

SCIENTIFIC REPORTS



OPEN

Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally

Verena Kersken^{1,2}, Klaus Zuberbühler^{1,2,3} & Juan-Carlos Gomez¹

Received: 16 November 2015

Accepted: 15 December 2016

Published: 25 January 2017

We present empirical evidence showing that the acoustic properties of non-linguistic vocalisations produced by human infants in different cultures can be used cross-culturally by listeners to make inferences about the infant's current behaviour. We recorded natural infant vocalisations in Scotland and Uganda in five social contexts; declarative pointing, giving an object, requesting an action, protesting, and requesting food. Using a playback paradigm, we tested parents and non-parents, who either had regular or no experience with young children, from Scotland and Uganda in their ability to match infant vocalisations of both cultures to their respective production contexts. All participants performed above chance, regardless of prior experience with infants or cultural background, with only minor differences between participant groups. Results suggest that acoustic variations in non-linguistic infant vocalisations transmit broad classes of information to listeners, even in the absence of additional cues from gesture or context, and that these cues may reflect universal properties similar to the 'referential' information discovered in non-human primate vocalisations.

In human speech, prosody changes the rhythm, stress, or intonation of an utterance and thereby conveys information beyond the semantic content of utterances, for example to indicate questions or make statements¹. The prosodic features of speech also function to convey basic motivational and emotional states, for example, joy, disgust, sadness, or contempt, which can be recognised from the acoustic structure of the speech signal alone². It has been suggested that this could be a human universal, as speakers of different languages can link differences in intonations in a fictitious language with specific emotions³. Alongside speech and its associated prosodic patterns, humans produce vocal signals that have no direct linguistic content, such as grunts, cries, screams, laughter, or gasps. However, the communicative functions of these non-linguistic vocal signals, despite their ubiquity in everyday human interaction, have rarely been studied.

Some developmental studies suggest that infant non-linguistic sounds (e.g., crying) are primarily expressions of affect and emotional states⁴. A significant proportion of the vocal signals produced by human infants in their first six months of life are of this type and appear to function to express the infant's primary needs, such as hunger or physical discomfort⁵. There is evidence that the acoustic properties of some of these sound types vary systematically with the context in which they are produced. For example, new-borns display acoustically different cry patterns when in pain as compared to when hungry^{6,7}. Similarly, within the first three months, the acoustic properties of infant vocalisations emitted in a positive or negative emotional state vary systematically^{7,8}. Furthermore, parents are able to distinguish these sounds, and make inferences about the infant's emotional state on the basis of acoustic information alone⁹. Thus, parents listening to the vocalisations of infants recorded in different settings (when the infant was hungry, when their nappy needed changing, or when she was content), were able to classify the sounds they heard on a specially designed infant-state "barometer"⁹.

As the infant matures, these classes of sounds do not disappear but continue to be produced in co-existence with speech. For example, 4–8 month-olds consistently produce acoustically similar vocal patterns during toy interactions¹⁰, and 11–22-month-old infants produce acoustically consistent structures to express affect, indicate an event or object, or request help^{11,12}. Whether these acoustic differences influence a listener's behaviour is a question that received considerably less research interest. One study reported that different auditory cues that accompanied a video still frame of an infant led parents to make different judgements about the activity the infant

¹School of Psychology, University of St Andrews, St Marys Quad, St Andrews, KY16 9JP, Scotland. ²Budongo Conservation Field Station, P.O. Box 362, Masindi, Uganda. ³Cognitive Science Centre, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland. Correspondence and requests for materials should be addressed to J.C.-G. (email: jg5@st-andrews.ac.uk)

Category	Description
Giving	Infant gives object to peer or caregiver
Action request	Infant requests an action or object from a caregiver
Protest	Infant vocalises in reaction to an unpleasant event or action
Declarative Pointing	Infant points at an interesting object or event and vocalises
Food requests	Infant is in the presence of food and requests some

Table 1. Descriptive contexts in which playback stimuli were recorded.

is engaged in ref. 13. However, a recent study by Lindova *et al.* suggests that listeners can make judgements about emotional salience when listening to infant vocalisations, but fail to draw correct inferences about the production context¹⁴.

In 12-month old infants, declarative pointing to direct someone's attention to an interesting event, and imperative pointing to request an object or action, are associated with acoustically different vocalisations^{15,16}, and 9–18 month old infants produce acoustically distinct grunt variants linked to different situations, such as physical effort, attention to objects, and attention to people^{17,18}. An interesting interpretation here has been that grunts are phylogenetically related to the grunts of non-human primates, and might facilitate the acquisition of referential words in humans¹⁷. While there is a wealth of research based on infants growing up in Western cultures, we have virtually no information about the use of non-linguistic vocal sounds in cross-cultural contexts and the type of information that they can transmit.

The aim of our study was to explore whether the non-linguistic vocalisations of human infants from different cultures convey referential information about situations and events independently of the linguistic and cultural background of participants. To address this, we investigated whether adult listeners from different cultures and with different degrees of experience with young children were able to interpret context-specific vocalisations produced by 11–18 month-old infants from their own or from another culture.

We hypothesized that a number of variables could potentially influence participants' performance on this task: 1) the level of experience participants had with young children¹⁹, 2) whether the infant vocalisation was recorded in the listener's own or another culture²⁰, and 3) in what behavioural context the infant vocalisation is produced¹⁴. We explored the influence of these variables on people's abilities to match infant vocalisations to their respective production context.

Results

We employed a playback paradigm in which vocalisations from five different behavioural contexts (protesting, requesting an action, declarative pointing, giving an object, and requesting food – see Table 1) were played back to 102 listeners. To investigate the influence of culture, we tested listeners from Scotland and rural Uganda with vocalisations that we previously recorded from infants in both cultures. To investigate the role of experience, we tested parents, non-parents who regularly interacted with young infants, and non-parents who had no direct experience with infants (the latter only in Scotland – all our participants in Uganda had smaller siblings or shared a compound with families with young children).

Audio stimuli were presented to participants, and they were asked to choose from three descriptions (for example “infant wants food”, “infant points to a car” or “infant gives toy to a peer”), which they thought best fit the audio sample. Overall 40 audio stimuli from different individuals were presented, 20 of which were recorded in Scotland, and 20 in Uganda. We analysed the frequency of correct matchings between recording context of the audio sample and the description offered.

In all of the five participant groups, participants scored a higher proportion of correct matchings of vocalisations and production contexts than would be expected by chance (i.e. 33% of correct responses, see Fig. 1 and Table 2) (one sample t-test on the proportion of correct responses: Scottish parents $t[19] = 33.41$, $p = 0.0001$, Scottish experienced $t[16] = 52.59$, $p = 0.0001$, Scottish inexperienced $t[19] = 30.84$, $p = 0.0001$, Ugandan parents $t[19] = 22.67$, $p = 0.0001$, Ugandan experienced $t[20] = 27.32$, $p = 0.0001$).

We chose a linear mixed-effects model fit by maximum likelihood, following Laird and Ware²¹. We tested a null-model (random factors: intercept and participant ID, nested within this variable were the following factors: participant origin, stimulus, context, participant group) against a full model that contained all predictor variables (fixed factors: participant origin, stimulus origin, context, participant group – see Table 3) to test if these would influence the participant's ability to correctly match recordings to their production context. The full model was significantly better at predicting the participants' success rate than the null-model (LRT: $\chi^2_1 = 61.66$, $p < 0.001$). Context and participant group were significant predictors of the success rate, whereas stimulus origin was not a significant predictor (see Fig. 2 and Table 4).

Discussion

In this playback study we show that Scottish and Ugandan adult participants were able to match audio samples of infant non-linguistic vocalisations to their corresponding behavioural contexts of emission, i.e., giving, declarative pointing, requesting actions, requesting food, and protesting. This was regardless of whether the vocalisations were recorded from infants in the listeners' own or a different culture, and regardless of listeners' previous experience with young infants.

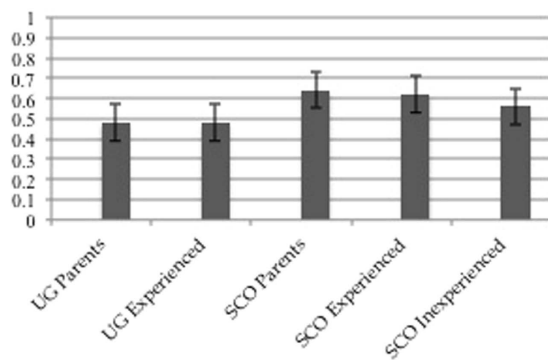


Figure 1. Proportion of correct responses in participant groups for each of the five stimuli categories.

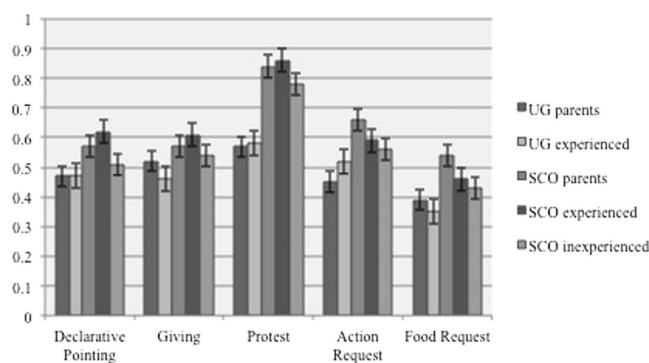


Figure 2. Proportion of correct responses for stimuli from both cultures for each participant group.

Participant group	Proportion of correct responses	Binomial Test p-values
UG parents	0.48	0.012
UG experienced non-parents	0.48	0.012
SCO parents	0.64	0.001
SCO experienced non-parents	0.62	0.001
SCO inexperienced non-parents	0.56	0.001

Table 2. Mean proportion of correct classifications for each participant group.

Random/Fixed	Effect	Levels
Random	Participant ID	
Fixed	Stimulus Origin	Uganda Scotland
Fixed	Participant Origin	Uganda Scotland
Fixed	Context	Giving, Action Request, Protest, Declarative Pointing, Food Request
Fixed	Group	Parents, Experienced, Non-experienced

Table 3. Fixed effects and their levels.

The results of this study confirm and extend existing evidence that non-linguistic infant vocalisations contain information that can be picked up by receivers^{11,13,14,16}. Importantly our study demonstrates for the first time that this information can be transmitted across cultures and is to some extent independent of the listener’s amount of experience with young infants.

An important issue is the nature of the information transmitted by the vocalizations. Although participants’ scores were significantly above chance level, classification rates were far from perfect, around 50–60% (compared to 33% expected by chance, see Table 2). This suggests that the information content of the vocalisations is broad and semantically restricted. These broad referential functions were consistent across cultures, despite some evidence of fine-tuning by individual experience and cultural background.

	b	SE b	CI 95%	p-value
Intercept	0.892	0.049	0.797 to 0.988	
Stimulus Origin	-0.026	0.016	-0.056 to 0.005	0.099
Participant Origin	-0.148	0.018	-0.183 to -0.112	<0.001
Context	-0.014	0.006	-0.025 to -0.003	0.010
Group	-0.025	0.012	-0.048 to -0.001	0.039

Table 4. Results of the GLMM testing factor that influence participants' performance on the playback task.

In everyday situations, listeners are likely to encounter vocalisations alongside additional information provided by other communicative signals, such as gestures or facial expressions^{22,23}, in addition to the situational context. These sources are likely to supplement the information contained in the vocalisations, and thereby increase the participants' ability to recognise and classify situations, but our results show that the vocalizations themselves contain enough information to infer the situations above chance.

The relative independence from culture and experience of the ability to extract broad information from the infant vocal sounds supports the idea that the vocalisations we recorded were truly non-linguistic, as an early influence of native speech has been reported in vocalisations directly related to language acquisition: babbling sequences reflecting intonation or melodic patterns, and frequently used syllables, from the native language^{20,24,25}.

The results of our study provide evidence that infants' non-linguistic vocalisations transmit referential information about social events in which the caller is involved, regardless of upbringing. These referential functions may be comparable to what has been reported for non-human primates, and raise similar issues as to the nature of the referential functions and information involved. Contrary to previous findings that cross-cultural recognition of vocalisations is only accurate in relation to negative emotions or basic positive emotions^{26,27}, the range of contexts recognised in our study is wider and richer, including information about subtle positive interactions like giving, showing, and cooperative requests of food and actions.

For many years, the default assumption, and still held by many, for primate vocalisations was that they purely reflect the caller's states of arousal.

Very few studies, however, directly measure the role that arousal plays in the production of non-human primate vocal signals. It is possible that the production of these signals is, at least to some extent, affect-based, but listeners can still make inferences about the state of the world on the basis of this information²⁸⁻³⁰. The exact role affect plays in the production and comprehension of these signals needs further investigation, but the presence of affective information is not incompatible with fulfilling referential functions³¹.

The same question applies to the non-linguistic infant vocalisations presented here. It is possible that the contexts we described provoke affective reactions in the infants, and that the listeners infer the most likely situation to have provoked each vocalization based on its affective information. However unlike alarm calls in non-human primates, it might be more difficult to match all the vocalisations in this study to distinctive emotional states. While some vocalisations, for example 'protests', might be more easily explained as being primarily affect-driven and therefore more recognizable by their affective information, this might be more difficult in categories such as 'declarative pointing' or 'giving', that would require much subtler emotional distinctions, or maybe something akin to what in prosody is known as "paralinguistic attitudes"³¹. However, beyond the unresolved problem of what types of information are conveyed in the production and perception of these vocalisations, our results show that human listeners were able to make inferences about events in the world on the basis of the vocalisations alone.

Crucially, in our study all participants were able to recognise and classify the different classes of vocalisations above chance, regardless of their own or the signaller's culture, suggesting that infant non-linguistic vocalisations are in this respect akin to those observed in non-human primates^{32,33}. As with our findings, playbacks of primate calls provoke consistent behavioural reactions in receivers, despite individual differences in call structures, suggesting that directly or indirectly these signals convey information about the situation that provoked the vocalization. There is an on-going debate about the exact nature of the information contained in primate calls and in what sense it is or not referential in their production and comprehension³⁴⁻³⁷. Our results indicate that non-linguistic human vocalisations should be included in this debate.

Methods

Participants. 102 adults volunteers took part in the study, 61 from Fife, Scotland, and 41 from the Masindi District, Uganda. The Scottish group consisted of 21 parents of infants older than two years, 20 participants with experience with children under the age of two years, and 20 with little or no experience. The Ugandan group consisted of 20 parents of infants older than two years, and 21 experienced participants (no non-experienced participants were found). The level of experience with infants was established through a self-report questionnaire, asking whether they had any children, and how old they were, or the open question of whether they had any experience with infants, and if so, to give examples of this (e.g. babysitting, job in nursery, younger relatives). If participants did not report any of these experiences, they were included in the group of "inexperienced non-parents". We can, however, not exclude that this group gained some experience with infants through the media or more irregular contact with young children. Three experienced participants from Scotland were excluded, as they did not provide enough information about their experience in the questionnaire. In Scotland, all participants reported English as their first language or as being bilingual. All participants have completed secondary education, the majority either held a University degree, or were postgraduate students. Most participants came from a European middle-class background. Parents had between 1 and 3 children. In Uganda, participants were often

multi-lingual, speaking Swahili, Alur or Acholi and all were able to read and understand written English. Formal education in Uganda is conducted in English, so the entire study was conducted in English for all participants. All Ugandan participants have completed at least primary education, some of the male participants also completed secondary education. All participants lived in rural villages in the Masindi district. In these communities the majority of people live in compounds shared with their large extended family and livestock. Some participants were professionals (teacher, shop-keeper), others were subsistence farmers who were occasionally employed. Parents had between 2 and 13 children.

Playback Stimuli. Stimuli were selected from pre-recorded vocalisations of Scottish and Ugandan infants between the ages of 11 and 18 months in five different contexts (Table 1). The contexts were chosen because they occurred frequently in the infant's everyday interactions in both cultures. Although we cannot completely rule out that some of these recordings carried traces of linguistic content, none of the calls revealed any resemblance to spoken words. Moreover, all infants were in the very early stages of speech development with a very small speech repertoire.

All audio stimuli were extracted from video recordings of natural interactions between the infants and their caregiver in a nursery environment (Scotland) or at home (Uganda). The video sequences were used to classify the stimuli according to context, type of interaction with persons and objects, and ongoing activities, using the broad categories of interactive behaviour listed in Table 1. Reliability of coding was ensured by asking two naïve coders to classify 15% of the video material from either culture into the five presented categories plus an additional 'unknown' category for cases that would not match any category. Inter-rater reliability was high (Cohen's kappa = 0.96), suggesting that the context in which the sound was recorded could be identified unambiguously.

We then randomly selected eight audio clips (between 2 and 10 seconds long) from each of the five categories from our database, four produced by Scottish infants and four by Ugandan infants. The samples were produced by different infants. Using Adobe Audition we removed any background noise that could provide clues to the infant's activity (e.g., hearing cutlery during food preparation). On some clips, the stimulus amplitude was enhanced to match other clips and to ensure that participants could hear the stimuli well. Otherwise the clips were not changed in any way.

Experimental Set-up. In the experiment, participants were presented with 40 different recordings and asked, for each one, to select one description of infant behaviour that would best fit the audio clip from three options. We chose three options to allow participants a variety of choices without making too many demands on memory (participants had to remember the vocalization they were trying to match), or introducing possible confounds due to limited attention, misreading, and differences in reading skills between Ugandan and Scottish participants. All descriptions were taken from the transcripts of the original video episode that contained the infant call sample. The distracters were chosen randomly among descriptions from two different categories than the matching description. Distracters were counter-balanced to ensure an even representation of each category accompanying the target description. Descriptions were of the type "infant sees more of a favourite food and requests some" or "infant gives an object to a friend". Descriptions removed cues to the cultural background of the infant, for example, for the food context we removed culturally specific descriptions of food such as: 'cheese', 'biscuits' for Scotland, or 'sweet potato', 'jackfruit' for Uganda; or objects (Scotland: 'soap bubbles', 'toys'; Uganda: 'bucket', 'jerry cans'), or events of interest (Scotland: pointing at a boat; Uganda: pointing at goats).

Audio stimuli and the possible answers were presented on an Apple Macbook Pro computer in Scotland, and in Uganda on an Apple iPad 2. Before starting the experiment participants received instructions on how to work the technical equipment and what the experiment required of them, that is, to choose the description that they thought best fit the sound they heard. Two practice clips presented at the start served to familiarize participants with the procedure, who then completed 40 experimental trials. In each trial, participants were first presented with the empty screen and one of the audio clips. They could replay the sound ad libitum by operating a replay icon at the bottom of the page, until they chose to be presented with the three response options. Participants were asked to confirm their choice and were then presented the next trial. Participants were unaware that the audio samples were recorded from two different cultural backgrounds. Audio clips were presented in random order to avoid effects due to presentation order. Participants' choices were recorded and whether or not these choices matched the recording context of the audio clip. Correct responses were coded when the participant's choice matched the original recording context.

Statistical Analysis. To investigate whether participants correctly matched a higher proportion of audio clips to their respective production context than expected by chance, we conducted a one-sample t-test on the proportion of correct response for each participant group.

To test whether participant's home country ('participant origin'), their experience with small children ('participant group'), the country in which the vocalisation was recorded ('stimulus origin'), or the recording context ('context') influenced participant's ability to correctly match vocalisations to their respective production contexts, we ran a linear mixed-effects model fit by maximum likelihood, following Laird and Ware²¹. The statistical analysis was conducted in R, version 3.3.0 (R Core Team, 2016), and the nlme package (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team). We tested whether the model predicted success in matching the playback stimuli to their respective production context and whether this is influenced by the fixed effects. The fixed effects, and their respective levels, are illustrated in Table 3. Our dependent variable was participant's success in matching a recording to the correct production context.

To confirm model validity, we used variance inflation factors (VIF, Fox and Weisberg 2011), which verified that collinearity was not an issue (maximum VIF = 1.48). Using a Likelihood Ratio Test (LRT), we tested our full model against a null model comprising the intercept and random effect. We conducted an Analysis of Variance

Participant Group	Context	Mean	Standard Error	Standard Deviation
Parents	Food Request	0.537	0.029	0.499
	Action Request	0.533	0.029	0.499
	Protest	0.706	0.026	0.456
	Declarative Pointing	0.550	0.028	0.498
	Giving	0.399	0.028	0.490
Experienced	Food Request	0.518	0.027	0.500
	Action Request	0.556	0.027	0.497
	Protest	0.693	0.025	0.461
	Declarative Pointing	0.556	0.027	0.497
	Giving	0.500	0.027	0.500
Inexperienced	Food Request	0.506	0.039	0.501
	Action Request	0.543	0.039	0.499
	Protest	0.781	0.032	0.414
	Declarative Pointing	0.562	0.039	0.497
	Giving	0.425	0.039	0.495

Table 5. Between-level comparisons for multi-level factors (participant group and context).

(ANOVA) on the baseline model to test how well different versions of the model describe the data and whether there are significant interactions between the fixed factors entered into the model. For the fixed factors with more than two levels (participant group and context), we conducted between-level comparisons (Table 5).

Ethics. Participation was entirely voluntary and with no financial incentives. Participants were informed about the aims of the study and what their participation would entail. All participants gave their written consent to take part. After completion, participants were debriefed about the nature of the study. The study has been performed in accordance to the rules and regulations for research with human subjects of the University of St Andrews Teaching and Research Ethics Committee, and the Ugandan National Council for Science and Technology. Both bodies approved the study.

References

- Kreiman, J. & Sidtis, D. V. L. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. (Wiley-Blackwell 2011).
- Banse, R. & Scherer, K. R. Acoustic profiles in vocal emotion expression. *J. of Personality and Social Psych.* **70**(3), 614–636 (1996).
- Scherer, K. R., Banse, R. & Wallbott, H. G. Emotion inferences from vocal expression correlate across languages and cultures. *J. of Cross-cultural Psych.* **32**(1), 76–92 (2001).
- Lester, B. M. & Boukydis, C. F. Z. *No language but a cry*. (eds H. Papoušek & U. Jürgens) *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (Cambridge University Press 1989).
- Oller, D. K. *The emergence of the speech capacity*. (Psychology Press 2000).
- Protopapas, A. & Eimas, P. D. Perceptual differences in infant cries revealed by modifications of acoustic features. *J. of the Acoustic Soc. of America* **102**(6), 3723–3734 (1997).
- Scheiner, E., Hammerschmidt, K., Jürgens, U. & Zwirner, P. Acoustic analysis of developmental changes and emotional expression in the preverbal vocalisations of infants. *J. of Voice* **16**(4), 509–529 (2002).
- Papoušek, M. *Vom ersten Schrei zum ersten Wort – Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation*. (Verlag Hans Huber 1992).
- Papoušek, H. Early ontogeny of vocal communication in parent- infant interactions. (eds H. Papoušek & U. Jürgens) *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (Cambridge University Press 1989).
- D’Odorico, L. & Franco, F. Selective production of vocalisation types in different communication contexts. *J. of Child Lang.* **18**, pp. 475–499 (1991).
- Dore, J., Franklin, M. B., Miller, R. T. & Ramer, A. L. H. Transitional phenomena in early language acquisition, *J. of Child Lang.* **3**(1), 13–28 (1976).
- Flax, J., Lahey, M., Harris, K. & Boothroyd, A. Relations between prosodic variables and communicative functions. *J. of Child Lang.* **18**(1), 3–19 (1991).
- Goldstein, M. H. & West, M. J. Consistent responses of human mothers to prelinguistic infants: the effect of prelinguistic repertoire size. *J. of Comp. Psych.* **113**(1), 52 (1999).
- Lindová, J., Špinková, M. & Nováková, L. Decoding of baby calls: can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *Plos One*, doi: 10.1371/journal.pone.0124317 (2015).
- Leroy, M., Mathiot, E. & Morgenstern, A. Pointing gestures, vocalisations and gaze: Two Case Studies. (eds J. Zlatev, M.J. Falck, C. Lundmark & M. Andren), *Studies in Language and Cognition*. (Cambridge Scholar Publishing 2009).
- Grünloh, T. & Liszkowski, U. Prelinguistic vocalizations distinguish pointing acts. *J. of Child Lang.* 1–25 (2014).
- McCune, L., Vihman, M. M., Roug-Hellichius, L., Delery, D. B. & Gogate, L. Grunt communication in human infants (Homo sapiens). *J. of Comp. Psych.* **110**(1), 27 (1996).
- Vihman, M. M. *Phonological development: The origins of language in the child*. (Blackwell 1996).
- Green, J. A. & Gustafson, G. E. Individual recognition of human infants on the basis of cries alone. *Dev. Psychobiol.* **16**(6), 485–493 (1983).
- de Boysson-Bardies, B. & Vihman, M. M. Adaptation to language: Evidence from babbling and first words in four languages. *Language* **67**(2), 297–319 (1991).
- Laird, N. M. & Ware, J. H. Random-Effects Models for Longitudinal Data, *Biometrics*. **38**, 963–974 (1982).
- Bruner, J. From Communication to language – A psychological perspective. *Cognition* **3**(3), 255–287 (1975).

23. Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. *The Emergence of Symbols – Cognition and Communication in Infancy*. (Academic Press 1979).
24. Whalen, D. H., Levitt, A. G. & Wang, Q. Intonational differences between the reduplicative babbling of French- and English-learning infants. *J. of Child Lang.* **18(03)**, 501–516 (1991).
25. Mampe, B., Friederici, A. D., Christophe, A. & Wermke, K. Newborns' cry melody is shaped by their native language. *Cur. Biol.* **19(23)**, 1994–1997 (2009).
26. Sauter, D. A., Eisner, F., Ekman, P. & Scott, S. K. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Procs. of the Nat. Acad. of Sci.* **107(6)**, 2408–2412 (2010).
27. Sauter, D. A. Are positive vocalizations perceived as communicating happiness across cultural boundaries? *Communicative & Integrative Biol.* **3(5)**, 440–442 (2010).
28. Seyfarth, R. M. *et al.* The central importance of information in studies of animal communication. *Animal Beh.* **80(1)**, 3–8 (2010).
29. Price, T. *et al.* Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Sci. Rep.* **5**, doi: 10.1038/srep13220 (2015).
30. Wheeler, B. & Fischer, J. Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology* **21**, 195–205 (2012).
31. Marler, P. Primate vocalization: affective or symbolic? *Progress in ape research* (ed Bourne, G. H.) (London: Academic Press 1977).
32. Slocombe, K., Townsend, S. W. & Zuberbuhler, K. Wild chimpanzees (*Pan troglodytes schweinfurthii*) distinguish between different scream types: evidence from a playback study. *Animal Cognition* **12(3)**, 441–449 (2009).
33. Seyfarth, R. M. & Cheney, D. L. Signallers and receivers in animal communication. *Ann. Rev. of Psych.* **54**, 145–173 (2002).
34. Fitch, W. T. & Zuberbuhler, K. Primate precursors to human language: Beyond discontinuity. (eds Altenmüller, E., Schmidt, S. & Zimmermann, E.) *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man* (Oxford University Press 2013).
35. Fedurek, P. & Slocombe, K. E. Primate vocal communication: a useful tool for understanding human speech and language evolution? *Hum. Biol.* **83**, 153–173 (2011).
36. Townsend, S. W., Owings, D. H. & Manser, M. B. Functionally Referential Communication in Mammals: The Past, Present and the Future. *Ethology* **119**, 1–11 (2013).
37. Owren, M. J. & Rendall, D. An affect-conditioning model of nonhuman primate vocal signaling. (eds Owings *et al.*) *Perspectives in Ethology* (Plenum Press New York 1997).

Acknowledgements

We are grateful to Santa Atim, and all parents and infants in Uganda and Scotland. We thank the staff of the Acorn Nursery, Kinaldy, and the WonderYears Nursery, St Andrews. We thank the Ugandan National Council for Science and Technology for permission and the Budongo Conservation Field Station for logistic support. We thank Katharina Maria Zeiner and Charlotte Brand for help with the statistical analysis. Study funded by Leverhulme Trust (F/00268/AP) and European Research Council (PRILANG 283871) Grants.

Author Contributions

All authors contributed equally to this project. J.C.G., V.K., and K.Z. jointly conceived the study. V.K. gathered the behavioural data. All authors jointly analysed the data, and wrote and edited the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kersken, V. *et al.* Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. *Sci. Rep.* **7**, 41016; doi: 10.1038/srep41016 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017