Vocalizations were selected from the recordings of thirty-three speakers pronouncing “Ah” neutrally as well as with anger, disgust, fear, happiness, sadness, and surprise. Recordings were made by the present investigators as well as Belin and colleagues (2008) in a soundproof chamber and digitized at 16 bits/44.1kHz.

Thirty participants (12 male, 18 female, mean age = 22.07), not contributing to the main experiment, classified each vocalization as angry, disgusted, fearful, happy, sad, surprised, neutral or “other” in case none of the aforementioned options seemed adequate. For sounds not categorized as neutral, the participants were prompted to rate emotion intensity and arousal on two 5-point scales ranging from 1 (“very weak”) to 5 (“very strong”).

For this project, we selected the 27 best recognized surprise expressions (mean accuracy = 73.70%, SD = 20.5; mean intensity = 3.3, SD = 0.5; mean arousal = 3.3, SD = 0.4) as well as 27 matching neutral sounds (mean accuracy = 76.91%, SD = 16.8). We decided to use one rather than multiple emotions because we wanted to control stimulus variation/homogeneity/probability between the neutral and the emotion condition. Moreover, we selected surprise because of its good recognition and high arousal value. All selected sounds were normalized at the same root-mean-square value and subjected to spectral rotation (http://www.phon.ucl.ac.uk/resource/software.php) resulting in an analogous set of non-vocal sounds (Warren et al., 2006).

To explore the effects of spectral rotation and emotion, all sounds were subjected to a series of acoustic analyses. Peak sound envelopes were extracted using a 10 ms running average and correlated between vocal and nonvocal as well as between surprised and neutral stimulus exemplars. The average Pearson coefficients were 0.81 and 0.33 for the voiceness and emotion factors, respectively. We, furthermore, computed the power spectral density for each sound using the Welch method with Hanning windowing of 256 samples. Again, resulting vectors for vocal/nonvocal and surprised.neutral sound pairs were subjected to Pearson correlations. The
average correlation coefficients were 0.51 and 0.70 for the voiceness and emotion factors, respectively. Note that differences as a function of voiceness emerged primarily in the lower end of the frequency spectrum as many sounds lost their fundamental frequency following spectral rotation. The harmonics-to-noise ratio (HNR) and specific formants (F1, F2) were analyzed with separate ANOVAs with Voiceness as a repeated measures factor and Emotion as a between items factor. HNR analysis revealed significant effects of Voiceness (F[1,52]=205.1, p<.0001) and Emotion (F[1,52]=8.92, p<.01) as well as an interaction (F[1,52]=11.13, p<.01) indicating that HNR was greater for vocal than for nonvocal and for neutral than for surprised sounds. Moreover, the Voiceness effect was larger for neutral than for surprised sounds. F1 analysis revealed an effect of Voiceness only (F[1,52]=21.48, p<.0001) showing that F1 was higher for vocal than for nonvocal sounds. F2 analysis was non-significant (p > .1).

A second stimulus rating served to explore whether and how spectral rotation affected the perceived human quality of the sounds and their emotionality. To this end, 12 participants (6 female, 6 male, mean age = 25.9) not involved in the main experiment listened to all experimental sounds in random order. For each sound, participants made three judgments. First, they indicated whether the sound was human, somewhat human, or non-human. Second, they indicated whether the sound was characterized by surprise or neutrality. Last, they rated the arousal level of the sound on a 5-point scale ranging from 0 for not aroused to 4 for very aroused. The data are illustrated in Figure 1.
Figure 1. Stimulus rating results from a separate group of participants. From left to right illustrated are the extent to which sounds were perceived as human, the sensitivity with which listeners discriminated between surprised and neutral exemplars, and the arousal emanating from the sound.

We subjected the humanness rating (dummy-coded responses as 3, 2, 1) to an ANOVA with Voiceness (vocal, non-vocal) and Emotion (neutral, surprised) as repeated measures factors and Sex as a between subjects factor. Analysis of humanness revealed effects of Voiceness ($F[1,10]=23.58, p<.0001$), Emotion ($F[1,10]=19.9, p<.0001$), and an Emotion by Voiceness interaction ($F[1,10]=7.6, p<.05$). On average, vocal sounds were perceived as human, whereas non-vocal sounds were perceived as non-human (Figure 1). However, non-vocal sounds were perceived as less non-human in the surprise relative to the neutral condition ($F[1,10]=22.8, p<.001$). Emotion did not influence humanness for vocal sounds ($p>.1$).

Analysis of emotion recognition accuracy was done by computing the d-prime sensitivity score. To this end, surprised sounds classified as surprised were treated as hits, whereas neutral sounds classified as surprised were treated as false alarms. The normalized probabilities of false alarms was subtracted from the normalized probabilities of hits and the resulting scores submitted to an ANOVA with Voiceness and Sex as factors. A Voiceness main effect indicated that participants were significantly better at discriminating surprised from neutral sounds when the sounds were
played in their vocal than in their non-vocal versions (F[1,10]=34.8, p<.0001). Notably, although emotion discrimination for non-vocal sounds was poor, it was significantly better than chance (i.e., 0) as demonstrated by a Welch's t-test (t(11)=3.1, p<.001).

Arousal scores were subjected to an ANOVA with Voiceness and Emotion as repeated factors and with Sex as a between subjects factor. This revealed again effects of Voiceness (F[1,11]=5.2, p<.05), Emotion (F[1,11]=337, p<.0001), and a Voiceness by Emotion interaction (F[1,11]=201.3, p<.0001). Follow-up analysis indicated that neutral sounds were rated as less arousing than surprised sounds and that this effect was greater in the vocal (F[1,11]=537.9, p<.0001) than the non-vocal condition (F[1,11]=29.2, p<.001).

Together these rating results show that spectral rotation significantly impairs the human quality of sounds but preserves some emotional sound characteristics. Most likely acoustic aspects linked to temporal course and intensity are relevant here (Schirmer, Ng, Escoffier, & Penney, 2016).

References

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*, 531–539.

Schirmer, A., Ng, T., Escoffier, N., & Penney, T. B. (2016). Emotional voices distort time: behavioral and neural correlates. *Timing & Time Perception, 4*, 79–98.

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., … Scott, S. K. (2006). Positive emotions preferentially engage an auditory–motor “mirror” system. *The Journal of Neuroscience, 26*, 13067–13075.
ERP Figure of Nose-Referenced Data

ERP Topographical Maps for Different Recording References
Mastoid Analysis of Nose-Referenced Data

Visual inspection suggested that some of the fronto-central Voiceness effects reversed polarity over the mastoids pointing to a possible generator in auditory cortex. To statistically probe this impression, we subjected average voltages for each of the time windows mentioned above to separate ANOVAs with Voiceness, Emotion, and Channel (left mastoid, right mastoid) as repeated measures factors and Sex as a between subjects factor.

For the N1, we obtained an interaction of Voiceness and Channel (F[1,30]=6.7, p<.05) indicating that vocal sounds elicited a larger negativity than non-vocal sounds over the right mastoid. The Voiceness effect was non-significant over the left mastoid. As this effect was of similar polarity as the right-lateralized scalp N1 effect, a generator in auditory cortex is unlikely.

For the P2/P3 complex and the N4, there was a Voiceness main effect (F[1,30]=20.5, p<.0001, F[1,30]=17.5, p<.001) indicating that the ERP to vocal stimuli was more negative than that to non-vocal stimuli over both mastoids. Thus, in both time windows, scalp effects reversed polarity in line with a possible auditory cortex generator.

For the LPP, an interaction of Voiceness and Sex (F[1,30]=4, p=.05) pointed to a significant Voiceness effect in women (F[1,30]=16.3, p<.01) but not in men (p>.1). In women only, the frontal LPP Voiceness effect was reversed in that vocal stimuli elicited a greater negativity than non-vocal stimuli.
ERP Analysis of Average-Referenced Data

Analyses were conducted for the same ERP components and time-windows as described in the main document.

N1. Analysis of N1 amplitudes revealed a significant interaction of Voiceness and Hemisphere (F[1,30]=5.19, p<.05) as well as Emotion and Region (F[2,60]=5.79, p<.05). Over right (F[1,30]=8.18, p<.01) but not left (p>.1) electrode sites, the N1 was larger for vocal as compared to nonvocal sounds. The N1 was marginally or significantly larger for neutral as compared to surprised sounds over anterior (F[1,30]=2.95, p=.09) and central (F[1,30]=5.88, p<.05) regions. This Emotion effect reversed polarity over posterior regions (F[1,30]=7.01, p<.05).

P2/P3 complex. Analysis of the P2/P3 complex revealed main effects of Emotion (F[1,30]=4.48, p<.05) and Voiceness (F[1,30]=61.31, p<.0001) as well as several interactions including Emotion and Region (F[2,60]=7.09, p<.01), Voiceness and Sex (F[1,30]=6.05, p<.05), Voiceness and Region (F[2,60]=54.67, p<.0001), and Voiceness, Region, and Sex (F[2,60]=3.16, p<.05).

The Emotion by Region interaction was explored for each level of Region. Over anterior (F[1,30]=5.91, p<.05) and central regions (F[1,30]=14.89, p<.001), surprised sounds elicited a greater positivity than did neutral sounds. The effect reversed polarity over posterior regions (F[1,30]=5.51, p<.05).

The interaction of Voiceness, Region, and Sex was pursued for female and male participants. In women (F[2,60]=46.48, p<.0001), vocal sounds elicited a greater positivity than nonvocal sounds over anterior (F[1,30]=71.43, p<.0001) and central regions (F[1,30]=55.43, p<.0001). The effect reversed polarity over posterior regions (F[1,30]=19.68, p<.001). A similar, although smaller, pattern was observed in men (interaction, F[2,60]=14.52, p<.0001; anterior, F[1,30]=14.85, p<.01; central, F[1,30]=34.14, p<.0001; posterior, F[1,30]=8.01, p<.05).
N4. Analysis of N4 mean voltages revealed a main effect of Emotion (F[1,30]=7.47, p<.05) and interactions of Emotion with Region (F[2,60]=3.13, p=.051) and with Hemisphere and Sex (F[1,30]=5.3, p<.05). Analysis of the former interaction revealed that the N4 was greater for neutral than for surprised sounds over central regions (F[1,30]=15.07, p<.001). Effects were non-significant over anterior and posterior regions (ps>.1). Analysis of the latter interaction showed that the Emotion effect differed by Hemisphere in men (F[1,15]=5.46, p<.05) but not in women (p>.1). In men, the Emotion effect was significant over right (F[1,15]=5.59, p<.05) but not left electrodes (p>.1).

There was a main effect of Voiceness (F[1,30]=43.8, p<.0001) and interactions of Voiceness with Region (F[1,30]=15.07, p<.001) as well as with Hemisphere, Region, and Sex (F[2,60]=15.07, p<.001). The latter effect was explored for men and women separately. In men, a significant Voiceness, Hemisphere, Region effect (F[2,30]=6.86, p<.01) indicated that Voiceness and Hemisphere interacted over posterior regions (F[1,15]=6.61, p<.05) in that vocal sounds elicited a larger N4 than nonvocal sounds over the left (F[1,15]=5.01, p<.05) but not the right hemisphere (p>.1). The Voiceness by Hemisphere interaction was non-significant over anterior and central regions. There, a Voiceness main effect with greater amplitudes for nonvocal relative to vocal sounds showed irrespective of laterality (anterior, F[1,15]=8.45, p<.05; central, F[1,15]=9.21, p<.01). As for men, there was an interaction of Voiceness, Hemisphere, and Region in women (F[2,30]=3.29, p=.051). However, the Voiceness by Hemisphere interaction was only marginal at central recording sites (F[1,15]=3.33, p=.088). At anterior (F[1,15]=33.44, p<.0001) and central (F[1,15]=15.03, p<.01) recording sites, nonvocal sounds elicited a larger N4 than vocal sounds and the effect reversed polarity at posterior recording sites (F[1,15]=7.56, p<.05).
LPP. The LPP was characterized by main effects of Emotion (F[1,30]=7.02, p<.05) and Voiceness (F[1,30]=51.85, p<.0001) as well as a number of interactions. The factor Emotion appeared in interactions with Voiceness (F[1,30]=4.98, p<.05), Voiceness and Hemisphere (F[1,30]=4.99, p<.05), as well as Hemisphere and Sex (F[1,30]=7.94, p<.01).

Pursuing the Emotion by Voiceness interaction for each level of Hemisphere revealed significant results for left (F[1,30]=8.31, p<.01) but not right electrodes (p>.1). At left electrodes, the LPP was larger for emotional as compared to neutral vocal sounds (F[1,30]=13.74, p<.001). The Emotion effect was non-significant for nonvocal sounds (p>.1). At right electrodes, main effects of Emotion (F[1,30]=5.13, p<.05) and Voiceness (F[1,30]=16.49, p<.0001) indicated that the LPP was larger from emotional as compared to neutral and vocal as compared to nonvocal sounds. Pursuing the Emotion by Hemisphere interaction for each level of Sex revealed significant results for men (F[1,30]=10.18, p<.01) but not women (p>.1). In men, LPP amplitudes were larger for surprised as compared to neutral sounds over right electrodes only (F[1,30]=7.94, p<.05). In women, an Emotion effect showed independently of Hemisphere (F[1,30]=4.3, p=.05).

The factor Voiceness interacted with Sex (F[1,30]=5.62, p<.05), Region (F[2,60]=32.29, p<.05), Sex and Region (F[2,60]=8.29, p<.001), as well as Sex, Region, and Hemisphere (F[2,60]=5.49, p<.01). Analysis of women revealed a significant interaction of Voiceness, Hemisphere, and Region (F[2,30]=3.92, p<.05). Over anterior (F[1,15]=37.87, p<.0001) and central (F[1,15]=47.53, p<.0001) regions, Voiceness main effects indicated that vocal sounds elicited a larger LPP than nonvocal sounds irrespective of laterality. Over posterior regions, the Voiceness effect reversed polarity and interacted with Hemisphere (F[1,15]=3.79, p<.05) indicating that it was greater over the right (F[1,15]=23.32, p<.001) as compared to the left (F[1,15]=14.1, p<.01) hemisphere. In men, the interaction of Voiceness, Hemisphere, and Region was non-significant (p>.1). However, the Voiceness effect differed as a function of Region (F[2,30]=29.94, p<.01) in that it was significant anteriorly (F[1,15]=23.32, p<.001) and centrally (F[1,15]=23.32, p<.001) but not posteriorly (p>.1).
ERP Analysis of Mastoid-Referenced Data

Analyses were conducted for the same ERP components and time-windows as described in the main document.

N1. Analysis of N1 amplitudes revealed a Voiceness main effect (F[1,30]=11.11, p<.01) as well as interactions of Voiceness with Region (F[2,60]=4.13, p<.05) and Hemisphere (F[1,30]=5.19, p<.05). N1 amplitudes were larger for vocal than for nonvocal sounds over central (F[1,30]=18.78, p<.001) and posterior (F[1,30]=15.96, p<.001) but not anterior regions (p>.1). Moreover, the Voiceness effect was greater over the right (F[1,30]=13.46, p<.001) than the left hemisphere (F[1,30]=7.41, p<.05). An interaction of Emotion and Region (F[2,60]=5.79, p<.01) indicated that N1 amplitudes were larger for neutral than for surprised sounds over anterior (F[1,30]=3.96, p=.05) and central (F[1,30]=4.34, p<.05) but not posterior regions (p>.1).

P2/P3 Complex. Amplitudes of the P2/P3 complex were modulated by Voiceness (F[1,30]=94.81, p<.0001) and interactions of Voiceness with Sex (F[1,30]=5.44, p<.05), Region (F[2,60]=54.67, p<.0001), as well as with Sex and Region (F[2,60]=3.16, p<.05). Analysis of the three-way interaction revealed a significant Voiceness by Region effect in both men (F[2,30]=14.52, p<.0001) and women (F[2,30]=46.48, p<.0001). In both sexes, vocal sounds elicited a larger amplitude than nonvocal sounds over anterior (men, F[1,15]=33.13, p<.0001; women, F[1,15]=102.1, p<.0001) and central regions (men, F[1,15]=35.69, p<.0001; women, F[1,15]=65.28, p<.0001). The effect reversed polarity over posterior regions (men, F[1,15]=8.82, p<.01; women, F[1,15]=13.84, p<.05). Voiceness effects were more prominent in women than in men over anterior and central regions.

In addition to Voiceness, Emotion modulated the P2/P3 complex. An Emotion main effect (F[1,30]=94.81, p<.0001) was accompanied by an interaction of Emotion with Region (F[2,60]=94.81, p<.0001) indicating that surprised sounds elicited a greater amplitude than neutral
sounds over anterior (F[1,30]=10.88, p<.01) and central (F[1,30]=12.47, p<.01) but not posterior regions (p>.1).

N4. The N4 was characterized by a Voiceness main effect (F[1,30]=44.33, p<.0001) and interactions of Voiceness with Region (F[2,60]=20, p<.0001) and with Hemisphere, Region, and Sex (F[1,30]=7.27, p<.01). Both men and women showed an interaction of Voiceness, Hemisphere and Region (men, F[2,60]=6.86, p<.01; women, F[2,60]=3.29, p=.05). Further analysis of male participants revealed that the Voiceness by Hemisphere interaction was non-significant at anterior, central and posterior sites (ps>.1). Instead, there was a Voiceness main effect that showed anteriorly (F[1,30]=10.84, p<.01) and centrally (F[1,30]=9.73, p<.01), but not posteriorly (p>.1). In women, effects were similar but larger over anterior (F[1,30]=62.19, p<.01), central (F[1,30]=33.24, p<.01) and posterior regions (F[1,30]=3.61, p=.08).

An Emotion main effect (F[1,30]=195.12, p<.0001) and an interaction of Emotion and Region (F[2,60]=3.13, p=.05) indicated that N4 amplitudes were larger for neutral than surprised sounds over anterior (F[1,30]=5.23, p<.05), central (F[1,30]=10.5, p<.01) and, although marginally, over posterior (F[1,30]=3.09, p=.08) regions.

LPP. There was an Emotion main effect (F[1,30]=7.43, p<.05) as well as marginal and significant interactions of Emotion with Voiceness (F[1,30]=2.95, p=.09) and Emotion with Voiceness and Hemisphere (F[1,30]=4.99, p<.05), respectively. Over the left hemisphere, the Emotion by Voiceness interaction (F[1,30]=5.01, p<.05) indicated that the LPP was larger for surprised as compared to neutral vocal (F[1,30]=10.76, p<.01) but not nonvocal sounds (p>.1). Over the right hemisphere, the Emotion by Voiceness interaction was non-significant (p>.1). Instead, independent effects of Emotion (F[1,30]=7.68 p<.01) and Voiceness (F[1,30]=31.86, p<.0001) indicated that the LPP was larger to surprised than neutral and to vocal than nonvocal sounds.

Emotion also interacted with Hemisphere and Sex (F[1,30]=7.94, p<.01). In men
(F[1,15]=10.18, p<.01), the Emotion effect was significant at right (F[1,15]=5.22, p<.05) but not left hemisphere sites (p>.1). In women, the Emotion by Hemisphere interaction was non-significant (p>.1). Instead, there was a marginal Emotion effect (F[1,15]=3.68, p=.07) distributed across the scalp.

The factor Voiceness showed in a main effect (F[1,30]=37.72, p<.0001) as well as in interactions with Sex (F[1,30]=7.57, p<.01), Region (F[2,60]=32.38, p<.0001), Sex and Region (F[2,60]=8.26, p<.001), as well as Sex, Region, and Hemisphere (F[2,60]=5.5, p<.01). The latter effect was pursued for men and women separately. In men, the interaction of Voiceness, Hemisphere, and Region was non-significant (p>.1). Instead, there was a Voiceness by Region interaction (F[2,30]=5.43, p<.01) pointing to a Voiceness main effect over anterior (F[1,15]=9.71, p<.01) and central (F[1,15]=6.01, p<.05) but not posterior regions (p>.1) with larger amplitudes for vocal than for nonvocal sounds. In women, a significant interaction of Voiceness, Hemisphere, and Region (F[2,30]=3.92, p<.01) was followed up by exploring the Voiceness by Hemisphere interaction for each level of Region. Anteriorly and centrally, this interaction was non-significant (ps>.1). Instead, Voiceness main effects revealed larger LPP amplitudes to vocal than to nonvocal sounds (anterior, F[1,15]=52.8, p<.0001; central, F[1,15]=55.97, p<.0001). Over posterior regions, the Voiceness by Hemisphere interaction reached significance (F[1,15]=6.14, p<.05), however, follow-up tests were non-significant (ps>.1).