Flexibility in wild infant chimpanzee vocal behavior

Guillaume Dezecache[a,b,c,d,*], Klaus Zuberbühler[a,b,e], Marina Davila-Ross[c] & Christoph D. Dahl[a,f,g,*]

[a]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland;

[b]Budongo Conservation Field Station, Masindi, Uganda;

[c]Department of Psychology, University of Portsmouth, Portsmouth, England, United Kingdom;

[d]Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France;

[e]School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland, United Kingdom;

[f]Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, Taipei, Taiwan;

[g]Brain and Consciousness Research Center, Taipei Medical University Shuang-Ho Hospital, New Taipei City, Taiwan

[*]Corresponding Authors:

Guillaume Dezecache <guillaume.dezecache@gmail.com>, LAPSCO CNRS 17 Rue Paul Collomp, 63000 Clermont-Ferrand, France;

Christoph D. Dahl <christoph.d.dahl@gmail.com>

Flexibility in wild infant chimpanzee vocal behavior

ABSTRACT

How did human language evolve from earlier forms of communication? One way to address this question is to compare prelinguistic human vocal behavior with nonhuman primate calls. An important finding has been that, prior to speech and from early on, human infant vocal behavior exhibits functional flexibility, or the capacity to produce sounds that are not tied to one specific function. This is reflected in human infants' use of single categories of protophones (precursors of speech sounds) in various affective circumstances, such that a given call type can occur in and express positive, neutral, or negative affective states, depending on the occasion. Nonhuman primate vocal behavior, by contrast, is seen as comparably inflexible, with different call types tied to specific functions and sometimes to specific affective states (e.g., screams mostly occur in negative circumstances). As a first step towards addressing this claim, we examined the vocal behavior of six wild infant chimpanzees during their first year of life. We found that the most common vocal signal, grunts, occurred in a range of contexts that were deemed positive, neutral and negative. Using automated feature extraction and supervised learning algorithms, we also found acoustic variants of grunts produced in the affective contexts, suggesting gradation within this vocal category. By contrast, the second most common call type of infant chimpanzees, the whimpers, was produced in only one affective context, in line with standard models of nonhuman primate vocal behavior. Insofar as our affective categorization reflect infants' true affective state, our results suggest that the most common chimpanzee vocalization, the grunt is not affectively bound. Affective decoupling is a prerequisite for chimpanzee grunts (and other vocal categories) to be deemed 'functionally flexible'. If later confirmed to be a functionally flexible vocal type, this would indicate that the evolution of this foundational vocal capability occurred before the split between the Homo and Pan lineages.

Keywords: language evolution, vocal flexibility, affective decoupling, wild chimpanzees, *Pan troglodytes*, grunts

1. INTRODUCTION

At some point in evolutionary history, there must have been a transition from primate-like to human-like acoustic communication, which may have coincided with the origins of speech. The evolutionary history of this transition continues to be vividly debated (Fitch, 2018), with a large range of comparative evidence from animal communication systems, and the consensus view that direct evolutionary homologies are generally absent in the primate order (Rendall & Owren, 2002). More recently, however, some vocal and neural equipment has been identified in different primate species that allow for the production of speech-like sounds (Boë et al., 2017; Fitch et al., 2016; Lieberman, 2017) and for some degree of control over vocal fold oscillation (Lameira & Shumaker, 2019). If the facial and gestural displays have undeniably played a crucial role in the evolution of language (Arbib et al., 2008; Pollick & Waal, 2007; Waal & Pollick, 2011), vocal production appears so strongly different in humans and other primates that the quest for evolutionary precursors of human vocal production has been and will continue to be particularly fruitful.

One key divergence between humans and other species, according to Oller and colleagues (2013), lie in the ontogenetic trajectories between non-human primate and human vocal behavior prior to speech. By the age of one month old (and possibly earlier, see Oller et al., 2019), human infants produce sounds that are not tied to the expression of one particular need, such that they can hold different illocutionary force on different occasions, and cause appropriate reactions in caregivers (Jhang & Oller, 2017; Oller et al., 2013). This is reflected in infants' use of squeals, vocants and growls in circumstances associated with positive, negative, or neutral affective states, such that those sounds are not bound to the experience of one particular type of affect (Oller et al., 2013). By contrast to those vocal types, human infants use laughter and cries in single affective contexts (positive and negative, respectively).

This capacity to produce one vocal unit under a variety of affective state (such that vocal production is independent from the experience of a specific affective state – a capacity termed 'affective decoupling') later allows infants to use those sounds as they wish, and to express a variety of needs on different occasions (Oller et al., 2013; Oller & Griebel, 2004). This decoupling between vocal production and affective experience is foundational for the ability to produce sounds that can later fulfil a variety of functions ('vocal functional flexibility'), that is, they can be used to invite a variety of responses in others (Oller & Griebel, 2004). For instance, a given utterance (such as 'the train is arriving') produced with neutral (a mere statement), negative (annoyance) and positive (exultation) tones have the same syntactic structure and semantic content, but are produced under antagonist affective states and cause vastly different responses in receivers. Without affective decoupling and vocal functional flexibility, basic speech acts cannot take place.

This decoupling of signal structure and affect in young infants' vocal repertoire has thus been identified as a major evolutionarily precursor to language (Oller et al., 2013). Because of their early ontogenetic onset, affective decoupling and vocal functional flexibility may be more foundational to human speech than other building blocks of the language faculty, such as proto-syntax or vocal elaboration (Oller et al., 2013). These capacities, in this view, are prerequisites for speech development, and major evolutionary departures from the affect-bound and functionally inflexible vocal behavior of non-human primates (Waal & Pollick, 2011). By contrast to their primate cousins, humans would have evolved in a social ecology conducive to the development of such vocal flexibility. Notably, it is proposed that functionally flexible vocalizations of young human infants have evolved in humans in relation to allo-maternity (Burkart et al., 2009, 2009, 2009; Hrdy, 2007; Kramer, 2010; Schaik & Burkart, 2010) or altriciality (Locke, 2006) and associated pressures on young infants to signal their needs and attract caregivers (Ghazanfar et al., 2019; Locke, 2006; Zuberbühler, 2012). Other species living in cooperative breeding systems (such as the marmosets (Burkart et al., 2007)) may display vocal functional flexibility in their vocal repertoire.

For long, primate (but also animal) communication systems have been characterized as affectively-biased, such that affect has been proposed to be both necessary and sufficient for vocal production to occur. According to Hammerschmidt and Fischer, there could even exist "[a] correspondence between non-verbal vocalizations in humans and non-human primates [such] that they both function to communicate the affective state of the signaler." (Hammerschmidt & Fischer, 2008, p. 103). In fact, a number of researchers have held the view that the equivalents of animal vocalizations are non-verbal affective expressions in humans, such as laughing, screaming and crying (see Gruber & Grandjean, 2017 and Marler, 1980 for a discussion). Examining the neural pathways of vocal production in squirrel monkeys, Jürgens (Jürgens, 1976; Jürgens, 1979) concluded that vocal production was mediated by affect. More recently in marmoset infants, Zhang & Ghazanfar (Zhang & Ghazanfar, 2016) found that fluctuations in cardiac rhythm shape respiratory patterns, which in turn contribute to vocal production, thereby attributing a central role to affect in early vocal production of this species (Tchernichovski & Oller, 2016). The vocal repertoire of species phylogenetically closer to humans (such as the chimpanzees) did not evade this conclusion. In her authoritative book on the behavior of Gombe chimpanzees, Goodall (Goodall, 1986) wrote that 'chimpanzee vocalizations are closely bound to emotion' and that 'the production of a sound in the *absence* of the appropriate emotional state seems to be an almost impossible task for a chimpanzee' (p. 125). As a consequence, she proposed a mapping between call and affect when describing the vocal repertoire of chimpanzees, with, for instance, a one-to-one correspondence between the experience of annoyance and the production of 'soft barks' (p .127).

To which degree is vocal production affectively bound in other species? To which extent can the developmental trajectory seen in humans (with early independence between certain sounds and particular affective states (Oller et al., 2013)) also be observed in other primates? In fact, are affective decoupling and vocal functional flexibility unique to human vocal ontogeny? In one relevant study, Clay et al. (Clay et al., 2015) examined 'peep' calls in mature bonobos (*Pan paniscus*), their most common vocalizations, and found that they are produced in a variety of contexts, ranging from seemingly positive (food provisioning) to neutral (travel and resting) and negative (agonistic and alarm) situations. Based on these findings, the authors concluded that bonobos have the capability to produce sounds that are not affectively biased (Clay et al., 2015), suggestive of affective decoupling in vocal production. Their peeps were, however, attributed to broad behavioral contexts (such as feeding or travelling) with no focus on more specific and transient behaviors that may help infer affective contexts, such as when individuals suddenly experience aggression during travelling and feeding bouts. In fact, peeps could well be bound to the expression of one particular affect, which could be common in both feeding and travelling contexts for example. As such, the bonobo data are indicative of their peeps occurring across broad behavioral contexts ('contextual flexibility') but may ultimately remain inconclusive in regards to whether affective decoupling and vocal functional flexibility are indeed present in species other than humans.

Similarly, the flexibility with which some call types are expressed in some primate species is only *suggestive* of 'vocal functional flexibility' (the use of one vocal type to convey various illocutionary forces on different utterances), and may only correspond to contextual flexibility (the use of one call type in different contexts, with core commonalities in the illocutionary force conveyed by all utterances). For example, Guinea baboons use a number of calls (e.g., grunts, roar grunts, barks and wahoos) in a diversity of contexts (Maciej et al., 2013). Guinea baboons' grunts are used in foraging and travelling contexts, but also affiliative, infant handling and greeting contexts. Does that mean that Guinea baboons show functional flexibility when producing grunts? It is a possibility. However, and in the absence of a methodological focus on potential affective states experienced by the animal, a temporary conclusion is that Guinea baboon grunts likely are 'contextually flexible'. The possibility that they also are not affectively bound (i.e., not bound to the experience and expression of a particular affective state) or 'functionally flexible' (i.e., not assigned to the fulfilment of one particular function) awaits empirical confirmation, for grunts in Guinea baboons could well be expressed under one particular affective state, and be used to meet one single function in a variety of contexts (e.g., maintaining contact with other troop members). In fact, a first step could be made by examining whether inferences about the affective state of animals (e.g., Guinea baboons) can be conducted using the behavioral contexts employed to describe the contextual occurrence of their vocalizations, and whether such analysis suggests that these vocalizations qualify as affectively decoupled.

A second study, also on bonobos (Oller et al., 2019), suggests protophone-like vocal behavior with bonobo infants producing calls that occur in both low or moderate arousal situations, implying no affective binding. This conclusion has been preliminary, however, for the affective quality of the contexts surrounding vocalizations (a reliable marker of illocutionary force and needs in human infants) has proven difficult to discern.

Here, we intended to provide a first evaluation of affective decoupling in infant chimpanzees' (*Pan troglodytes schweinfurthii*) vocal behavior at a very early age (< 12 months). Given the recent studies in both immature and mature bonobos, focusing on the other closest living relatives, the chimpanzees, is crucial to test hypotheses about the evolutionary origins of functionally flexible vocal behavior. What's more, examination of *early* vocal production is critical for a more direct comparison with findings on human infants (Oller et al., 2013). We focused on two call types, the grunts and the whimpers, as they are acoustically very distinct vocalization categories that are common in young infants (Plooij, 1984). Finally, we tried to approach the affective dimension of the context of calling by focusing on transient behavioral cues (e.g., the infant escaping a situation) rather than broader behavioral contexts (e.g., travelling context).

Grunt calls are of particular importance as they develop into a central component of the vocal repertoire of chimpanzees and contribute to a variety of vocal sequences produced by juveniles, sub-adults and adults (Crockford & Boesch, 2005). For example, grunts complement panting elements during laughter (Leavens, 2009) and when encountering dominant individuals ('pant-grunts') (Laporte & Zuberbühler, 2011; Laporte & Zuberbühler, 2010). They are also produced upon encountering a food patch or when joining a foraging party ('rough grunts') (Fedurek & Slocombe, 2013; Schel et al., 2013; Slocombe et al., 2010; Slocombe & Zuberbühler, 2005; Watson et al., 2015). Finally, they are routinely produced throughout resting or in relaxed social activities (Goodall, 1986). Grunts are produced from the first days of life in chimpanzees. Their ontogenetic development has already been studied to some degree in chimpanzees, which has shown some flexibility in usage (Laporte & Zuberbühler, 2011). It has been suggested that at least two types of grunts could be distinguished. First, uh-grunts are short, tonal sounds, resembling human vowels {u}, {o} and {a} (and possible homologous to quasi-vowels in humans), sometimes produced in short series (staccato-grunts) (Kojima, 2003; Plooij, 1984). The second type are the so-called 'effort' grunts, which are common in immature chimpanzees (Plooij, 1984) and are also present in adult chimpanzees, mature and immature humans and other mammals (McCune et al., 1996). So-called ‘effort grunts’, are very soft and require the close presence of observers to be reliably heard (Plooij, 1984). They received their name from their presence during locomotor activities. Despite Plooij’s suggestion that they could be mere by-products of locomotor activities, he also noted they can occur in the absence of movements (Plooij, 1984). So far, no study has yet offered an acoustical validation of the existence of these diverse types, such that we (and others, see Laporte & Zuberbühler, 2011) cannot rely on this distinction.

Another common vocal utterance produced by chimpanzee infants is whimpers (Dezecache et al., 2020; Levréro & Mathevon, 2013; Plooij, 1984). They are short, tonal and often produced in series with an upward shift in fundamental frequency. Contrarily to grunts, whimpers preferentially occur in aversive contexts, likely homologous to human crying or distress calls in other mammals (Plooij, 1984). Previous research (e.g., Plooij, 1984) has suggested the presence of whimper subtypes (single, serial and human-like whimpers), but again, we are not aware of any systematic acoustical analysis that would justify this nomenclature. Whimpers are also present in the repertoire of adult chimpanzees, notably in alarm (Tsukahara, 1993), food begging (Crockford & Boesch, 2005; Slocombe & Newton-Fisher, 2005), and physical separation (Crockford & Boesch, 2005) contexts.

To start addressing the hypothesis that affective decoupling and vocal functional flexibility evolved before the split between *Pan* and *Homo* lineages, we examined the vocal behavior of six wild chimpanzee infants aged between 0-12 months old from the Sonso community of Budongo Forest, Uganda. We analyzed the extent to which vocal production of grunt-like and whimper-like vocalizations were occurring with so-called positive, negative or neutral behaviors, as a first step towards evaluating the affective quality of the vocalization contexts.

We also took advantage of recent developments of machine learning techniques to the study of animal communication (Fedurek et al., 2016; Mielke & Zuberbühler, 2013) to evaluate acoustical differences between calls produced with positive, negative and neutral markers.

2. METHODS

2.1 Ethics

Permission to conduct the study was obtained from the Ugandan Wildlife Authority (UWA) and the Uganda National Council for Science and Technology (UNCST).

2.2 Subjects and data collection

Data were collected in the Sonso community of the Budongo Forest Reserve, Uganda (Reynolds, 2005) between February-June 2014, December 2014 and March-June 2015. This community comprises around 70 individuals well habituated to human observers. The natural behavior of N = 7 infants was video recorded continuously during focal animal sampling (Altmann, 1974), using Panasonic HC X909/V700 cameras, with a Sennheiser MKE-400 shotgun microphone. Six of those infants produced enough calls to be further considered for data analysis (see Table 1 for details).

2.3 Behavioral data analysis

Videos were inspected for the presence of infant vocalizations. We defined vocal behavior as the occurrence of single sound units or series of sounds produced by the infant's vocal apparatus, separated by a least 5 seconds of silence.

As of today, there is no definitive repertoire of infant chimpanzee vocal behaviors, only suggestive classifications (Plooij, 1984; Plooij et al., 2014). The categories used in this research are based on First Author's assessment. This assessment proved reliable when confronted to an independent assessment with Derry Taylor, using vocalizations from infant and juvenile semi-wild chimpanzees from the Chimfunshi Wildlife Orphanage, Zambia, collected by DT. One hundred-and-sixty vocalizations were indeed classified as belonging to either the 'grunt', 'whimper', 'scream' or 'laughter' category. Agreement was excellent (k = 0.77) and even better when considering only 'grunts' and 'whimpers' (k = 0.92).

For each vocal occurrence, we coded infant behavior from a list of mutually exclusive behaviors (summarized in Table 2). This list was established following data collection, with some inspiration from the behavioral categories established by Plooij during his study with the infant chimpanzees of the Gombe community between 1971 and 1973 (Plooij, 1984). As in the original human study (Oller et al., 2013), we reckoned the behavior of the infants could offer a reliable source of information unto their affective state, as a first step towards establishing affective descriptions of contexts. In fact, we originally aimed at mimicking their coding strategy, using categories appropriate to the study of wild infant chimpanzees. The affective quality of the infants' behavior was classified as 'positive' if it showed one of the following four behaviors: (1) 'play' (2) giving or receiving 'grooming' (note that allo-grooming was never observed in our infants); (3) 'feeding', and (4) 'social approach'. See Table 2 for details.

The affective context was classified as 'neutral' if it showed one of the following behaviors: (5) 'resting'; (6) 'moving'; (7) 'manipulating objects' without playful postures, or (8) 'greeting without approach'. See Table 2 for details.

Infant behavior was classified as 'negative' if it showed one of the following behaviors: (9) 'nuzzling'; (10) 'begging'; (11) 'hiding'; (12) 'contact mother/kin' was coded if infants were urgently seeking contact with the mother or a kin when contact was not already established between them; (13) ‘escaping’. See Table 2 for details.

We performed intra-coder reliability tests on the affective contexts coded as positive, neutral and negative. For this, we randomly selected 200 video clips (around 19% of the coded dataset composed of the 7 infants), which were coded independently during two coding sessions more than a year apart (November 2015 and February 2017), so that the second coding was, notably, naïve. We found strong agreement between the two coding sessions (k = 0.73).

In order to evaluate the evenness of the distributions of grunts and whimpers across affective contexts, we calculated, for each infant, and for grunts and whimpers separately, the dominance of one affect over the two others, using the Berger-Parker Dominance index (see Morris et al., 2014):

$$dominance = N_{max} / N$$

where N_{max} is the number of calls in the most abundant affective context; N the total number of calls across all affective contexts. Dominance values range from 1 / number of affects (= equiprobability of calls across affects; here 1 / 3 = 0.33) to 1 (= complete dominance of one affective context over the others).

Dominance values (one per infant per call type) were compared between grunts and whimpers using a paired Wilcoxon Sign-Ranked test. These analyses were carried out using R (version 3.6.1; R Core Team, 2018) and R Studio (version 1.2.1335; RStudio Team, 2015).

2.4 Acoustic analysis

Acoustic data analysis focused on grunts for they were the only vocal category for which at least two of the affective contexts were well represented. The acoustic structure of whimpers has been analyzed as part of another study (see Dezecache et al., 2020). N = 180 grunts were extracted from independent vocal behaviors. For each affective context, 60 were randomly selected. Following extraction, we used MATLAB (MathWorks Inc., Natick, MA, USA) for the acoustic data analysis, consisting of features extraction, feature selection and call classification. We first pre-processed the audio files by applying a band pass filter from 50 to 4000 Hz and normalized the signals using the following function:

signal = (signal - mean(signal)) / max(abs(signal - mean(signal)))

2.4.1 Feature extraction and selection

We first ran a feature extraction algorithm to reduce redundancy of information and computational efforts in classifying the grunts and to maximize the generalization ability of the classifier (Tajiri et al., 2010). A popular method is extraction of mel frequency cepstral coefficients (MFCCs) (Supplementary Figure 2). MFCCs represent the envelope of the short-time power spectrum, as determined by the shape of the vocal tract (Logan, 2000). The idea behind the extraction of MFCCs is to obtain a comprehensive representation of the frequencies that compose an audio bout, while putting emphasis on certain frequency bands. While a typical spectrogram linearly scales frequencies (i.e., each frequency bin is spaced an equal number of Hertz apart), the mel-frequency scale is a logarithmical spacing of frequencies. MFCCs is routinely used in speech recognition and is gaining prominence in the field of animal communication (see for instance Fedurek et al., 2016 in chimpanzees). The use of MFCCs to represent sounds can be considered to be a solution preferable to the selection of a limited set of parameters to describe acoustical phenomena (such as these related to the shape of the fundamental frequency) for it offers a more comprehensive representation of sounds. In the context of our work (the aim of which was to evaluate potential distinctiveness between grunts occurring in so-called positive, neutral and negative contexts), MFCCs appeared as the optimal solution to the problem of a false negative conclusion.

We divided the calls into segments of 25ms length and 10ms steps between two successive segments. We warped 26 spectral bands and returned 13 cepstra, which resulted in feature dimensions of 13 values each. We then took the mean and co-variances of each cepstra over the collection of feature segments, resulting in a 13-value vector and a 13 x 13-value matrix, respectively, and concatenated to 104-unit vectors (Mandel & Ellis, 2005, p. 594-599) (Figure 3). We applied feature scaling to [0 to 1] and mean normalization.

Second, we performed a feature selection procedure: too many feature dimensions are not useful for producing reliable classification systems, whereas low sample numbers can lead to over-fitting to noisy feature dimensions. We therefore selected a subset of the original feature dimensions and evaluated classification performance based on sequentially selected feature sets until there was no improvement in performance. At this end, we subdivided the entire data set into a training (75%) and a test data set (25%) and applied a *t*-test on each feature dimension, comparing values of given feature dimension sorted by predefined class labels (e.g., grunts occurring with negative (1) vs. positive (2) affects) and used *p*-values as a measure separability of the two classes. We plotted the *p*-values as an empirical cumulative distribution function (eCDF) to get an understanding of how well each feature separated the two classes and how many features contributed to a significant separation (5%-level). We ran this procedure 20 times for each comparison and plotted the results individually (gray lines) and the mean of all repetitions (black line) (Figure 2A). The classification routines were then independently run either on feature dimensions selected according to the discrimination power (decreasing order) (orange lines in Figure 2B), as shown in the eCDF plots (Figure 2A). Such procedure is referred to as a simple filter approach on feature selection, where general characteristics of the extracted features are taken into consideration when selecting feature dimensions, without subjecting them to a classifier. We also applied a more extensive procedure of feature selection by sequentially selecting feature dimensions by adding (forward search) feature dimensions, referred to as sequential feature selection (black lines in Figure 2B). As part of this method, the algorithm searched the best feature dimensions (predictors) according to their individual classification performance in the given subset of data. For each candidate feature subset (predictor), the algorithm performed a 10-fold cross-validation procedure with different training and test subsets. After computing the mean performance values for each candidate feature subset, the algorithm chooses the candidate feature subset with minimal misclassification. For both methods, we systematically varied the number of features used for classification (x-axis in Figure 2B). The selected features from a single run of the sequential search algorithm are illustrated in Figure 2C. Scales reflect the feature-scaled and normalized values, as a result of feature extraction, from which the grand means (i.e. for each feature dimensions across all data) were subtracted. This measure was used to visually highlight differences and was not used in further analyses.

2.4.2 Classification

We used support vector machine (SVM) with a radial basis function (RBF) Kernel (Vert et al., 2004) for the classification of calls according to the class labels (so-called negative, neutral and positive affective contexts). A classification procedure contains a training phase followed by a test phase. We separated training samples and labelled them according to an attribute of interest (e.g., negative (1) vs. positive (2) affective contexts). The algorithm then created a model that optimally separates the two classes. In the test phase, samples without attribute labels were fed into the model to measure its generalization performance. We used the SVM implementation from LIBSVM toolbox (Chang & Lin, 2011). To evaluate how the classification results generalize to a novel and independent data set, we 10-fold cross-validated the classification process and optimized the parameters C and gamma (Fedurek et al., 2016), with the C taking values in a range of [2^{-1}, 2^{3}] and gamma in a range of [2^{-4}, 2^{1}]. In addition, to ensure that no single individuals contributed solely to the classification outcome, we ran a leave-one-out algorithm, where the procedure described above was re-run six times, excluding one of the individuals in each run. We applied one-sample *t*-tests to compare the classification scores with a 50% baseline condition. The 50% baseline results from the pairwise comparisons of affective contexts (positive, neutral, negative). To ensure samples were normally distributed (a key assumption behind the use of one-sample *t*-tests), we used Lilliefors test prior to each comparison at a significance level of 5%. In cases where data samples were not normally distributed, we used a one-sample Kolmogorov-Smirnov test. All reported *p*-values were adjusted for multiple comparisons using Bonferroni corrections.

2.4.3 Feature evaluation

To evaluate whether certain feature dimensions are particularly critical for the classification of grunts, we assessed whether feature dimensions have been repeatedly used by the classifier overall in the classification of grunts. We therefore considered the three types of comparisons, positive vs neutral, positive vs negative and neutral vs negative grunts, as well as the two feature evaluation algorithms (simple feature selection and sequential feature selection). Each comparison was ten-fold cross-validated. We then calculated the empirical distribution of the ten features with best classification power, as determined by the feature selection algorithms (see above). Also, we determined a random distribution of "best features" for each comparison by randomly selecting 10 out of 104 features. The frequency distribution across all comparisons were determined and 95% confidence intervals were calculated by running the procedure 1,000 times. We then traced back the significant feature dimensions to the underlying frequency bands in Hertz.

3. RESULTS

3.1 Types of vocal utterances

We inspected N = 1,016 vocal occurrences, of which N = 967 could be classified as either 'grunts' (N = 833) (corresponding to a rough, harsh and noisy sound) or 'whimpers' (N = 134) (usually a series of low-pitch tonal calls with increase in fundamental frequency throughout the series). Other types of calls were identified as 'hoos' (n = 23), 'pants' (n = 15), 'screams' (n = 2), 'squeaks' (n = 2), 'barks' (n = 4) and 'laughter' (defined as grunting and panting) (n = 3).

3.2 Distribution of grunts and whimpers across so-called affective contexts

Grunts: 44.8% of grunt-like vocalizations co-occurred with contexts we classified as 'positive', 40.9% with 'neutral', and 14.3% with 'negative'. When considering each individual separately, a similar picture emerged (Figure 1), with most grunt-like vocalizations co-occurring with 'positive' and 'neutral' contexts. We found dominance to be relatively low in grunts, varying from 0.37 and 0.63 (mean = 0.53; SD = 0.10), suggesting a stable and relative evenness in the affective distribution of grunts, such as defined by our coding system (see Table 2).

Whimpers: 94.8% of whimpers co-occurred with negatively classified contexts, and rarely with neutral (4.5%) or positive (0.7%) affects. Inspection of individual distributions revealed the same pattern with whimper-like vocalizations systematically co-occurring with negatively classified contexts (Figure 1). The dominance of one affective context over the others in whimpers was relatively high, ranging from 0.89 to 1 (mean = 0.96; SD = 0.05), indicating low evenness in the affective distribution of whimpers.

Grunts vs. Whimpers: When comparing the distributional evenness of grunts vs. whimpers, we found dominance to be statistically higher in whimpers than in grunts (paired Wilcoxon signed rank test: V = 21, *p* = .031).

3.3 Acoustic variants of grunts

We classified the N = 180 grunts (N = 60 per affective contexts) according to their association with so-called positive, neutral, negative contexts in order to test for the presence of acoustic variants. In the first step, we followed the feature extraction procedure by extracting the means and covariances of MFCCs for each call, and compared these values according to the calls' associations (e.g. positive vs. negative) using *t*-tests. We displayed the resulting *p*-values in an empirical cumulative distribution function (eCDF) (Figure 2A). We found that 5-10% of all features showed significant differences between the class labels at a 5%-significance level. In other words, 5-10 of 104 feature dimensions had strong discrimination power to distinguish between grunts pertaining to the various affective contexts.

With the simple feature selection algorithm, the SVM correctly discriminated between classes at up to 80% (positive vs. neutral: M = 78.99, SD = 3.53, *t*(59) = 63.69, *p* < .001; positive vs. negative: M = 79.58, SD = 1.83, *t*(59) = 125.37, *p* < .001; neutral vs. negative: M = 80.44, SD = 2.06, *t*(59) = 114.26, *p* < .001; orange lines in Figure 2B). A substantial improvement was found when sequentially selecting feature dimensions: SVM correctly classified samples at up to 95% (positive vs. neutral: M = 89.56, SD = 4.84, *t*(59) = 143.42, *p* < .001; positive vs.

negative: M = 88.72, SD = 4.49, *t*(59) = 153.11, *p* < .001; neutral vs. negative: M = 84.27, SD = 5.23, *t*(59) = 124.91, *p* < .001; black lines in Figure 2B). For all comparisons chance levels were 50% due to the two-class comparisons applied. We, therefore, used one-sampled *t*-tests. The classification scores in all (but one) comparisons fulfilled the requirement of normal distribution. The first comparison (feature-selection algorithm, positive vs. neutral) was not conform with a normal distribution and was, thus, re-evaluated using a one-sampled Kolmogorov-Smirnov test, resulting in the following values (ks = 0.17; *p* < .001).

We further illustrated the simple feature selection outcomes by highlighting the feature dimensions selected (circles in Figure 2C) among the feature dimensions not selected (gray dots). Further, the features selected via the sequential feature selection are marked with x's. The sequential feature selection yields better performance through sequential combinations of feature dimensions that, on average, fall more distal to the diagonal mid-line than the feature dimensions selected by the simple feature selection process. Sequential feature selection, to a large extent, included feature dimensions not selected by the simple feature selection method.

We further ensured that each individual was not contributing solely to the classification results of various contrasts. As can be seen in Supplementary Figure 1, the classification performance did not improve nor deteriorate systematically when one individual was removed at a time, suggesting no effect due to caller identity (the average *t*-value of one-sample *t*-tests is 97.52 +/- 30.25 (SD); all *p*-values were smaller than .001).

The use of means and covariances of cepstra yielded relatively high-performance scores in the classification routines at low computational loads. To assess whether certain feature dimensions (means and covariances of cepstra) occurred above chance across all comparisons, we determined the empirical distribution of occurrences of feature dimensions and contrasted it with a random distribution. While the use of the same feature dimension in up to 33% of the comparisons was not significantly different in the empirical distribution from the random distribution, the use of the same feature dimension in 50% of comparisons was significantly increased in the empirical distribution (Figure 3A).

To describe the frequency bands explaining significant variances between classes of calls, we traced back the frequency bands underlying the significant feature dimensions, i.e., covariances of cepstra, and determined the sign of the covariances. We found negative covariances between the following frequency bands (Figure 3B): (1) band 2 (196.30 to 488.89 Hz) and band 4 (488.89 to 927.78 Hz), (2) band 4 (488.89 to 927.78 Hz) and band 8 (1074.07 to 1366.67 Hz), band 6 (781.48 to 1074.07 Hz) and band 9 (1220.37 to 1512.96 Hz). We found a positive covariance between the frequency bands 9 (1220.37 to 1512.96 Hz) and 10 (1366.67 to 1659.26 Hz). Mean cepstra were significantly contributing in the frequency bands from (1) 50 to 342.59 Hz, (2) 196.30 to 488.89 Hz, (3) 927.78 to 1220.37 Hz.

4. DISCUSSION

Oller and colleagues (Jhang & Oller, 2017, 2017; Oller et al., 2013, 2016; Oller & Griebel, 2004) posit that speech emerged from pre-linguistic vocalizations that are free of predetermined biological function, a precursor called ‘vocal functional flexibility’. One capacity foundational to vocal functional flexibility is the ability to use sounds that are not affectively-bound, a capacity we call ‘affective decoupling’. Modern human infants regularly vocalize in such a way, in supposed contrast to the relative inflexibility of vocalizations in non-human primates (e.g., Pollick & Waal, 2007). Indeed, human infants can use sounds (‘protophones’) that can be uttered into a diversity of affective circumstances on diverse occasions, such that these sounds are not tied to the experience and expression of one particular affective state (Oller et al., 2013; Oller & Griebel, 2004). By contrast primate (and more largely, 'animal') vocal behavior would be affectively bound, with particular calls being used to express particular affective state, ultimately constraining their signaling function. The view that primate vocalizations are read-outs of the affective states of the animal has otherwise long been held in the literature (Goodall, 1986; Gruber & Grandjean, 2017; Hammerschmidt & Fischer, 2008; Marler, 1980).

In the current study, we specifically looked at one of our closest living relative species, the chimpanzees. We focused on the grunt-like and whimper-like calls of young chimpanzee infants, using novel coding strategies and state-of-the-art acoustic analysis tools. We elaborated a workable coding system, which was meant to provide first insights into the affective state of infant chimpanzees, as seen in Oller et al. (Oller et al., 2013), and so as to allow for a first comparison between human and chimpanzee infants. We found that grunt-like calls are produced frequently by chimpanzee infants with both contexts we deemed positive and neutral, and less commonly also with the so-called negative affective context. Importantly, the presence of grunts in contexts of low-to-mild arousal is consistent with the hypothesis of vocal functional flexibility (Oller et al., 2019), and so is the finding that grunts occur in similar proportion with contexts we deemed positive and neutral (Oller et al., 2013).

On the other hand, whimper-like vocalizations seem to be confined to behaviors and contexts we associated with negative affective states in the infants. Their near absence with positive and neutral contexts suggests that they represent an affectively bound vocalization that has evolved to signal a narrow range of needs and one single (negative) affective valence, similar to cries in humans (Oller et al., 2013), to which they may functionally correspond (Goodall, 1986). Our results therefore suggest that grunts are not bound to one particular affective context in chimpanzees. They may also further qualify as a functionally flexible vocal unit, consistent with the observations of the circumstances of production of squeals, vocants and growls in young human infants (Oller et al., 2013). This, however, requires further examination, notably by improving our capacity to produce inferences about animals' transient affective states, and measuring whether recipients respond to these calls in a way consistent with the affect they are meant to convey.

Indeed, vocal functional flexibility requires not only affective decoupling (or the independence between particular vocalization and one affective dimension) but also evidence for consistent functionality. In human infants, the findings have been that infants use protophones with a diversity of affects, with mothers reacting consequently, showing that infants' calls are indeed fully functionally flexible (Oller et al., 2013). In these studies, the mothers' behavior could be examined, although protophones are not always socially directed (Oller & Griebel, 2004). Protocols where mothers may be asked to interact with toddlers may yield to responsiveness from the mothers whichever the affective state of the infant is (Yoo et al., 2018), which is critical in determining the function of the calls. In the course of spontaneous behavior, though, we expected little intervention from the chimpanzee mothers, except in situations where the infant is in danger. In our sample, responsiveness of the mother (tentatively defined in pilot coding as being either proactive, protective or neutral by the observer) was relatively low, a pattern which might be due to differences in mothering style between chimpanzees and humans, or a difference between our own study (where no particular demand is put on the mother) and others (where mothers may be interacting with their infant, e.g., Oller et al., 2013). This leaves us with the impossibility to conclude on whether mothers would react in ways consistent with the affective dimension of the vocal production, as seen in the human studies. Although playback of infant grunts to the mother may appear like a methodological possibility to further establish their functionality (Fischer et al., 2013; Fischer, 2016; Zuberbühler, 2014), this would require either playing the infants' calls in its own presence (which is ethically inappropriate) or playing the calls of another infant to a mother (which may not trigger any reaction at all in the non-genetically related mother). Another possibility is that the sounds we examined are not meant to be fully functional, and could be considered to be vegetative sounds. The fact that they may not appear socially directed should, however, not speak against the hypothesis that they are affectively decoupled, for the fact that a given vocal unit is independent from one particular affective valence is orthogonal with the fact that it is social directed or not. Our results are compatible with grunts being a functionally flexible call type in young chimpanzees, but do not yet demonstrate this, for the reactions of the mothers (and therefore, the function of the calls) could not be directly assessed.

Grunts (and other close calls (Oller & Griebel, 2004)) are a promising class of vocalizations to investigate the evolutionary origins of vocal functional flexibility. In a number of species (such as the vervet monkeys (Cheney & Seyfarth, 1982), western gorillas (Salmi et al., 2013), sooty mangabeys (Range & Fischer, 2004), chacma (Meise et al., 2011), Guinea (Faraut et al., 2019; Maciej et al., 2013) and olive baboons (Ey & Fischer, 2011; Silk et al., 2018)), grunts are used flexibly and can occur in a variety of contexts. So far, such evidence speaks in favor of grunts being a contextually flexible vocal unit (that is, a vocal unit whose function can be fulfilled in a diversity of contexts). Future research should try delving into the affective state animals likely experience and express when producing grunts, to confirm whether these also displays affective decoupling (i.e., the independence between grunt production and the experience of one particular affective valence) and functional flexibility (i.e., the capacity of grunts to fulfil a variety of functions on different occasions). If the term 'functional flexibility' could appear misleading, its use in the field of child development should encourage the animal communication community to employ it, such that more fruitful cross-disciplinary work can best take place.

Our second main finding was systematic acoustic differences between grunts given with so-called positive, neutral and negative behaviors, which enabled us to segregate acoustic variants of grunts into these categories. Acoustical differences linked to the affect surrounding vocal production are common in humans as in other animals (Arias et al., 2018; Aucouturier et al., 2016; Banse & Scherer, 1996; Briefer, 2012; Goupil et al., 2019; Ponsot et al., 2018; Williams & Stevens, 1972). Our data suggest that there is inter-gradation between grunt-types, with differences in acoustics relating to differences in contexts. Grunts, in other words, represent a coherent and unified call type that can manifest itself in acoustic variants in relation to the affective contexts in which they are produced. It is possible that grunts acoustically vary with arousal of the animal (as seen in other primate species (Rendall, 2003)), although positive and negative circumstances could, in principle, be equally arousing.

How exactly functionally flexible vocalizations produced by human infants transition into speech sounds has been described in previous studies (Boysson-Bardies, 2001; de Boysson-Bardies, 1993; de Boysson-Bardies & Vihman, 1991; Elbers & Ton, 1985; Nathani et al., 2006; Oller, 2000; Oller et al., 1976). Chimpanzee infants may produce grunts in ways consistent with the functional flexibility hypothesis but they of course never produce speech sounds and, historically, have failed to acquire human speech utterance even after extensive training (Hayes & Hayes, 1951). Instead, infant chimpanzee grunts may gradually develop into call variants with seemingly relatively narrow biological functions (Laporte & Zuberbühler, 2011;. Slocombe & Zuberbühler, 2010; Slocombe & Zuberbühler, 2005; Watson et al., 2015), with clear acoustical boundaries notably between grunts used to greet conspecifics ('pant-grunts'

(Laporte & Zuberbühler, 2011)) and those produced upon encountering food ('rough' or 'food grunts' (Slocombe & Zuberbühler, 2005)). It is possible that the acoustic boundaries we identified between the grunts produced across affective states (under our nomenclatures and coding system) are the foundation of acoustic diversification in adults, although the categories used here (for instance, feeding and social approach are together considered 'positive') are not consistent with the vocal differentiation seen in adults (the grunts produced in feeding vs. social approach situations are acoustically distinct in adults (Crockford, in press; Goodall, 1986)). Alternatively, those calls may simply disappear and be absent from the adult repertoire, one causal factor being the relative absence of social reinforcement (including contingent vocal responses (Ghazanfar et al., 2019)) associated with grunt production, as compared to the frequent maternal reactions to distress calls (Dezecache et al., 2020).

Our tentative to explore the affective state of the infant may be seen as preliminary, insofar as the categories we have used do not represent read-outs of physiological states. This being said, the acoustical differentiation we found speak in favor of the appropriate character of our affective distinctions. Ideally, other cues should be considered, such as the infants' facial expressions or the mothers' behavior. This approach would however face considerable challenges. We found that infant facial movements are extremely fast and fluid, which prevented us from reliable coding particularly in the wild. For this reason, the behavioral context of the infant alone (although imperfect and probably still questionable) was the most relevant available cue to approach the affective dimension of the situation. While we must again acknowledge the limitations pertaining to the fact that judgments of infants' affect were made based on the infants' behavioral contexts and done so by a human observer, the results of the acoustic analysis are providing support for the approach used to categorize affect in the present work. Future studies should investigate the affective impact of other communicative signals used by infants, such as gesture and facial behavior, and their combinations (Fröhlich et al., 2018; Fröhlich & Hobaiter, 2018).

Besides the limitations pertaining to our coding system (and its shortcomings with respect to the production of inferences regarding infants' affective states), one other limitation of this study is the small sample size, as we could only collect enough data from 6 infants. One particular difficulty with collecting data from such young chimpanzee infants is that some of their calls (notably a large part of their grunts) are very soft (a point also acknowledged by Plooij (Plooij, 1984)) and can only be heard from close, limiting the number of individuals whose mothers are unwary enough of continuous and long-lasting human observational efforts. We could not use already published data, because, to the best of our knowledge, no previous studies on the vocal behavior of wild infant chimpanzees (such as Laporte & Zuberbühler, 2011, Plooij et al., 2014 or Plooij, 1984) used a coding system amenable to inferences about the affective state of the infant.

In latest research, the comparative volubility (quantity of sounds produced in a given period of time) of human infants and other animals (Ghazanfar & Takahashi, 2014; Oller et al., 2019; Takahashi et al., 2015), and the privileged function of protophone-like vocalizations to increasingly elicit social interactions and vocal turn-taking with caregivers (Oller et al., 2019; Yoo et al., 2018). In humans, non-affectively bound vocalizations appear to occur more often than affectively bound vocalizations (such as crying) (Oller et al., 2019). They occur in solitary contexts where infants invest in practice and vocal exploration. They also occur in interactive contexts, so as to elicit and regulate social interactions with caregivers. Caregivers appear to detect the functional difference between protophones (as potentially interactive calls) and other calls (such as cries), where caregiver intervention is solicited (Yoo et al., 2018). Comparison with bonobo infants suggested much higher rate of production of non-affectively bound vocalizations and much higher vocal investment in social interactions in human infants (Oller et al., 2019). Whether human infants also are comparably more 'talkative' than their chimpanzee counterparts is a question we need to be exploring. This should be preferably investigated in captive or semi-captive settings, where true calling rate can be assessed, for video monitoring is less likely to be interrupted and for levels of ambient noise could be comparatively less problematic. Such problems have already been acknowledged by Oller and colleagues (2019) regarding previous report on the flexible development of grunting behavior in wild chimpanzees as well as their rate of occurrence (Laporte & Zuberbühler, 2011). Data from the vocal development of one captive chimpanzee indicated lower volubility than in humans (Kojima, 2003). Future studies should evaluate this fact with a larger sample.

Our study suggests that, insofar as one can delve into the affective state of infants using our coding system, chimpanzees may possess a feature that is fundamental to the development of speech in humans, the ability to produce vocalizations that are not strongly bound to the experience and expression of one particular affective valence. However, we should expect that future research will reveal further examples. For instance, coo calls in several macaque species (Hsu et al., 2005; Owren & Casale, 1994), wahoos of baboons (Maciej et al., 2013) or grunts of a number of primate species seem to be given in a variety of contexts, a precondition for affective decoupling in vocal production, itself a prerequisite for vocal functional flexibility. More largely, close calls appear to be excellent candidates (Oller & Griebel, 2004). Importantly, methodologically efforts to infer the affective states of the animals should be made in order for affective decoupling to be hypothesized.

Future research will have to address the question of how selection favored acoustic diversification of functionally flexible vocal behavior into speech in humans. The main driver for this transition, it has been argued, may have been the highly cooperative breeding system of humans, with infants regularly looked after by individuals other than the mother, which requires infants to become more active agents in forming social bonds from a much younger age than in great ape infants (Ghazanfar et al., 2019; Zuberbühler, 2012).

Cooperative breeding, in this view, may thus have transformed a functionally flexible vocal system into the uniquely human way of using vocal signals to interact socially. Another complementary reasoning is that humans' high altriciality selected for the most vocal individuals, capable of attracting caregivers (Locke, 2006). The relative contribution of both factors through mapping the phylogenetic distribution of affective decoupling and vocal functional flexibility remains to be investigated.

REFERENCES

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3–4), 227–266.

Arbib, M. A., Liebal, K., Pika, S., Corballis, M. C., Knight, C., Leavens, D. A., Maestripieri, D., Tanner, J. E., Arbib, M. A., & Liebal, K. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076.

Arias, P., Belin, P., & Aucouturier, J.-J. (2018). Auditory smiles trigger unconscious facial imitation. *Current Biology*, *28*(14), R782–R783. https://doi.org/10.1016/j.cub.2018.05.084

Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, *113*(4), 948–953. https://doi.org/10.1073/pnas.1506552113

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y., Rey, A., & Fagot, J. (2017). Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. *PLOS ONE*, *12*(1), e0169321. https://doi.org/10.1371/journal.pone.0169321

Boysson-Bardies, B. de. (2001). *How Language Comes to Children: From Birth to Two Years*. MIT Press.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Burkart, J. M., Fehr, E., Efferson, C., & Schaik, C. P. van. (2007). Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences*, *104*(50), 19762–19766. https://doi.org/10.1073/pnas.0710310104

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, *18*(5), 175–186.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27. https://doi.org/10.1145/1961189.1961199

Cheney, D. L., & Seyfarth, R. M. (1982). How vervet monkeys perceive their grunts: Field playback experiments. *Animal Behaviour*, *30*(3), 739–751.

Clay, Z., Archbold, J., & Zuberbühler, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ*, *3*, e1124. https://doi.org/10.7717/peerj.1124

Crockford, C. (in press). Why Does the Chimpanzee Vocal Repertoire Remain Poorly Understood? And What Can Be Done About It. In *The Tai Chimpanzees: 40 years of Research. Eds: Boesch C. and Wittig R.* Cambridge University Press.

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*(4), 397–421. https://doi.org/10.1163/1568539054012047

de Boysson-Bardies, B. (1993). Ontogeny of Language-Specific Syllabic Productions. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 353–363). Springer Netherlands. https://doi.org/10.1007/978-94-015-8234-6_29

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to Language: Evidence from Babbling and First Words in Four Languages. *Language*, *67*(2), 297–319. JSTOR. https://doi.org/10.2307/415108

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2020). A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. *Animal Cognition*. https://doi.org/10.1007/s10071-020-01437-5

Elbers, L., & Ton, J. (1985). Play pen monologues: The interplay of words and babbles in the first words period. *Journal of Child Language*, *12*(3), 551–565. https://doi.org/10.1017/S0305000900006644

Ey, E., & Fischer, J. (2011). Keeping in contact: Flexibility in calls of olive baboons. In *Primates of Gashaka* (pp. 413–436). Springer.

Faraut, L., Siviter, H., Dal Pesco, F., & Fischer, J. (2019). How life in a tolerant society affects the usage of grunts: Evidence from female and male Guinea baboons. *Animal Behaviour*, *153*, 83–93.

Fedurek, P., & Slocombe, K. E. (2013). The social function of food-associated calls in male chimpanzees. *American Journal of Primatology*, *75*(7), 726–739.

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*, 38226. https://doi.org/10.1038/srep38226

Fischer, J. (2016). Playback Experiments. *The International Encyclopedia of Primatology*, 1–2. https://doi.org/10.1002/9781119179313.wbprim0140

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology*, *75*(7), 643–663. https://doi.org/10.1002/ajp.22153

Fitch, W. T. (2018). The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics*, *4*(1), 255–279. https://doi.org/10.1146/annurev-linguistics-011817-045748

Fitch, W. T., Boer, B. de, Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Fröhlich, M., & Hobaiter, C. (2018). The development of gestural communication in great apes. *Behavioral Ecology and Sociobiology*, *72*(12), 194.

Fröhlich, M., Wittig, R. M., & Pika, S. (2018). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental Science*, e12716.

Ghazanfar, A. A., Liao, D. A., & Takahashi, D. Y. (2019). Volition and learning in primate vocal behaviour. *Animal Behaviour*, *151*, 239–247. https://doi.org/10.1016/j.anbehav.2019.01.021

Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: Vision, rhythm, cooperation. *Trends in Cognitive Sciences*, *18*(10), 543–553. https://doi.org/10.1016/j.tics.2014.06.004

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Harvard University Press.

Goupil, L., Johansson, P., Hall, L., & Aucouturier, J.-J. (2019). *Influence of Vocal Feedback on Emotions Provides Causal Evidence for the Self-Perception Theory*. https://doi.org/10.1101/510867

Gruber, T., & Grandjean, D. (2017). A comparative neurological approach to emotional expressions in primate vocalizations. *Neuroscience & Biobehavioral Reviews*, *73*, 182–190.

Hammerschmidt, K., & Fischer, J. (2008). Constraints in primate vocal production. In *Evolution of communicative flexibility: Complexity, creativity and adaptability in human and animal communication*. MIT press.

Hayes, K. J., & Hayes, C. (1951). The Intellectual Development of a Home-Raised Chimpanzee. *Proceedings of the American Philosophical Society*, *95*(2), 105–109.

Hrdy, S. (2007). Evolutionary context of human development: The cooperative breeding model. In *Family Relationships: An Evolutionary Perspective*. Oxford University Press.

Hsu, M. J., Chen, L.-M., & Agoramoorthy, G. (2005). The vocal repertoire of Formosan macaques, Macaca cyclopis: Acoustic structure and behavioral context. *Zoological Studies*, *44*(2), 275.

Jhang, Y., & Oller, D. K. (2017). Emergence of Functional Flexibility in Infant Vocalizations of the First 3 Months. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00300

Jürgens, U. (1976). Reinforcing concomitants of electrically elicited vocalizations. *Experimental Brain Research*, *26*(2), 203–214.

Jürgens, Uwe. (1979). Vocalization as an emotional indicator. *Behaviour*, *69*(1–2), 88–117.

Kojima, S. (2003). *A search for the origins of human speech: Auditory and vocal functions of the chimpanzee*. Kyoto University Academic Press.

Kramer, K. L. (2010). Cooperative Breeding and its Significance to the Demographic Success of Humans. *Annual Review of Anthropology*, *39*(1), 417–436. https://doi.org/10.1146/annurev.anthro.012809.105054

Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membranophone. *Scientific Reports.*, *9*, 12289. https://doi.org/10.1038/s41598-019-48760-7

Laporte, M. N. C., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. *Developmental Science*, *14*(5), 1220–1234. https://doi.org/10.1111/j.1467-7687.2011.01069.x

Laporte, M. N., & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour*, *80*(3), 467–473.

Leavens, D. A. (2009). Animal communication: Laughter is the shortest distance between two apes. *Current Biology*, *19*(13), R511–R513.

Levréro, F., & Mathevon, N. (2013). Vocal Signature in Wild Infant Chimpanzees. *American Journal of Primatology*, *75*(4), 324–332. https://doi.org/10.1002/ajp.22108

Lieberman, P. (2017). Comment on "Monkey vocal tracts are speech-ready." *Science Advances*, *3*(7), e1700442. https://doi.org/10.1126/sciadv.1700442

Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature*, *17*(2), 155–168. https://doi.org/10.1007/s12110-006-1015-x

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. *Ismir*, *270*, 1–11.

Maciej, P., Ndao, I., Hammerschmidt, K., & Fischer, J. (2013). Vocal communication in a complex multi-level society: Constrained acoustic structure and flexible call usage in Guinea baboons. *Frontiers in Zoology*, *10*(1), 58.

Mandel, M. I., & Ellis, D. P. (2005). Song-level features and support vector machines for music classification. *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, 594–599.

Marler, P. (1980). Primate vocalization: Affective or symbolic? In *Speaking of apes* (pp. 221–229). Springer.

McCune, L., Vihman, M. M., Roug-Hellichius, L., Delery, D. B., & Gogate, L. (1996). Grunt communication in human infants (Homo sapiens). *Journal of Comparative Psychology*, *110*(1), 27.

Meise, K., Keller, C., Cowlishaw, G., & Fischer, J. (2011). Sources of acoustic variation: Implications for production specificity and call categorization in chacma baboon (Papio ursinus) grunts. *The Journal of the Acoustical Society of America*, *129*(3), 1631–1641.

Mielke, A., & Zuberbühler, K. (2013). A method for automated individual, species and call type recognition in free-ranging animals. *Animal Behaviour*, *86*(2), 475–482.

Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., Meiners, T., Müller, C., Obermaier, E., & Prati, D. (2014). Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution*, *4*(18), 3514–3524.

Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics*, *20*(5), 351–369. https://doi.org/10.1080/02699200500211451

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Psychology Press. https://doi.org/10.4324/9781410602565

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Oller, D. K., & Griebel, U. (2004). Contextual freedom in human infant vocalization and the evolution of language. In *Evolution of communicative flexibility: Complexity, creativity and adaptability in human and animal communication* (p. 135). MIT Press.

Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language Origins Viewed in Spontaneous and Interactive Vocal Rates of Human and Bonobo Infants. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00729

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal Development as a Guide to Modeling the Evolution of Language. *Topics in Cognitive Science*, *8*(2), 382–392. https://doi.org/10.1111/tops.12198

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*(1), 1–11. https://doi.org/10.1017/S0305000900001276

Owren, M. J., & Casale, T. M. (1994). Variations in fundamental frequency peak position in Japanese macaque (Macaca fuscata) coo calls. *Journal of Comparative Psychology*, *108*(3), 291. https://doi.org/10.1037/0735-7036.108.3.291

Plooij, F. X. (1984). *The behavioral development of free-living chimpanzee babies and infants.* Ablex.

Plooij, F. X., Van De Rijt-plooij, H., Fischer, M., & Pusey, A. (2014). Longitudinal recordings of the vocalizations of immature Gombe chimpanzees for developmental studies. *Scientific Data*, *1*(1), 1–10.

Pollick, A. S., & Waal, F. B. M. de. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, *104*(19), 8184–8189. https://doi.org/10.1073/pnas.0702624104

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Range, F., & Fischer, J. (2004). Vocal repertoire of sooty mangabeys (Cercocebus torquatus atys) in the Taï National Park. *Ethology*, *110*(4), 301–321.

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, *113*(6), 3390–3402.

Rendall, D., & Owren, M. J. (2002). Animal vocal communication: Say what? In *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 307–313). MIT Press.

Reynolds, V. (2005). *The chimpanzees of the Budongo Forest: Ecology, behaviour, and conservation*. Oxford University Press. http://books.google.fr/books?hl=fr&lr=&id=C6hzM5lQJ6YC&oi=fnd&pg=PR11&dq=budongo+reynolds&ots=OOfJtMfycP&sig=X1c6kEzGs8ZlzhXdNH3aK3foPrA

RStudio Team. (2015). RStudio: Integrated development for R. *RStudio, Inc., Boston, MA URL Http://Www. Rstudio. Com*, *42*.

Salmi, R., Hammerschmidt, K., & Doran-Sheehy, D. M. (2013). Western gorilla vocal repertoire and contextual use of vocalizations. *Ethology*, *119*(10), 831–847.

Schaik, C. P. van, & Burkart, J. M. (2010). Mind the Gap: Cooperative Breeding and the Evolution of Our Unique Features. In *Mind the Gap* (pp. 477–496). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02725-3_22

Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, *86*(5), 955–965.

Silk, J. B., Roberts, E. R., Städele, V., & Strum, S. C. (2018). To grunt or not to grunt: Factors governing call production in female olive baboons, Papio anubis. *PloS One*, *13*(11), e0204601.

Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. *The Mind of the Chimpanzee: Ecological and Experimental Perspectives. University of Chicago Press, Chicago*, 192–207.

Slocombe, Katie E., Kaller, T., Turman, L., Townsend, S. W., Papworth, S., Squibbs, P., & Zuberbühler, K. (2010). Production of food-associated calls in wild male chimpanzees is dependent on the composition of the audience. *Behavioral Ecology and Sociobiology*, *64*(12), 1959–1966.

Slocombe, Katie E., & Newton-Fisher, N. E. (2005). Fruit sharing between wild adult chimpanzees (Pan troglodytes schweinfurthii): A socially significant event? *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *65*(4), 385–391.

Slocombe, Katie E., & Zuberbühler, K. (2005). Functionally Referential Communication in a Chimpanzee. *Current Biology*, *15*(19), 1779–1784. https://doi.org/10.1016/j.cub.2005.08.068

Tajiri, Y., Yabuwaki, R., Kitamura, T., & Abe, S. (2010). Feature Extraction Using Support Vector Machines. In K. W. Wong, B. S. U. Mendis, & A. Bouzerdoum (Eds.), *Neural Information Processing. Models and Applications* (pp. 108–115). Springer. https://doi.org/10.1007/978-3-642-17534-3_14

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, *349*(6249), 734–738. https://doi.org/10.1126/science.aab1058

Tchernichovski, O., & Oller, D. K. (2016). Vocal Development: How Marmoset Infants Express Their Feelings. *Current Biology*, *26*(10), R422–R424. https://doi.org/10.1016/j.cub.2016.03.063

Tsukahara, T. (1993). Lions eat chimpanzees: The first evidence of predation by lions on wild chimpanzees. *American Journal of Primatology*, *29*(1), 1–11.

Vert, J.-P., Tsuda, K., & Schölkopf, B. (2004). A primer on kernel methods. In *Kernel methods in computational biology* (Vol. 47, pp. 35–70). MIT Press.

Waal, F. B. M. de, & Pollick, A. S. (2011). Gesture as the most flexible modality of primate communication. *The Oxford Handbook of Language Evolution*. https://doi.org/10.1093/oxfordhb/9780199541119.013.0006

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015). Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology*, *25*(4), 495–499. https://doi.org/10.1016/j.cub.2014.12.032

Williams, C. E., & Stevens, K. N. (1972). Emotions and Speech: Some Acoustical Correlates. *The Journal of the Acoustical Society of America*, *52*(4B), 1238–1250. https://doi.org/10.1121/1.1913238

Yoo, H., Bowman, D. A., & Oller, D. K. (2018). The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01510

Zhang, Y. S., & Ghazanfar, A. A. (2016). Perinatally influenced autonomic system fluctuations drive infant vocal sequences. *Current Biology*, *26*(10), 1249–1260.

Zuberbühler, K. (2012). Cooperative breeding and the evolution of vocal flexibility. In *The Oxford Handbook of Language Evolution*. Oxford University Press.

Zuberbühler, K. (2014). Experimental field studies with non-human primates. *Current Opinion in Neurobiology*, *28*, 150–156. https://doi.org/10.1016/j.conb.2014.07.012

CONFLICTS OF INTEREST

No conflicts of interest.

FIGURES

Figure 1 Proportion of grunt-like (GR) and whimper-like (WH) vocal behaviors recorded with negative (NEG), neutral (NEU) and positive (POS) affective categories of behaviors, for each

individual separately. Number between brackets indicate the number of GR and WH calls contributed by each individual.

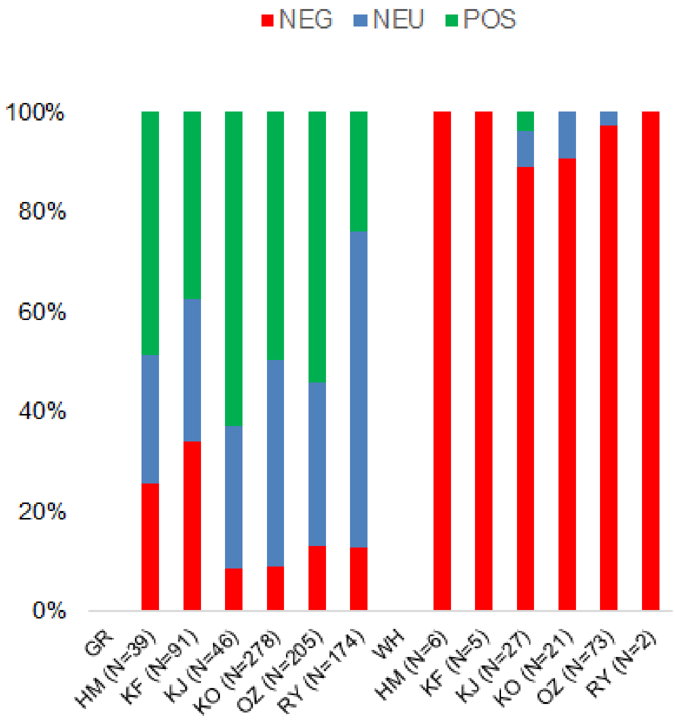
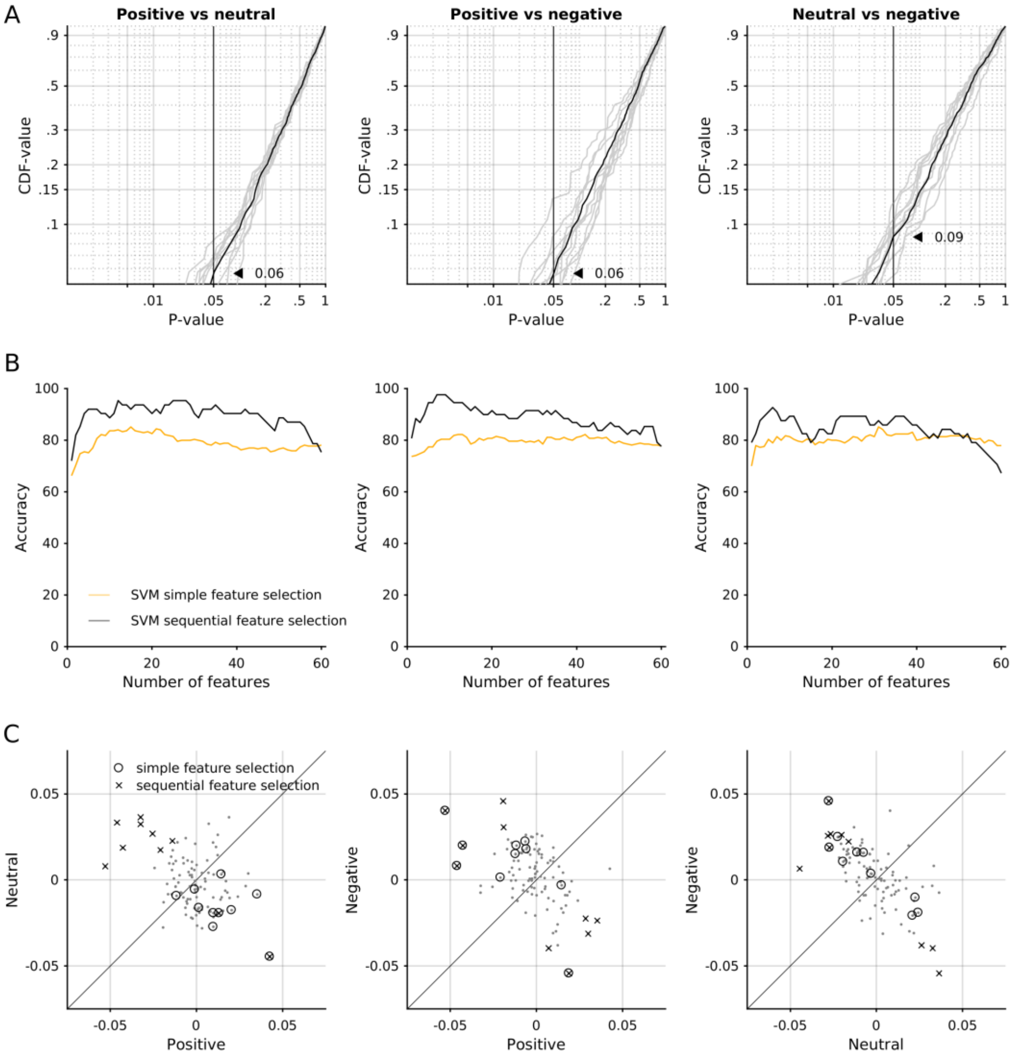


Figure 2 Feature selection and classification performances. The columns represent the comparisons of affects during which the vocal utterance occurred.

A. For each feature dimensions the discrimination power of the two classes (e.g. positive vs. neutral) was evaluated using a t-test. P-values are shown as an empirical cumulative distribution function (eCDF). Gray lines show the results of individual runs of evaluation; black lines show the means of individual runs. Indicated with arrow heads are the proportions of feature dimensions that significantly discriminate between the two classes tested.

B. The classification performances are shown for the SVM classifier relying on feature dimensions extracted through a simple feature selection (orange lines) and a sequential feature selection procedure (black lines).

C. Feature selection outcomes are shown for simple (circles) and sequential feature selection (blue x-s) as overlays on all feature dimensions (gray dots).


A
Positive vs neutral
Positive vs negative
Neutral vs negative
CDF-value
.9
.5
.3
.2
.15
.1
.01
.05
.2
.5
1
P-value
0.06
0.06
0.09
B
Accuracy
100
80
60
40
20
0
20
40
60
Number of features
SVM simple feature selection
SVM sequential feature selection
C
simple feature selection
sequential feature selection
Neutral
Negative
Negative
Positive
Positive
Neutral
0.05
0
-0.05

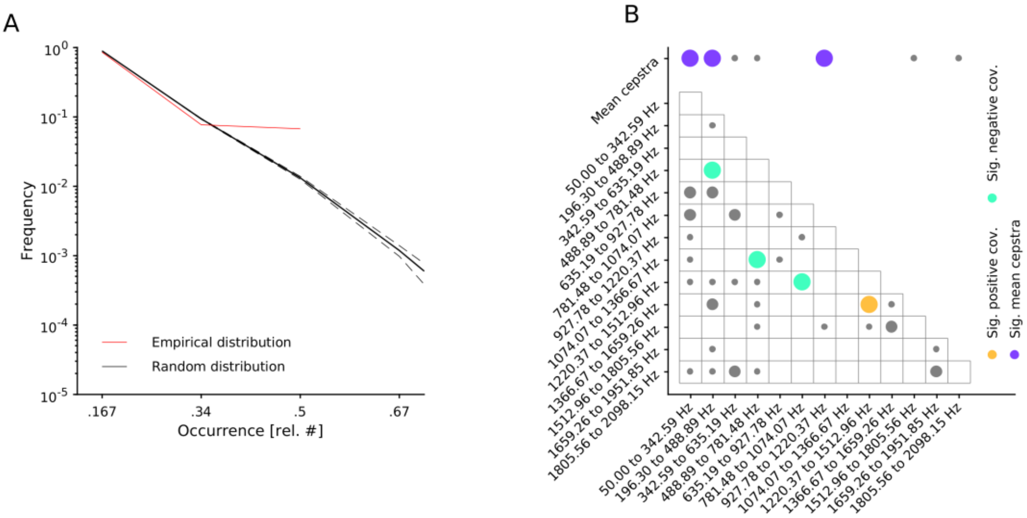
983

984

Figure 3 Overall feature importance.

A. The empirical distribution of feature dimensions across all comparisons.

B. Significant feature dimensions are shown in colors, according to their sign: in orange positive covariances, in mint negative covariance. The means of cepstra are shown in violet. The marker size indicates the occurrence: small = 1, medium-large = 2, large = 3 (significant). Gray-colored markers are non-significant feature dimensions.



TABLES

Table 1 List of focal animals, with their name (ID), sex and minimum and maximum age in months. Also given are the number of grunt-like and whimper-like vocal behaviors collected, as well as grunt-like vocalizations acoustically analyzed.

ID	*Sex*	*Min. Age (in months)*	*Max. Age (in months)*	*N whimper-like vocalizations*	*N grunt-like vocalizations*	*N of grunt-like vocalizations used in acoustical analysis*
HM	F	3.41	6.85	6	39	10
KF	M	<1	11.87	5	91	20
KJ	M	6.98	10.52	27	46	7
KO	M	3.08	8.46	21	278	67
OZ	M	1.38	8.16	73	205	32
RY	M	4.75	8.16	2	174	44

Table 2 Affective coding of infant behavior

Affect	Behavior	Description
POSITIVE	Play	Relaxed movements without obvious purpose. Can be solitary (shaking, biting and gnawing vegetation, swinging) or social (wrestling, gentle biting, gentle hitting, chasing or being chased).
POSITIVE	Grooming	Giving or receiving ‘grooming’, i.e., defined following Plooij (1984) as ‘picking through the fur of another individual’, using one’s hands or lips.
POSITIVE	Feeding	Breastfeeding or swallowing an edible element
POSITIVE	Social approach	Greeting a conspecific whilst moving (locomotion or clear leaning of the body) towards this individual
NEUTRAL	Resting	Remaining within a limited area, may involve some degree of moving around, marked by relative idleness
NEUTRAL	Moving	Locomotion not directed towards a specific individual, and not involving play
NEUTRAL	Manipulating objects	Manipulating objects (leaves, branches, rocks)
NEUTRAL	Greeting without approach	Calling upon the approach of a conspecific without showing approach (as in Social approach) or avoidance behavior towards it
NEGATIVE	Nuzzling	Unsuccessfully trying to access the mother’s nipple
NEGATIVE	Begging	Unsuccessfully attempting to access food other than breast milk
NEGATIVE	Hiding	Increased gripping or seeking contact with the mother when contact already established between them
NEGATIVE	Contact mother	Seeking contact with the mother when contact not established between them
NEGATIVE	Escaping	Showing movements meant to avoid or withdraw from a certain situation (play, grooming) or a physical position (such as moments

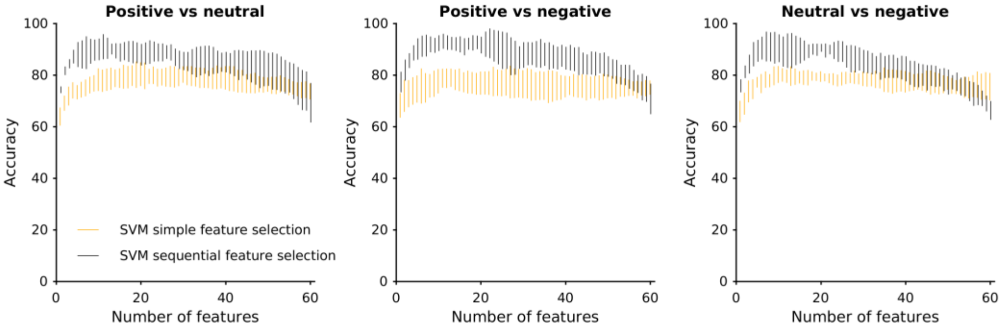
		of discomfort when the infant is suddenly pressed against the belly of the mother) the infant is in

SUPPLEMENTARY INFORMATION

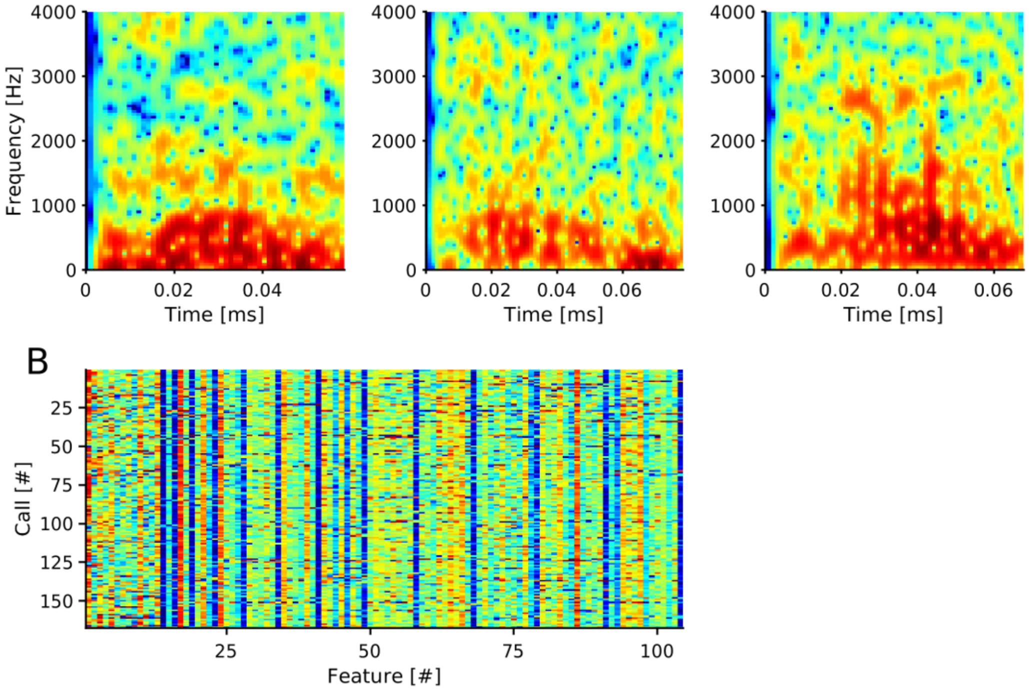
Supplementary Table 1 Number of calls per infant per affective category.

	Negative	**Neutral**	**Positive**	**Grand Total**
Grunt-like	**119**	**341**	**373**	**833**
HM	10	10	19	39
KF	31	26	34	91
KJ	4	13	29	46
KO	25	115	138	278
OZ	27	67	111	205
RY	22	110	42	174
Whimper-like	**127**	**6**	**1**	**134**
HM	6			6
KF	5			5
KJ	24	2	1	27
KO	19	2		21
OZ	71	2		73
RY	2			2
Grand Total	**246**	**347**	**374**	**967**

Supplementary Figure 1 Leave-one-out method to account for subject effects. The accuracies of the three comparisons of grunt types are shown as function of number of features. These graphs illustrate the variability of accuracy caused by leaving out one of the 6 individuals per each separate classification procedure. The vertical bars indicate the minimum and maximum scores.



Supplementary Figure 2 MFCCs extracted from example calls and extracted feature matrix. A. Time-frequency spectra of three arbitrarily chosen calls. B. From each call 26 spectral bands and 13 cepstra were extracted. Feature vectors containing the means and covariances of cepstra are shown for each call. Means are shown as features 1 to 13 on the x-axis, followed by covariances (91 values).


Frequency [Hz]
Time [ms]
Time [ms]
Time [ms]
B
Call [#]
Feature [#]