Human Screams Occupy a Privileged Niche in the Communication Soundscape

Highlights

- We provide the first evidence of a special acoustic regime ("roughness") for screams
- Roughness is used in both natural and artificial alarm signals
- Roughness confers a behavioral advantage to react rapidly and efficiently
- Acoustic roughness selectively activates amygdala, involved in danger processing

Authors
Luc H. Arnal, Adeen Flinker, Andreas Kleinschmidt, Anne-Lise Giraud, David Poeppel

Correspondence
luc.h.arnal@gmail.com (L.H.A.), david.poeppel@nyu.edu (D.P.)

In Brief
Arnal et al. show that, unlike speech, screams exploit a privileged acoustic attribute: "roughness." Sounds in this modulation regime specifically target subcortical brain areas involved in danger processing and improve behavior in various ways, suggesting that this acoustic niche may be preserved to insure efficient warning.
Human Screams Occupy a Privileged Niche in the Communication Soundscape

Luc H. Arnal,1,2,* Adeen Flinker,2 Andreas Kleinschmidt,1 Anne-Lise Giraud,3 and David Poeppel2,4,*
1Department of Clinical Neurosciences, University Hospital (HUG) and University of Geneva, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva, Switzerland
2Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
3Department of Neuroscience, University of Geneva, Biotech Campus, 9 Chemin des Mines, 1211 Geneva, Switzerland
4Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany
*Correspondence: luc.h.arnal@gmail.com (L.H.A.), david.poeppel@nyu.edu (D.P.)
http://dx.doi.org/10.1016/j.cub.2015.06.043

SUMMARY
Screaming is arguably one of the most relevant communication signals for survival in humans. Despite their practical relevance and their theoretical significance as innate [1] and virtually universal [2, 3] vocalizations, what makes screams a unique signal and how they are processed is not known. Here, we use acoustic analyses, psychophysical experiments, and neuroimaging to isolate those features that confer to screams their alarming nature, and we track their processing in the human brain. Using the modulation power spectrum (MPS [4, 5]), a recently developed, neurally informed characterization of sounds, we demonstrate that human screams cluster within a restricted portion of the acoustic space (between ~30 and 150 Hz modulation rates) that corresponds to a well-known perceptual attribute, roughness. In contrast to the received view that roughness is irrelevant for communication [6], our data reveal that the acoustic space occupied by the rough vocal regime is segregated from other signals, including speech, a pre-requisite to avoid false alarms in normal vocal communication. We show that roughness is present in natural alarm signals as well as in artificial alarms and that the presence of roughness in sounds boosts their detection in various tasks. Using fMRI, we show that acoustic roughness engages subcortical structures critical to rapidly appraise danger. Altogether, these data demonstrate that screams occupy a privileged acoustic niche that, being separated from other communication signals, ensures their biological and ultimately social efficiency.

RESULTS AND DISCUSSION
Screams result from the bifurcation of regular phonation to a chaotic regime, thereby making screams particularly difficult to predict and ignore [2]. While previous research in humans suggested that acoustic parameters such as “jitter” and “shimmer” [7–9] are modulated in screams, whether such dynamics and parameters correspond to a specific acoustic regime and how such sounds impact receivers’ brains remain unclear.

To characterize the spectro-temporal specificity of screams, we used the modulation power spectrum (MPS) (Figure 1). The MPS, beyond classical representations such as the waveform and spectrogram (Figures 1A and 1B, upper and middle panels), displays the time-frequency power in modulation across both spectral and temporal dimensions (Figures 1A and 1B, lower panels). The MPS has become a particularly useful tool in auditory neuroscience because it provides a neurally and ecologically relevant parameterization of sounds [5, 6, 15].

In speech, spectro-temporal attributes encode distinct categories of information, which in turn occupy distinct areas of the MPS (Figures 1B and 1C). For instance, whereas the fundamental frequency of the voice informs the listener about the gender of the speaker [5, 10, 16] (Figure 1C, blue region), slow temporal fluctuations carry cues such as the syllabic or prosodic information that underlie parsing and decoding speech to extract meaning [11, 12, 17] (Figure 1C, green region). Interestingly, the large region of the MPS that corresponds to temporal modulations between 30 and 150 Hz (orange zones in Figure 1C) has, to date, not been associated with any ecological function—and is generally considered irrelevant for human communication [6]. This spectro-temporal region corresponds to a perceptual attribute called roughness [13, 14]. Sounds in this region correspond to amplitude modulations ranging from 30 to 150 Hz and typically induce unpleasant, rough auditory percepts.

To ensure communication efficacy, screams should be acoustically well segregated from other communication signals. Conventional features that can further modulate or accentuate speech, such as increased loudness or high pitch, contribute to potentiate fear responses [18–20] but are not sufficiently distinctive, as these attributes accompany a wide range of utterances. Therefore, we conjectured that screams might occupy a dedicated part of the MPS, so that false alarms, i.e., confusions with non-alarm signals, are unlikely to occur. The roughness region (Figure 1C) is unexploited by speech, and therefore constitutes a plausible candidate space to encode alarm communication signals.
compared the MPS of screamed and spoken utterances with equivalent communicative content. We analyzed the MPS of four types of vocalizations, recorded from 19 participants, according to two factors: “scream” and “sentence” (Figures 2A and 2B). A two-way repeated-measures ANOVA was performed using the MPS of each vocalization. As hypothesized, screamed vocalizations contain stronger temporal modulations in the 30–150 Hz roughness window than do non-screamed ones (Figure 2C, left; averaged clusters statistic: F = 64.8, p = 2.5 × 10^-30). A two-way repeated-measures ANOVA was performed according to two factors: “scream” and “sentence” (Figures 2A and 2B). We next tested the hypothesis that roughness in screams is compared the MPS values in the roughness range of artificial instruments (e.g., strings or keyboards), which also have spectro-temporally complex structure but are not a priori designed to trigger danger-related reactions. This comparison (Figure 3A, center) reveals that alarm, but not musical, sounds exploit scream-like rough modulations (unpaired t test: p = 9 × 10^-10).

Roughness Is Exploited in Both Natural and Artificial Alarm Signals

We next tested the hypothesis that roughness in screams is selectively used to signal danger and should therefore not be exploited to the same degree in other kinds of communication signals. We performed a series of comparisons with other, vocal and non-vocal, stimuli. We first compared the average magnitude of temporal modulations in the roughness range (30–150 Hz) between sentential vocalizations (normal speaking), musical vocalizations (a cappella singing), and screaming (Figure 3A, left). The MPS values in the roughness range were significantly stronger in screams than in sung (unpaired t test: p = 6 × 10^-10) and spoken (unpaired t test: p = 8 × 10^-27) vocalizations. In order to explore whether rough sound modulations might be used in other languages, we compared the roughness index between English, French, and Chinese (Mandarin) neutrally spoken sentences. We found that roughness indices did not differ across languages (F = 0.04, p = 0.957; Figure S2) and were consistently smaller than those of screamed sentences in English (F = 24.97, p = 9 × 10^-14). Together, these results suggest that, regardless of communicative intention, only screamed vocalizations (whether sentential or not) maintain their invariant niche in the rough modulation regime.

If sound roughness is an effective feature for screams to constitute an alarm signal, it might also be exploited by man-made technological devices that generate non-biological acoustic signals to alert humans to danger. To address this, we compared the MPS values in the roughness range of artificial alarm signals (buzzers, horns, etc.; Table S1) to that of musical instruments (e.g., strings or keyboards), which also have spectro-temporally complex structure but are not a priori designed to trigger danger-related reactions. This comparison (Figure 3A, center) reveals that alarm, but not musical, sounds exploit scream-like rough modulations (unpaired t test: p = 9 × 10^-10).
The fact that roughness appears to be used in the design of artificial alarm signals in human culture, perhaps unwittingly, underlines both the perceptual salience and ecological relevance of rough sounds. This discovery is intriguing, as roughness is barely ever mentioned as a relevant feature in the applied acoustics literature on alarm signals [21].

Dissonant Intervals Elicit Temporal Modulations in the Rough Regime

The observation that there is a surprising convergence between roughness, screams, and dissonance.

Screams Roughness Confers a Behavioral Advantage to React Efficiently

We next addressed whether roughness is merely incidentally and epiphenomenally stronger in screams or whether this modulation window is universally exploited because of its causal relevance to behavior. We conjectured that if roughness informs conspecifics about danger, rough screams should induce more fearful subjective percepts than less rough vocalizations. To

Figure 2. Acoustic Characterization of Screamed Vocalizations

(A) Example spectrograms of the four utterance types, produced by one participant: screamed vocalizations, vowel [a] (top left); sentence (top right); neutral vocalizations, vowel [a] (bottom left); and spoken sentence (bottom right). (B) Average MPS across participants (n = 19) for each type. For the factorial analysis, the “sentence” factor (vertical dashed line) determines whether the utterance contains sentential information or the vowel [a]; the “scream” factor (horizontal dashed line) determines whether the utterance was screamed or neutral. (C) Main effect of “scream” (left) and main effect of “sentence” (right).

In (B) and (C), contours delimit statistical thresholds of p < 0.001 (Bonferroni corrected). See also Figure S1.
address this hypothesis, we asked 20 participants to rate the fear induced by screams and neutral vocalizations [a] on a subjective scale, ranging from neutral (1) to fearful (5). To assess the effect of rough modulations on perceived fear, we tested two additional conditions in which (1) we low-pass filtered screams’ temporal modulations in the roughness range (<20 Hz) and (2) we added rough temporal modulations to neutral vocalizations (see the Supplemental Experimental Procedures). As expected, the data showed (Figure 3B, left) that screams were perceived as more fearful than neutral vocalizations (paired t test: p = 4 × 10−3). Furthermore, screams were perceived as more fearful than filtered screams (paired t test: p = 4 × 10−4); in complementary fashion, modulation of neutral vocalizations in the roughness range increased perceived fear (paired t test: p = 0.045).

To test whether this effect generalizes to artificial alarm signals, we performed a similar experiment using the same acoustic alteration procedures on the set of artificial sounds. Thirteen participants rated the perceived “alarmness” on a subjective scale, ranging from neutral (1) to alarming (5). As found for human vocalizations, the data show (Figure S3) that alarm sounds are perceived as more alarming than instrument sounds (paired t test: p = 3 × 10−4). Also, alarm sounds were perceived as more alarming than filtered alarm sounds (paired t test: p = 0.035), whereas musical-instrument sounds modulated in the roughness range yielded increased perceived alarmness ratings (paired t test: p = 5 × 10−9).

Taken together, these results are consistent with the hypothesis that roughness contributes to induce an aversive percept, regardless of the nature (vocal or artificial) of the sound.

We further tested whether screams’ roughness scaled with subjective ratings, querying 11 participants who rated the perceived fear induced by scream recordings (Table S3). The data reveal (Figure 3B, middle) that the rougher the screams, the more fearful the induced emotional reaction (Pearson’s r = 0.65, p = 10−8). Interestingly, the speed of behavioral responses (Figure 3B, right) also scaled with scream roughness (Pearson’s r = −0.35, p = 0.005). Roughness hence not only increases the perceived fear valence of screams, but also enables a faster appraisal of danger.

Rapid, accurate evaluation of danger (as indexed by the valence of screams) is presumably crucial for adaptive behavior. In that context, the precise location of the scream source in the environment is of critical relevance. To assess whether roughness improves the ability to localize vocalizations, we implemented a spatial localization behavioral experiment. We measured in 21 participants the speed and accuracy to detect whether normal vocalizations and screams were presented on their left or right sides using inter-aural time-difference cues. In addition to natural vocalizations, we also tested a control set of synthetic screams, constructed by modulating neutral vocalizations in the roughness range (Figure S4). As anticipated, accuracy and speed varied as a function of vocalization type (Figure 3C, left and center panels; repeated-measures ANOVA, for accuracy: F(2,40) = 7.01, p = 0.004; reaction speed: F(2, 40) = 5.8, p = 0.006). Participants were both more accurate and faster at localizing natural (paired t test, for accuracy: p = 3 × 10−6;
The relevance of roughness for auditory processing of danger, Responses in the Amygdala

Rough Temporal Modulations Induce Selective vocalizations. These findings plausibly suggest that rough vocalizations recruit dedicated neural processes that prioritize fast reactions to danger versus neutral sounds (regardless of sound category). This analysis revealed that rough sounds induce larger hemodynamic responses in the bilateral anterior amygdala and primary auditory cortices (Figure 4A and Table S4). To determine whether these regions encode specific subparts of the MPS, we implemented a reverse-correlation approach and related single-trial blood-oxygen-level-dependent response estimates with the MPS of the corresponding sound (after removal of the variance explained by the valence of the stimuli, as indexed by individual participant ratings; see [26]). We found that the amygdala—but not auditory cortex—is specifically sensitive to temporal modulations in the roughness range (Figure 4B). These results demonstrate that rough sounds specifically target neural circuits involved in fear/danger processing [27, 28] and hence provide evidence that roughness constitutes an efficient acoustic attribute to trigger adapted reactions to danger.

In this series of acoustic, behavioral, and neuroimaging experiments, we characterized the spectral modulation of various natural and artificial sounds and demonstrated the ecological, behavioral, and neural relevance of roughness, a well-known perceptual attribute hitherto unrelated to any specific communicative function. The findings support the view that roughness, as featured in screams, improves the efficiency of warning signals, possibly by targeting sub-cortical neural circuits that promote the survival of the individual and speed up reaction to danger.

EXPERIMENTAL PROCEDURES

A bank of sounds containing several types of human vocalizations (screams and sentences), artificial sounds (alarm and instrument sounds), and sound intervals (pure tone intervals) was constructed for subsequent acoustic characterization. Sounds were edited to last 1,000 ms and were root-mean-square normalized. In order to quantify the power in temporal and spectral modulations, the two-dimensional Fourier transform of the spectrogram was calculated to obtain the MPS of each sound [6].

A repeated-measures ANOVA (n = 19 speakers) was performed on the vocalizations’ MPS to test for specific scream and sentence effects. After identifying a restricted window in the roughness domain (30–150 Hz) for screamed vocalizations, we compared the averaged MPS values in this window between the different categories of the sound bank using ANOVAs and unpaired t tests.

The influence of MPS values in the roughness range was assessed in four behavioral experiments. The first three experiments tested the relationship between roughness and behavioral ratings in both natural and artificial sounds. The fourth experiment tested the influence of roughness on the spatial localization of vocalizations. We measured the localization performance, reaction times, and efficiency during the perception of lateralized vocalizations [a], screams, and synthetic screams (100-Hz amplitude modulated vocalizations [a]).

Finally, we used fMRI to explore the neural structures implicated in the processing of such sounds. We executed a sparse-sampling experiment in which participants rated the unpleasantness (on a 1–5 scale) of three types of sounds (human vocalizations, artificial sounds, and tone intervals). After identifying the
brain regions that responded to the unpleasantness of these sounds, we used a reverse-correlation approach to investigate the relative hemodynamic sensitivity of these regions to sub-regions of the MPS.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.06.043.

AUTHOR CONTRIBUTIONS

L.H.A. designed the experiments, performed the research, analyzed the data, and wrote the manuscript. A.F. contributed to analysis tools. A.K., A.-L.G., and D.P. wrote the manuscript. Correspondence and requests for materials should be addressed to L.H.A. and D.P.

ACKNOWLEDGMENTS

We thank Jan Manent for useful discussions; Jess Rowland, Tobias Overath, Jean M. Zarate, Mariane Haddad, and Josh Barocas for technical assistance; and Gregory Hickok, Shihab A. Shamma, Gregory B. Cogan, Nai Ding, and Keith B. Doelling for comments on the manuscript. This work was supported in part by the Fondation Fyssen and the Philippe Foundation (L.H.A.), the Fondation Louis Jeantet (L.H.A. and A.K.), 1F32DC011985 (A.F.), and 2R01DC05660 (D.P.).

Received: March 12, 2015
Revised: May 21, 2015
Accepted: June 17, 2015
Published: July 16, 2015

REFERENCES