirably high f_0 (Anikin et al., 2021). A more radical solution is to evolve specific anatomical adaptations for vocalizing loudly at very low frequencies, as observed in roaring cats (Titze et al., 2010).

The proposed notion of a loudness-frequency tradeoff requires a fundamental rethinking of the evolutionary forces shaping the form of acoustic signals, especially regarding the widely accepted frequency code in vocal intimidation (Morton, 1977; Ohala, 1984). It will be essential to replicate our results in different experimental setups and in non-human species, particularly with respect to the association between vocal loudness and physical strength, and to investigate the ability of listeners to estimate the original loudness of a vocalization when the caller is invisible or when the recording has been normalized. Above all, our findings highlight the urgent need to obtain reliable loudness estimates in field recordings: despite the technical challenges, this information is clearly essential for a full understanding of the principles and functions of vocal communication.

Constraints on Generality

The participants in both voice production (Experiment 1) and voice perception (Experiments 2-5) studies were literate adults fluent in Indo-European languages and residing in Europe. It is important to test the extent to which these results generalize across cultures including marginalized and traditional small-scale societies. In addition, we studied vocal intimidation in ecologically valid but imagined contexts: neither the vocalizers nor the listeners were engaged in an actual confrontation. While shared by an overwhelming majority of voice studies and emotion research in general, this is an important limitation that calls for future replications outside the lab. Despite the challenges of obtaining audio recordings of real-world confrontations (but see Aung et al., 2021; Schild & Zettler, 2021; Šebesta et al., 2019), there is some evidence that the vocalizations produced in such situations are indeed very rough and loud (Anikin & Persson, 2017) and can sometimes be recognizably different from vocalizations produced on demand in an experimental setting (Anikin & Lima, 2018). At the same time, studies have shown that agonistic vocalizations produced on demand in the lab often share key acoustic similarities, and thus form-function mappings, with those produced in real-world contexts (e.g., Kleisner et al., 2021; Raine et al., 2019).

References

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods*, *51*(2), 778–792.

Anikin, A. (2023). The honest sound of physical effort. *PeerJ*, *11*, e14944. https://doi.org/10.7717/peerj.14944

Anikin, A., & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics*, *81*(3), 764–777.

Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Quarterly Journal of Experimental Psychology*, *71*(3), 622–641.

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, *49*(2), 758–771. https://doi.org/10.3758/s13428-016-0736-y

Anikin, A., Pisanski, K., Massenet, M., & Reby, D. (2021). Harsh is large: Nonlinear vocal phenomena lower voice pitch and exaggerate body size. *Proceedings of the Royal Society B*, *288*(1954), 20210872.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, *25*(15), 2051–2056.

August, P. V., & Anderson, J. G. (1987). Mammal sounds and motivation-structural rules: A test of the hypothesis. *Journal of Mammalogy*, *68*(1), 1–9.

Aung, T., Goetz, S., Adams, J., McKenna, C., Hess, C., Roytman, S., Cheng, J. T., Zilioli, S., & Puts, D. (2021). Low fundamental and formant frequencies predict fighting ability among male mixed martial arts fighters. *Scientific Reports*, *11*(1), 1–10.

Aung, T., & Puts, D. (2020). Voice pitch: A window into the communication of social power. *Current Opinion in Psychology*, *33*, 154–161.

Barreda, S. (2016). Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics*, *55*, 1–18.

Behrman, A. (2021). *Speech and voice science. Fourth edition*. Plural publishing.

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

Borch, D. Z., Sundberg, J., Lindestad, P.-Å., & Thalen, M. (2004). Vocal fold vibration and voice source aperiodicity in \ lqdist\rqtones: A study of a timbral ornament in rock singing. *Logopedics Phoniatrics Vocology*, *29*(4), 147–153.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication. Second edition*. Sinauer Associates.

Briefer, E. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20.

Bryant, G. A. (2020). The evolution of human vocal emotion. *Emotion Review*, 1754073920930791.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*, 12739.

Demartsev, V., Gordon, N., Barocas, A., Bar-Ziv, E., Ilany, T., Goll, Y., Ilany, A., & Geffen, E. (2019). The “Law of Brevity” in animal communication: Sex-specific signaling optimization is determined by call amplitude rather than duration. *Evolution Letters*, *3*(6), 623–634.

Ekström, A. G. (2022). What’s next for size-sound symbolism? *Frontiers in Language Sciences*, *1*. https://doi.org/10.3389/flang.2022.1046637

Emanuel, A., & Ravreby, I. (2022). *Sounds hard: Speech features reflect effort level and related affect during strenuous tasks*. https://psyarxiv.com/huq2m/download?format=pdf

Fastl, H., & Zwicker, E. (2006). *Psychoacoustics: Facts and models. Third edition*. Springer.

Feinberg, D. R., Jones, B. C., & Armstrong, M. M. (2018). Sensory exploitation, sexual dimorphism, and human voice pitch. *Trends in Ecology & Evolution*, *33*(12), 901–903.

Fischer, J., Kitchen, D. M., Seyfarth, R. M., & Cheney, D. L. (2004). Baboon loud calls advertise male quality: Acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology*, *56*(2), 140–148.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222.

Fitch, W. T. (2016). Sound and meaning in the world’s languages. *Nature*, *539*(7627), 39–40.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522.

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*(3), 407–418.

Fletcher, N. H. (2004). A simple frequency-scaling rule for animal communication. *The Journal of the Acoustical Society of America*, *115*(5), 2334–2338.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition*, *175*, 101–108.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, *25*(20), 5004–5012.

Gillooly, J. F., & Ophir, A. G. (2010). The energetic basis of acoustic communication. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1686), 1325–1331.

González, J. (2006). Research in acoustics of human speech sounds: Correlates and perception of speaker body size. *Recent Research Developments in Applied Physics*, *9*, 1–15.

Gramming, P., & Sundberg, J. (1988). Spectrum factors relevant to phonetogram measurement. *The Journal of the Acoustical Society of America*, *83*(6), 2352–2360.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, *67*(2), 274–284.

Green, J. A., Whitney, P. G., & Potegal, M. (2011). Screaming, yelling, whining, and crying: Categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. *Emotion*, *11*(5), 1124–1133.

Gustison, M. L., & Townsend, S. W. (2015). A survey of the context and structure of high-and low-amplitude calls in mammals. *Animal Behaviour*, *105*, 281–288.

Herzel, H., Berry, D., Titze, I., & Steinecke, I. (1995). Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(1), 30–34.

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, *21*(4), 406–427.

Hunter, E. J., & Titze, I. R. (2005). Overlap of hearing and voicing ranges in singing. *Journal of Singing: The Official Journal of the National Association of Teachers of Singing*, *61*(4), 387.

Jakobsen, L., Christensen-Dalsgaard, J., Juhl, P. M., & Elemans, C. P. (2021). How Loud Can you go? Physical and Physiological Constraints to Producing High Sound Pressures in Animal Vocalizations. *Frontiers in Ecology and Evolution*, *9*, 325.

Johnstone, T., & Scherer, K. R. (1999). The effects of emotions on voice quality. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 2029–2032.

Kleisner, K., Leongómez, J. D., Pisanski, K., Fiala, V., Cornec, C., Groyecka-Bernard, A., Butovskaya, M., Reby, D., Sorokowski, P., & Akoko, R. M. (2021). Predicting strength from aggressive vocalizations versus speech in African bushland and urban communities. *Philosophical Transactions of the Royal Society B*, *376*(1840), 20200403.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*(1), 1–11.

Ladd, R., & Silverman, K. E. (1984). Vowel intrinsic pitch in connected speech. *Phonetica*, *41*(1), 31–40.

Lagier, A., Legou, T., Galant, C., Amy de La Bretèque, B., Meynadier, Y., & Giovanni, A. (2017). The shouted voice: A pilot study of laryngeal physiology under extreme aerodynamic pressure. *Logopedics Phoniatrics Vocology*, *42*(4), 141–145.

Lamesch, S., Doval, B., & Castellengo, M. (2012). Toward a more informative voice range profile: The role of laryngeal vibratory mechanisms on vowels dynamic range. *Journal of Voice*, *26*(5), 672.e9-672.e18.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*(3), 377–394.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73*(4), 653–676.

Leinonen, L., Linnankoski, I., Laakso, M.-L., & Aulanko, R. (1991). Vocal communication between species: Man and macaque. *Language & Communication*, *11*(4), 241–262.

Maynard Smith, J., Harper, D., & others. (2003). *Animal signals*. Oxford University Press.

Mercer, E., & Lowell, S. Y. (2020). The low mandible maneuver: Preliminary study of its effects on aerodynamic and acoustic measures. *Journal of Voice*, *34*(4), 645.e1−645.e9.

Mittal, V. K., & Yegnanarayana, B. (2013). Effect of glottal dynamics in the production of shouted speech. *The Journal of the Acoustical Society of America*, *133*(5), 3050–3061.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869.

Neuhoff, J. G. (2018). Adaptive biases in visual and auditory looming perception. In T. Hubbard (Ed.), *Spatial biases in perception and cognition* (pp. 180–190). Cambridge University Press.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, *45*(1), 53–75.

Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F_0 of Voice. *Phonetica*, *41*(1), 1–16.

Ohala, J. J., & Eukel, B. W. (1978). Explaining the intrinsic pitch of vowels. In *Report of the Phonology Laboratory* (Vol. 2, pp. 118–125). University of California Berkeley.

Ohms, V. R., Beckers, G. J., Ten Cate, C., & Suthers, R. A. (2012). Vocal tract articulation revisited: The case of the monk parakeet. *Journal of Experimental Biology*, *215*(1), 85–92.

Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge University Press.

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In *Communication* (pp. 299–346). Springer.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3–4), 319–333.

Perlman, M., Paul, J., & Lupyan, G. (2022). Vocal communication of magnitude across language, age, and auditory experience. *Journal of Experimental Psychology: General*, *151*(4), 885.

Pisanski, K., Anikin, A., & Reby, D. (2022). Vocal size exaggeration may have contributed to the origins of vocalic complexity. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200401.

Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. Meizel (Eds.), *Oxford handbook of voice studies* (pp. 269–300). Oxford University Press.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C., & Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour*, *95*, 89–99.

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: Evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, *33*(5), 509–519.

Pisanski, K., & Reby, D. (2021). Efficacy in deceptive vocal exaggeration of human body size. *Nature Communications*, *12*(1), 1–9.

Pitcher, B. J., Mesoudi, A., & McElligott, A. G. (2013). Sex-biased sound symbolism in English-language first names. *PLoS One*, *8*(6), e64825.

Pouw, W., & Fuchs, S. (2022). Origins of vocal-entangled gesture. *Neuroscience & Biobehavioral Reviews*, 104836.

Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296.

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., Welling, L. L., Dawood, K., Cárdenas, R., Burriss, R. P., & others. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829), 20152830.

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344.

Raine, J., Pisanski, K., Bond, R., Simner, J., & Reby, D. (2019). Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS One*, *14*(3), e0213034.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519-530.

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

Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal vibratory mechanisms: The notion of vocal register revisited. *Journal of Voice*, *23*(4), 425–438.

Salais, L., Arias, P., Le Moine, C., Rosi, V., Teytaut, Y., Obin, N., & Roebel, A. (2022). Production Strategies of Vocal Attitudes. *Interspeech 2022*, 4985–4989. https://doi.org/10.21437/Interspeech.2022-10947

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225–239.

Schild, C., & Zettler, I. (2021). Linking voice pitch to fighting success in male amateur mixed martial arts athletes and boxers. *Evolutionary Human Sciences*, *3*, e46.

Šebesta, P., Třebickỳ, V., Fialová, J., & Havlíček, J. (2019). Roar of a champion: Loudness and voice pitch predict perceived fighting ability in MMA fighters. *Frontiers in Psychology*, *10*, 859.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., Krauss, A., & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3509–3518.

Shipp, T. (1987). Vertical laryngeal position: Research findings and application for singers. *Journal of Voice*, *1*(3), 217–219.

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*, 1619–1643.

Slocombe, K. E., & Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (Pan troglodytes schweinfurthii) vary as a function of social role. *Journal of Comparative Psychology*, *119*(1), 67–77.

Smith, L. B., & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*, *24*(1), 99–142.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995.

Stout, B. (1938). The harmonic structure of vowels in singing in relation to pitch and intensity. *The Journal of the Acoustical Society of America*, *10*(2), 137–146.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, *236*(3), 82-91.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion*, *10*(2), 216–229.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236.

Teghtsoonian, M. (1980). Children's scales of length and loudness: A developmental application of cross-modal matching. *Journal of Experimental Child Psychology*, *30*(2), 290–307.

Titze, I. R. (1992). Acoustic interpretation of the voice range profile (phonetogram). *Journal of Speech, Language, and Hearing Research*, *35*(1), 21–34.

Titze, I. R. (2000). *Principles of voice production. Second printing*. National Center for Voice and Speech.

Titze, I. R., Fitch, W. T., Hunter, E. J., Alipour, F., Montequin, D., Armstrong, D. L., McGee, J., & Walsh, E. J. (2010). Vocal power and pressure–flow relationships in excised tiger larynges. *Journal of Experimental Biology*, *213*(22), 3866–3873.

Titze, I. R., & Palaparthi, A. (2018). Radiation efficiency for long-range vocal communication in mammals and birds. *The Journal of the Acoustical Society of America*, *143*(5), 2813–2824.

Tokuda, I. T., Zemke, M., Kob, M., & Herzel, H. (2010). Biomechanical modeling of register transitions and the role of vocal tract resonators. *The Journal of the Acoustical Society of America*, *127*(3), 1528-1536.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*, *107*(6), 3438–3451.

Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., Hsiao, T.-Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception: An Interdisciplinary Journal*, *27*(3), 209–222.

Valentova, J. V., Tureček, P., Varella, M. A. C., Šebesta, P., Mendes, F. D. C., Pereira, K. J., Kubicová, L., Stolařová, P., & Havlíček, J. (2019). Vocal parameters of speech and singing covary and are related to vocal attractiveness, body measures, and sociosexuality: A cross-cultural study. *Frontiers in Psychology*, 2029.

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1), 79.

Zahavi, A. (1982). The pattern of vocal signals and the information they convey. *Behaviour*, *80*(1), 1–8.

Funding

AA was supported by grant 2020-06352 from the Swedish Research Council (Vetenskapsrådet). CC, DV, KP, and DR were supported by the French National Research Agency (ANR) grant ["SCREAM", ANR-21-CE28-0007-01] to DR and KP and by the University of Lyon IDEXLYON project as part of the 'Programme Investissements d'Avenir' (ANR-16-IDEX-0005) to DR.

Author contributions

Conceptualization: AA, DV, KP, GB, DR
Investigation: AA, DV, CC
Formal analysis, writing – original draft: AA
Writing – review and editing: AA, KP, GB, DR

Competing interests

Authors declare that they have no competing interests.

Data and materials availability

Audio recordings (except for background speech with non-anonymizable personal information), datasets, and R scripts for data analysis can be downloaded from https://osf.io/ngwcp/.