Supplementary materials for article “The perceptual effects of manipulating nonlinear phenomena in synthetic nonverbal vocalizations”

Experiment 1

Generation of stimuli

The *temperature* hyper-parameter in *soundgen* was set to 0.001. This made the sounds highly reproducible while enabling a minimum level of stochasticity to enhance authenticity (e.g., by adding extra formants above the user-specified ones). Pitch jumps were created by manually coding nearly instantaneous (~1 ms) changes in f_0, aiming to preserve the overall intonation and median f_0.

In *soundgen* each *f*- and *g*-harmonic is synthesized separately, with one vectorized control parameter for g_0 and another for the width of sidebands. Since f_0 is forced to be an integer multiple of g_0, changes in f_0 and/or g_0 naturally lead to bifurcations. For example, as f_0 rises, g_0 changes from $f_0/2$ to $f_0/3$ and so on. With only two control parameters, it is thus possible to create a complex, natural-sounding vocalization with static or dynamic subharmonics and sidebands. Original human recordings were used to determine reasonable values of subharmonic frequency g_0 (lower than f_0) and the width of sidebands for each sound. Often the prototype sound already had one or more nonlinearity, which was simply reproduced and modified (boosted or removed) in the synthetic version; and if not, other comparable sounds were used to derived reasonable parameters for synthesis. The average value of g0 was ~170 Hz (SD 160, range 75 to 850 Hz).

Chaos was emulated by adding strong jitter (all 28 sounds), sometimes combined with shimmer (5/28 sounds) and some turbulent noise (1/28 sound). Mixed nonlinear effects were created by simultaneously adding chaos, subharmonics, and (for longer sounds) pitch jumps at approximately the same intensity and duration as for isolated effects (“Mixed level 1”) or at increased intensity and/or duration (“Mixed level 2”). These manipulations are illustrated in Figure S1.

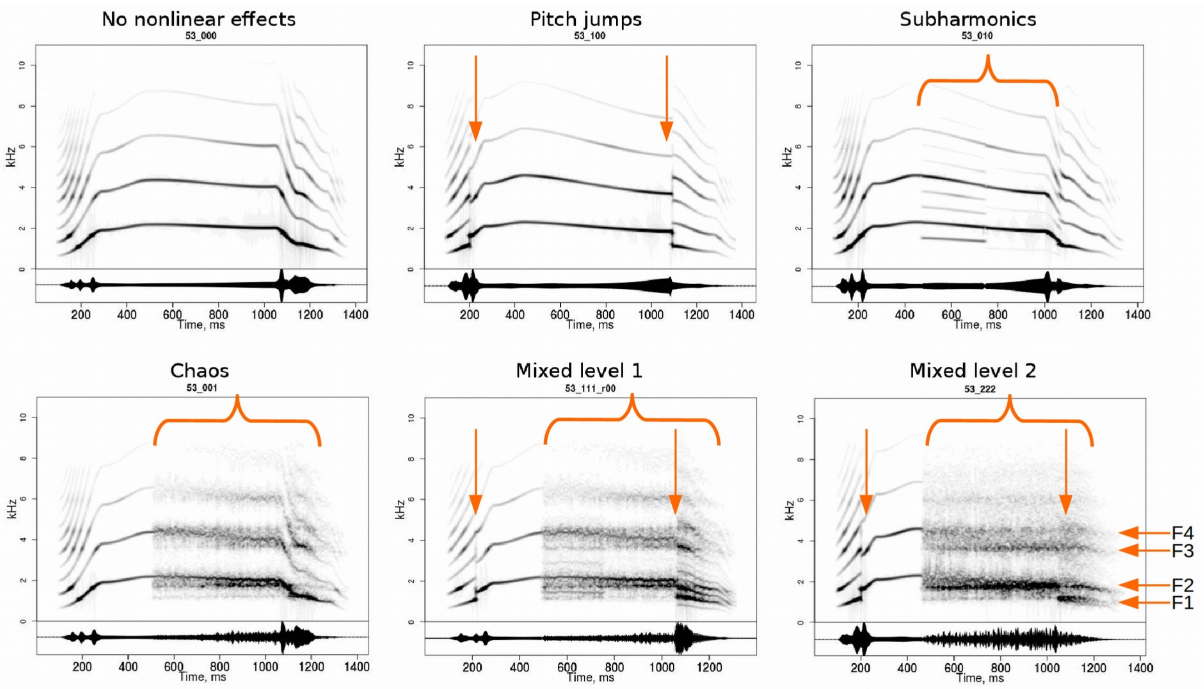


Figure S1. Spectrograms illustrating the addition of nonlinear effects to synthetic scream #53. Pitch jumps are marked with vertical arrows. Note how harmonics become less visible as the strength of chaos increases (Mixed level 2), while formants (marked F1 to F4) become more prominent. AUDIO #3 in Supplements.

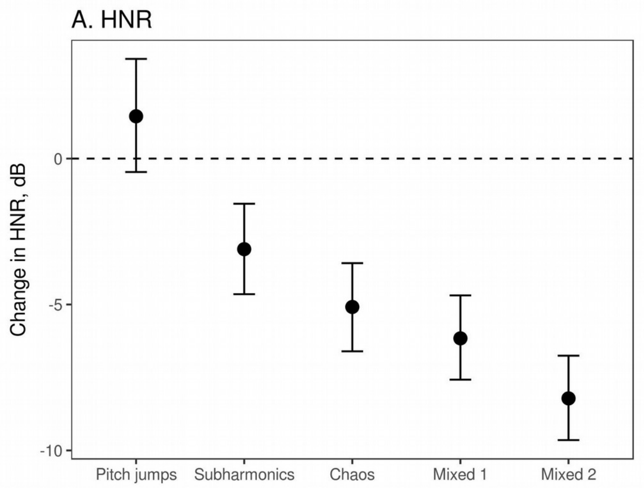


Figure S2. Change in harmonics-to-noise ratio (HNR) after manipulating nonlinear effects compared to unmanipulated sounds. Median of posterior distribution and 95% CI from mixed models with a random intercept per prototype sound. N = 28 prototypes with 144 sounds for (A) and 84 sounds for (B).

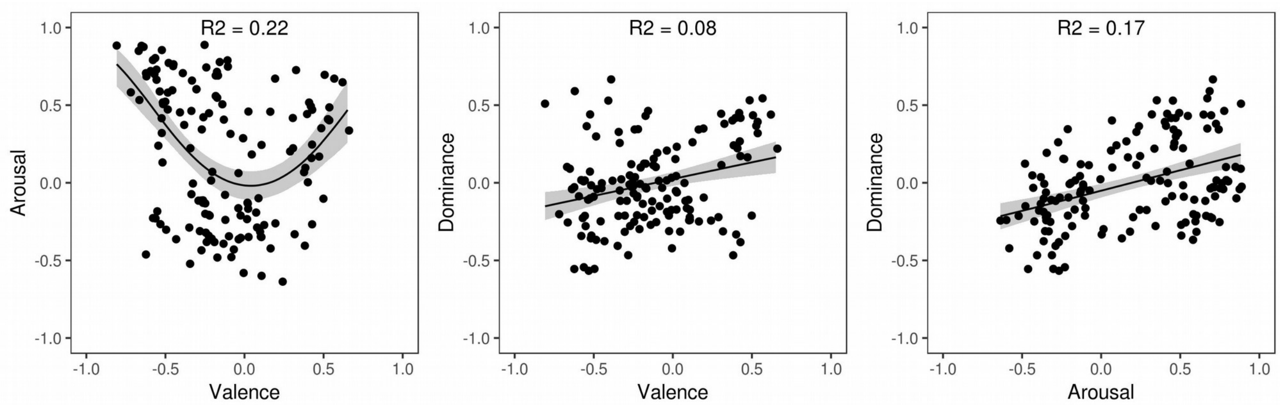


Figure S3. Correlation between valence, arousal, and dominance ratings averaged per stimulus in Experiment 1: scatterplots and the best fit with 95% CI. Each point represents one stimulus.

Table S1. The effect of adding nonlinear phenomena per call type.

Scale	Call type	Predicted difference: median [95% CI]				
		Any vs. none	Pitch jumps vs. none	Subharmonics vs. none	Chaos vs. none	Mixed vs. none
Valence	Cry	**-0.09 [-0.16, -0.03]**	-0.02 [-0.15, 0.06]	-0.02 [-0.13, 0.04]	**-0.15 [-0.27, -0.03]**	**-0.13 [-0.21, -0.05]**
	Gasp	**-0.20 [-0.33, -0.08]**				**-0.20 [-0.33, -0.08]**
	Grunt	**-0.06 [-0.16, 0.06]**		-0.03 [-0.16, 0.11]	-0.02 [-0.18, 0.16]	-0.09 [-0.20, 0.05]
	Laugh	**-0.09 [-0.17, -0.02]**	-0.03 [-0.18, 0.09]	0.00 [-0.10, 0.11]	-0.10 [-0.22, 0.02]	**-0.16 [-0.27, -0.07]**
	Moan	**-0.18 [-0.31, -0.06]**		**-0.28 [-0.46, -0.07]**	0.00 [-0.17, 0.18]	**-0.23 [-0.37, -0.10]**
	Roar	**-0.12 [-0.22, -0.04]**	-0.01 [-0.14, 0.13]	-0.06 [-0.20, 0.05]	**-0.22 [-0.36, -0.09]**	**-0.17 [-0.28, -0.07]**
	Scream	**-0.19 [-0.30, -0.09]**	**-0.25 [-0.39, -0.1]**	-0.01 [-0.13, 0.11]	**-0.20 [-0.33, -0.08]**	**-0.25 [-0.37, -0.13]**
Arousal	Cry	0.00 [-0.03, 0.04]	0.00 [-0.1, 0.04]	0.02 [-0.02, 0.11]	0.01 [-0.02, 0.11]	0.00 [-0.05, 0.02]
	Gasp	0.00 [-0.06, 0.12]				0.00 [-0.06, 0.12]
	Grunt	0.00 [-0.07, 0.06]		0.03 [-0.04, 0.16]	0.01 [-0.09, 0.11]	-0.02 [-0.13, 0.03]
	Laugh	0.00 [-0.07, 0.07]	-0.01 [-0.13, 0.09]	0.07 [-0.02, 0.27]	0.02 [-0.06, 0.13]	-0.03 [-0.16, 0.02]
	Moan	0.02 [-0.04, 0.13]		0.03 [-0.03, 0.21]	0.00 [-0.14, 0.10]	0.02 [-0.04, 0.15]
	Roar	0.00 [-0.05, 0.06]	0.00 [-0.11, 0.11]	0.03 [-0.04, 0.15]	0.02 [-0.05, 0.14]	-0.02 [-0.12, 0.03]
	Scream	0.00 [-0.03, 0.05]	-0.01 [-0.12, 0.03]	0.00 [-0.08, 0.08]	0.01 [-0.03, 0.10]	0.00 [-0.04, 0.08]
Dominance	Cry	0.01 [-0.02, 0.05]	-0.01 [-0.11, 0.03]	0.02 [-0.02, 0.11]	0.00 [-0.05, 0.05]	0.02 [-0.01, 0.08]
	Gasp	0.07 [-0.02, 0.19]				0.07 [-0.02, 0.19]
	Grunt	0.04 [-0.01, 0.13]		0.03 [-0.06, 0.14]	0.01 [-0.06, 0.15]	0.05 [-0.01, 0.18]
	Laugh	0.01 [-0.05, 0.06]	-0.03 [-0.20, 0.05]	0.05 [-0.02, 0.19]	0.00 [-0.08, 0.08]	0.01 [-0.06, 0.08]
	Moan	0.05 [-0.02, 0.14]		0.04 [-0.03, 0.18]	-0.01 [-0.14, 0.06]	0.09 [-0.01, 0.21]
	Roar	0.02 [-0.04, 0.09]	-0.05 [-0.22, 0.03]	0.09 [-0.01, 0.26]	0.01 [-0.06, 0.15]	0.02 [-0.04, 0.09]
	Scream	0.01 [-0.04, 0.06]	-0.01 [-0.10, 0.10]	0.00 [-0.13, 0.09]	-0.01 [-0.11, 0.06]	0.04 [-0.02, 0.12]

Table S2. Sensitivity analysis: the effects of nonlinearities in samples of 86 or 97 participants.

Model	Nonlinear phenomena	Predicted difference: median [95% CI]*		
		Valence	Arousal	Dominance
Reproduced from Table 1 in the main text (86 participants)	Any vs. none	**-0.18 [-0.24, -0.11]**	0.02 [-0.05, 0.08]	**0.05 [0.00, 0.11]**
	Pitch jumps vs. none	**-0.14 [-0.26, -0.03]**	-0.03 [-0.14, 0.08]	-0.02 [-0.11, 0.07]
	Subharmonics vs. none	**-0.11 [-0.19, -0.02]**	**0.08 [0.00, 0.16]**	**0.09 [0.01, 0.16]**
	Chaos vs. none	**-0.18 [-0.27, -0.09]**	0.05 [-0.04, 0.14]	0.03 [-0.05, 0.10]
	Mixed vs. none	**-0.23 [-0.3, -0.15]**	-0.01 [-0.08, 0.07]	**0.08 [0.02, 0.14]**
	Mixed level 2 vs. 1	-0.05 [-0.13, 0.03]	-0.01 [-0.10, 0.07]	0.04 [-0.03, 0.11]
	Chaos vs. subharmonics	-0.07 [-0.16, 0.02]	-0.02 [-0.12, 0.06]	-0.06 [-0.14, 0.02]
Full sample of 97 participants	Any vs. none	**-0.16 [-0.22, -0.10]**	0.01 [-0.05, 0.07]	0.04 [-0.01, 0.09]
	Pitch jumps vs. none	**-0.13 [-0.23, -0.03]**	-0.03 [-0.14, 0.07]	-0.03 [-0.12, 0.06]
	Subharmonics vs. none	**-0.07 [-0.16, 0.01]**	**0.07 [-0.01, 0.15]**	**0.08 [0.01, 0.15]**
	Chaos vs. none	**-0.18 [-0.26, -0.09]**	0.04 [-0.03, 0.12]	0.01 [-0.06, 0.07]
	Mixed vs. none	**-0.21 [-0.28, -0.15]**	-0.01 [-0.08, 0.06]	**0.07 [0.01, 0.13]**
	Mixed level 2 vs. 1	-0.05 [-0.13, 0.02]	-0.03 [-0.1, 0.05]	0.05 [-0.02, 0.11]
	Chaos vs. subharmonics	**-0.1 [-0.19, -0.02]**	-0.02 [-0.1, 0.06]	**-0.07 [-0.14, 0.00]**

Experiment 2

Generation of stimuli

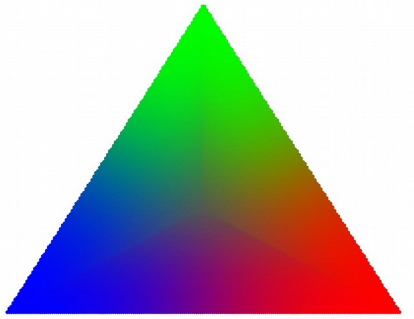
For relatively high-pitched sounds, subharmonics were added in two versions – with higher or lower g_0. Only a single version with subharmonics was created for low-pitched sounds. The duration of the episode with subharmonics and strength of g_0 relative to f_0 varied per sound, but it was always the same for both values of g_0. The average value of g0 was ~150 Hz (SD 110, range 70 to 750 Hz). Chaos was generated with jitter of 1.5 semitones, shimmer of 15%, and the same duration as for subharmonics. Pitch jumps were added only to screams, attempting to preserve the median f_0 and overall intonation contour. In order to standardize nonlinear effects across different versions of the same prototype, *soundgen* was modified to allow the user to specify the exact timing of subharmonics and chaos (version 1.2.0). Minor variation in sound duration dependent on the nature of nonlinear effects was also corrected. These measures insured that experimental stimuli differed only in the manipulated characteristic (nonlinear effects or rolloff) and were otherwise even more strictly identical than in Experiment 1.

Table S3. The algorithm for transforming the marker's coordinates inside an equilateral triangle into relative weights of the vertices.

Taking as input Euclidean coordinates of each vertex and of the marked point, which is assumed (in this case forced) to lie within the triangle, do the following:

1. Calculate the Euclidean distance from the marker to each vertex $D\_vertex_{1\text{-}3}$.
2. Calculate the Euclidean distance from the marker to each edge $D\_edge_{1\text{-}3}$. If the marked point has coordinates (x0, y0) and the two relevant vertices have coordinates (x1, y1) and (x2, y2), the distance from the marker to the edge D_edge_i is given by:
   $D\_edge_i = |(y2 - y1) * x0 - (x2 - x1) * y0 + x2 * y1 - y2 * x1| / sqrt((x1 – x2) \wedge 2 + (y1 – y2) \wedge 2))$,
   where |...| is the absolute value (modulus).
3. Convert distances to edges into weights. For distances $D\_edge_{1\text{-}3}$,
   $W\_edge\_raw_i = (sum(D\_edge_1, D\_edge_2, D\_edge_3) / (D\_edge_i + 0.0001)) \wedge 2$
4. Normalize weights $W\_edge\_raw_{1\text{-}3}$ to range from 0 to 100%:
   $W\_edge_i = W\_edge\_raw_i / (W\_edge_1 + W\_edge_2 + W\_edge_3) * 100$
5. Find the edge closest to the marker and then adjust the distance to the vertex D_vertex_f which lies opposite to the closest edge:
   $D\_vertex\_adj_f = D\_vertex_f \; / sqrt((100 – W\_edge_c) / (100 – 33.3))$,
   where W_edge_c is the weight of the closest edge. Note that $D\_vertex\_adj_f$ approaches infinity as W_edge_c approaches 100, and therefore the weight of the vertex opposite to the closest edge approaches zero as the marker approaches the edge. The weights of edges are not allowed to become exactly zero due to the addition of a small constant in (3). If the marker is exactly equidistant to all three vertices, the weight of each vertex is ~33.3% and this adjustment has no effect.
6. Convert the adjusted distances $D\_edge\_adj_{1\text{-}3}$ into weights as in (3) and normalize to range from 0 to 100% as in (4).

   The resulting distribution of weights can be represented with color gradients:



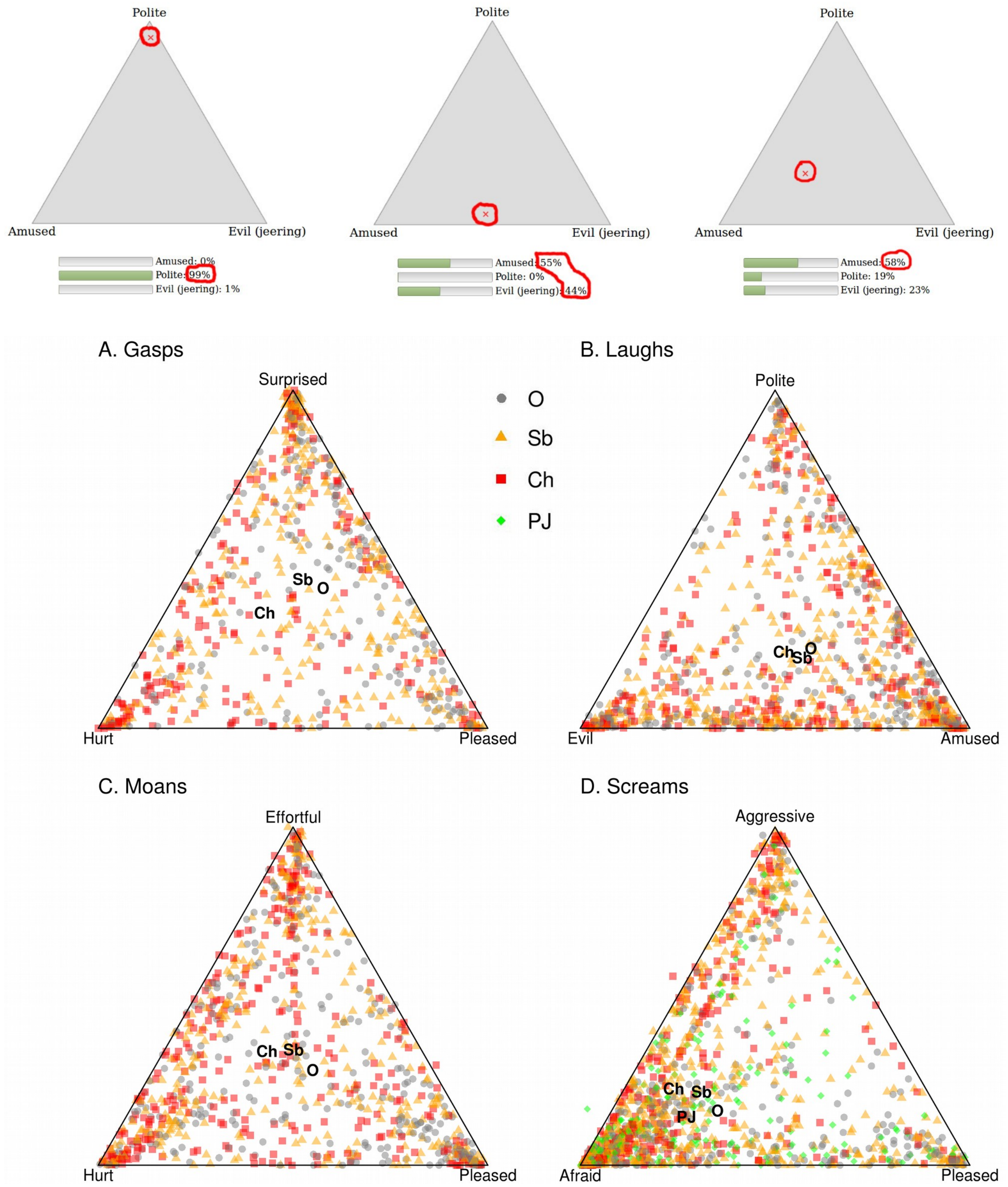


Figure S4. An example of using the triadic scale for rating laughs as it was displayed to participants during training (above) and the observed pattern of individual responses (below). Each point represents one trial (one click), color-coded in the same way as in Figure 4 in the main text. Text labels show the center of gravity for each manipulation. O = original, Sb = subharmonics, Ch = chaos, PJ = pitch jumps.

While the vertices attracted most clicks, there were also many ratings along the edges and closer to the middle of the triangle. Some combinations of emotions were more common than others: for example, for screams there were many responses between *Afraid* and *Aggresive* and between *Afraid* and *Pleased*, but few between *Pleased* and *Aggressive*. This is consistent with response patterns observed with similar sounds when each emotion had its own scale (Anikin & Persson, 2017).

Table S4. Contrasts between the effect of different acoustic manipulations on the weight of emotion categories for each call type, controlling for harmonics-to-noise ratio (HNR): median of posterior distribution (%) and 95% CI. Cf. Table 5 in the main text.

	Gasps				Laughs		
Contrast	**Hurt**	**Surprised**	**Pleased**	**Contrast**	**Evil**	**Polite**	**Amused**
Sb vs. O*	4.1 [-2.2, 10.2]**	2.4 [-6.3, 11.3]	**-6.6 [-12.9, -0.1]**	Sb vs. O	0.2 [-7.4, 8.0]	-1.5 [-6.9, 4.1]	1.2 [-4.2, 6.7]
Ch vs. O	**18.8 [11.5, 26.1]**	-5.6 [-16.0, 4.6]	**-13.2 [-20.4, -5.8]**	Ch vs. O	1.9 [-6.9, 10.7]	0.4 [-5.7, 6.8]	-2.3 [-8.6, 3.9]
Ch vs. Sb	**14.8 [8.9, 20.7]**	**-8.1 [-16.2, 0.0]**	**-6.7 [-12.3, -0.9]**	Ch vs. Sb	1.6 [-5.3, 8.5]	1.9 [-3.1, 6.9]	-3.5 [-8.4, 1.4]

	Moans				Screams		
Contrast	**Hurt**	**Effortful**	**Pleased**	**Contrast**	**Afraid**	**Aggressive**	**Pleased**
Sb vs. O	-0.1 [-6.1, 5.8]	**7.4 [1.3, 13.5]**	-7.3 [-15.8, 1.3]	Sb vs. O	4.3 [-0.9, 9.4]	4.5 [-0.6, 9.5]	**-8.8 [-15.9, -1.8]**
Ch vs. O	5.1 [-3.1, 13.4]	6.2 [-2.0, 14.5]	-11.4 [-23.0, 0.4]	Ch vs. O	**14.5 [5.6, 23.3]**	3.3 [-5.3, 11.8]	**-17.8 [-29.6, -5.9]**
Ch vs. Sb	5.2 [-0.7, 11.2]	-1.2 [-7.2, 4.9]	-4.1 [-12.6, 4.6]	Ch vs. Sb	**10.2 [4.5, 15.7]**	-1.1 [-6.7, 4.4]	**-9.0 [-16.5, -1.4]**
				PJ vs. O	**8.2 [3.5, 13]**	0.9 [-3.7, 5.6]	**-9.1 [-15.7, -2.5]**

* O = original, R- = less energy in harmonics, R+ = more energy in harmonics, Sb = subharmonics, Ch = chaos, PJ = pitch jumps
** Where the models with and without controlling for HNR and spectral centroid disagree, in the sense that the magnitude of effect or the location of the 95% CI lead to substantively different interpretations, the cell is highlighted. For example, the model with HNR and spectral centroid is highly uncertain about the effect of chaos on laughs, whereas the model without these variables predicts that chaos shifts the interpretation from "amused" to "evil".

Table S5. Sensitivity analysis: same as Table 4 in the main text, but with the full sample of 102 participants instead of 83 participants.

	Gasps				Laughs		
Contrast	**Hurt**	**Surprised**	**Pleased**	**Contrast**	**Evil**	**Polite**	**Amused**
Sb vs. O*	**6.1 [1.1, 10.9]**	-0.9 [-8.2, 6.4]	**-5.2 [-10.1, -0.2]**	Sb vs. O	5.3 [-1.1, 11.5]	-2.0 [-6.5, 2.6]	-3.2 [-7.5, 1.1]
Ch vs. O	**20.0 [14.4, 25.5]**	-7.6 [-15.5, 0.4]	**-12.4 [-17.8, -6.8]**	Ch vs. O	**9.8 [2.9, 16.4]**	-2.9 [-7.5, 2.1]	**-6.9 [-11.6, -2.0]**
Ch vs. Sb	**13.9 [8.4, 19.0]**	-6.8 [-14.0, 0.4]	**-7.1 [-12.2, -2.2]**	Ch vs. Sb	4.5 [-1.6, 10.5]	-0.8 [-5.1, 3.4]	-3.6 [-7.9, 0.7]

	Moans				Screams		
Contrast	**Hurt**	**Effortful**	**Pleased**	**Contrast**	**Afraid**	**Aggressive**	**Pleased**
Sb vs. O	4.0 [-0.6, 8.6]	**4.8 [0.5, 9.4]**	**-8.9 [-15.1, -2.6]**	Sb vs. O	1.7 [-1.9, 5.2]	**5.8 [2.3, 9.2]**	**-7.4 [-12.2, -2.5]**
Ch vs. O	**11.8 [6.8, 16.7]**	3.6 [-1.3, 8.4]	**-15.3 [-22.3, -8.5]**	Ch vs. O	**10.0 [6.1, 13.9]**	**6.3 [2.3, 10.1]**	**-16.3 [-21.8, -10.7]**
Ch vs. Sb	**7.8 [2.8, 12.6]**	-1.3 [-6.2, 3.5]	-6.4 [-13.3, 0.4]	Ch vs. Sb	**8.4 [5.0, 11.7]**	0.6 [-2.9, 4.0]	**-8.9 [-13.6, -4.1]**
				PJ vs. O	**7.2 [2.9, 11.5]**	1.3 [-3.1, 5.6]	**-8.5 [-14.7, -2.4]**

* O = original, R- = less energy in harmonics, R+ = more energy in harmonics, Sb = subharmonics, Ch = chaos, PJ = pitch jumps
** Where the models with and without controlling for HNR and spectral centroid disagree, in the sense that the magnitude of effect or the location of the 95% CI lead to substantively different interpretations, the cell is highlighted.