

RESEARCH ARTICLE

Acoustic models of orangutan hand-assisted alarm calls

Bart de Boer^{1,*}, Serge A. Wich^{2,3}, Madeleine E. Hardus⁴ and Adriano R. Lameira³

ABSTRACT

Orangutans produce alarm calls called kiss-squeaks, which they sometimes modify by putting a hand in front of their mouth. Through theoretical models and observational evidence, we show that using the hand when making a kiss-squeak alters the acoustics of the production in such a way that more formants per kilohertz are produced. Our theoretical models suggest that cylindrical wave propagation is created with the use of the hand and face as they act as a cylindrical extension of the lips. The use of cylindrical wave propagation in animal calls appears to be extremely rare, but is an effective way to lengthen the acoustic system; it causes the number of resonances per kilohertz to increase. This increase is associated with larger animals, and thus using the hand in kiss-squeak production may be effective in exaggerating the size of the producer. Using the hand appears to be a culturally learned behavior, and therefore orangutans may be able to associate the acoustic effect of using the hand with potentially more effective deterrence of predators.

KEY WORDS: Orangutan, Kiss-squeak, Sound production, Size exaggeration, Deception

INTRODUCTION

An interesting but rarely studied topic is whether species use certain body parts (that are not part of the vocal apparatus) or tools to alter the acoustic characteristics of their calls. Recently, a study (Hardus et al., 2009a) showed that Bornean orangutans, *Pongo pygmaeus wurmbii* (Tiedemann 1808), alter the acoustic characteristics of an alarm call, the kiss-squeak, by positioning a hand in front of their lips during call production, and in some cases holding detached leaves in their hand simultaneously. Orangutans produce hand and leaf kiss-squeaks in the same context of their unaided version of the call – moments of distress and towards disturbances on the forest ground, such as humans, sun bears and snakes, as well as to realistic and unrealistic predator models (Lameira et al., 2013a). Physically, kiss-squeaks are produced with lip vibration as a sound source, through ‘a sharp intake of air through pursed trumpet-like lips’ (Hardus et al., 2009b) and thus do not depend on the vocal folds.

Earlier work (Hardus et al., 2009a,b) that measured the spectral center of gravity indicated that orangutan hand and leaf kiss-squeaks have more energy at lower frequencies; that is, they are lower pitched than their unaided counterpart. This has been interpreted as a potential way to deceive listeners that the sound is emitted by an animal of larger size (Hardus et al., 2009a). This is in accordance with

the observation (Fitch and Hauser, 1995, 2002; Ohala, 1984) that lower frequencies are in general associated with larger body size, as it is indeed the case in orangutan kiss-squeaks (Hardus et al., 2009a; Lameira et al., 2013b). It also is congruent with the observation that modified kiss-squeaks were emitted in circumstances perceived to be more dangerous (Hardus et al., 2009a).

However, there are different methods by which the spectral center of gravity can be shifted – for example, by lengthening the acoustic filter, by altering source characteristics or by selective damping of higher frequencies. Not all of these methods are equally effective to exaggerate size: only methods that increase the number of resonances (i.e. formants) per kilohertz are truly effective at exaggerating size (Fitch and Hauser, 1995, 2002). Of the above-mentioned methods, only lengthening the acoustic filter increases the number of resonances per kilohertz. As the acoustic effect of placing a hand in front of the mouth is not directly obvious, a precise acoustic analysis is warranted in order to clarify the nature of the shift in acoustic characteristics observed during hand kiss-squeaks. Another reason to analyze hand kiss-squeaks is that, unlike any other non-human vertebrate species (Fitch and Reby, 2001; Matrosova et al., 2007), orangutans achieve this acoustic effect via external manipulation.

Direct empirical testing of the proposed deceptive function of modified kiss-squeaks is virtually impossible in the wild as it would require the undisturbed observation by humans of several hunting attempts by natural predators on orangutans, primarily Bornean clouded leopard, *Neofelis diardi*, and the Sumatran tiger, *Panthera tigris sumatrae* (Rijksen, 1978), and such circumstances have, to our knowledge, never been reported. Thus, as a first step, we needed to assess whether hand-modified kiss-squeaks have the right acoustic properties to exaggerate size. Therefore, we modeled them mathematically and computationally.

Here, we investigated, by means of two acoustic models and by analysis of kiss-squeaks recorded in the wild, what the precise acoustic effect is of putting a hand in front of the mouth when producing a kiss-squeak, and we discuss the implications for putative mechanisms of call acquisition and the size exaggeration hypothesis. The effect of leaf use was not investigated here, as too little is known about the precise configurations used, and because the theory for very thin and light obstacles (such as leaves) is very different from that for relatively thick and heavy obstacles (such as a hand). Thick obstacles reflect waves, whereas thin obstacles also transmit them, while in addition they may exhibit complex vibrational patterns of their own (comparable to those of a thin membrane).

The effect of putting a hand in front of the mouth during kiss-squeak production alters the acoustic system fundamentally. Rather than the sound being produced by a plane wave propagating in a tube or a spherical wave in a cavity, the acoustic system partly depends on the propagation of a cylindrical wave. The acoustic wave, confined between the hand and the face, propagates analogously to ripples on a pond. To the best of our knowledge, cylindrical waves have hitherto not been described for other animal

¹Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium. ²Research Centre in Evolutionary Anthropology and Palaeoecology, School of Natural Sciences & Psychology, Liverpool John Moores University, Byrom Street, Liverpool, UK. ³Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, Amsterdam 1098 XH, The Netherlands. ⁴Pongo Foundation, Papenhoeftaan 91, Oudewater 3421XN, The Netherlands.

*Author for correspondence (bart@ai.vub.ac.be)

calls. Plane waves exist in tubular structures such as the human vocal tract (e.g. Flanagan, 1965), while spherical waves exist in large volumes, such as the air sacs found in many mammalian species (de Boer, 2009).

This paper consists of a theoretical part and an observational part. In the theoretical part, two approximations (one mathematical and one computational) to the acoustic system for the hand and unaided kiss-squeak are investigated. In the observational section, kiss-squeaks recorded in the wild are analyzed and compared with the theoretical predictions. Finally, the broader consequences of the effects of using a hand in kiss-squeak production, and the potential implications for understanding its acquisition, use and potential function are discussed.

RESULTS

Theoretical predictions

In order to understand the physics of the hand kiss-squeak, two theoretical models were investigated: a one-dimensional approximation that was studied analytically in order to obtain basic theoretical insights and a two-dimensional approximation that was studied computationally for making predictions that can be compared with real recorded (hand) kiss-squeaks.

Analytical one-dimensional approximation

For a qualitative understanding of the effect of the hand while producing kiss-squeaks, a simple approximation of the acoustical filter can be insightful. The simplest possible approximation consists of a short tube of constant diameter (modeling the lip opening from the sound source at the interior of the lips to the exterior end of the lips) of length L and radius R_1 , while the hand in combination with the face can be modeled as a cylindrical flange with radius R_2 and width d (that is, the distance between the face and the hand). This configuration is illustrated in Fig. 1. The figure also illustrates the constants used in the text.

The model has resonances when the acoustic impedance at the end of the tube, $z(L)$ is equal to the impedance at the beginning of the cylindrical flange, $z(R_1)$. A derivation of this condition and more exact formulations of $z(L)$ and $z(R_1)$ are given in Appendix 1. There is no analytical solution for this condition in the general case, but plotting (the imaginary part of) $z(L)$ and $z(R_1)$ on the same graph helps to analyze the effect of holding a hand in front of the mouth

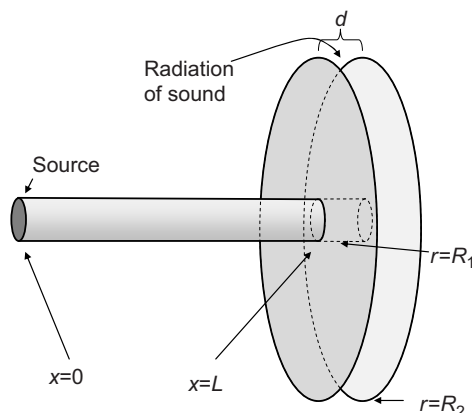


Fig. 1. Simplified model of hand kiss-squeak production, as used in the mathematical analysis. Note that sound propagates as a plane wave in the tubular section, but as a cylindrical wave in the disk-like section. d , width of the cylindrical flange; r , position along the radius of the cylinder; R_1 and R_2 , radius; L , length; x , position along the length of the tube.

when producing a kiss-squeak. This is illustrated in Fig. 2. The resonances of the complete tract are found where the line representing the tubular part [$z(L)$] and the line representing the cylindrical flange [$z(R_1)$] intersect. For reference, the line $y=0$ represents the case where the cylindrical flange is absent.

It can be observed that the presence of the cylindrical flange increases the number of intersections (i.e. resonances), resulting in a doubling when the length of the tube and the radius of the cylinder are approximately equal. The effect of a hand in front of the mouth is in fact very similar to lengthening the tube with an extra section (Flanagan, 1965; section 3.74) of about the same length as the radius of the cylindrical flange. This conclusion remains true even if the tubular section has a non-constant area function: in this case, the asymptotes of the impedance (where it goes to infinity) will be in different places, but the number of intersections with the impedance function of the cylindrical flange will remain the same, and thus the number of resonances will remain the same. As the average number of formants per kilohertz is proportional to the size of an acoustic system, it appears that holding a hand in front of the mouth is an effective way of lengthening the acoustic system, at least on theoretical grounds.

Computational two-dimensional approximation

Although the one-dimensional approximation aids in understanding the effect of putting a flat obstacle in front of an acoustic tube, a somewhat more accurate model is needed in order to make a comparison with real calls. There are three points in which the one-dimensional approximation is unrealistic. First of all, it assumes that the part of the acoustic system that corresponds to the lips is a tube with uniform diameter, while in reality it has a cone-like shape (this could in fact be incorporated into the one-dimensional model, but as the other points of inaccuracy cannot, and as it would considerably increase the mathematical complexity while not adding much to the theoretical understanding, we decided not to do this). The second inaccuracy is in the interaction between the tubular and the cylindrical sections. In the one-dimensional approximation, it is assumed that the radius of the tube (R_1) and the width of the cylindrical flange (d) are infinitesimally small. This is of course not the case in the real system. The final inaccuracy is that the one-dimensional model does not model radiation of sound. This is not so important if one wants to

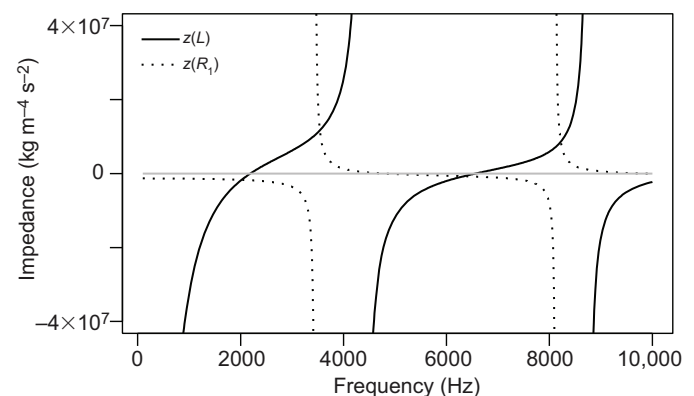


Fig. 2. Nomogram for determining resonances of the combination of a tubular section and a cylindrical flange. The imaginary part of the acoustic impedance is plotted against frequency. Parameters are: $L=4$ cm, $R_1=0.5$ cm, $d=1$ cm and $R_2=4$ cm. The intersections between the black lines and the horizontal gray line indicate the resonances of the tubular section alone. Note how the line representing the cylindrical flange [$z(R_1)$] intersects the line representing the tubular section [$z(L)$] approximately twice as often as the horizontal line.

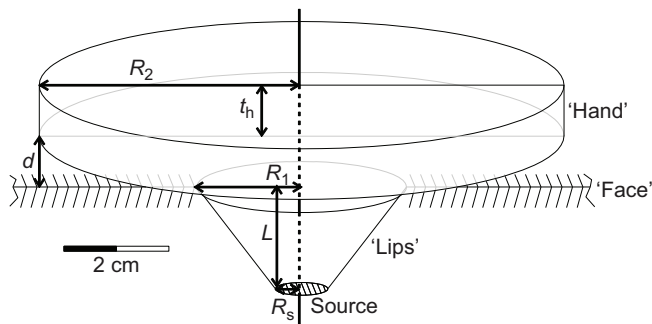


Fig. 3. Dimensions of the two-dimensional model. Note that the model is cylindrically symmetrical around the vertical axis. t_h , thickness of the hand.

have a qualitative understanding of the number and approximate location of the resonances, but it is important if one wants to compare theoretical predictions with observed sounds. The effect of radiation is to widen and lower the resonance peaks. This may result in peaks overlapping and therefore becoming indistinguishable in an observed spectrum. The two-dimensional approximation addresses all these problems.

The two-dimensional approximation is an axially symmetrical model that consists of a conical section representing the lips and a cylindrical plate (optional) representing the hand. The conical section has length L with a radius of R_s at its base (representing the sound source) and R_1 at the opening. The cylindrical plate has width (distance to the opening) d , radius R_2 and thickness t_h of 1 cm. It is assumed that the conical section is set in an infinite plane, representing the face. The system is illustrated in Fig. 3.

Radiated power for different parameter settings was calculated with a finite element approximation, and is presented in Fig. 4. It can be observed that for the system without the cylindrical plate (which

would represent the hand), there are only very broad peaks at high frequencies. The broadness of the peaks indicates that all frequencies are radiated approximately equally efficiently. For the system with the cylindrical plate that models the hand, the spectrum is much more complicated, with more, narrower peaks and most notably an extra peak at lower frequency. The increased number of peaks and the presence of peaks at lower frequencies corroborates the one-dimensional analysis. The increased sharpness of the peaks indicates that some frequencies are radiated more efficiently than others. Variations of parameters influence the position and broadness of the peaks somewhat, but the main distinction between the system with and without the cylindrical plate remains.

Observational work

The theoretical predictions were compared with a set of kiss-squeaks recorded in the wild. For the detailed analysis, we used data from the individual for which most kiss-squeaks were recorded. We only averaged over one individual because we expect the peaks for different individuals to occur at different frequencies (because of body size, among other factors). When averaging over multiple individuals, it is therefore likely that peaks of one individual's spectra would overlap with dips in another individual's spectra, causing finer spectral detail to be obscured. Data from two more individuals can be found in Appendix 2, and are used in the principal components analysis presented below. For these recordings, it was first tested whether duration and power differed between the hand and unaided kiss-squeaks. As there is no reason to assume that these quantities are normally distributed, medians and quartiles are reported and the Wilcoxon rank sum test was used. The median duration for hand kiss-squeaks was 0.30 s, with a first quartile of 0.20 s and third quartile of 0.34 s. For unaided kiss-squeaks, these values were 0.31 s (0.22 and 0.39 s). The difference

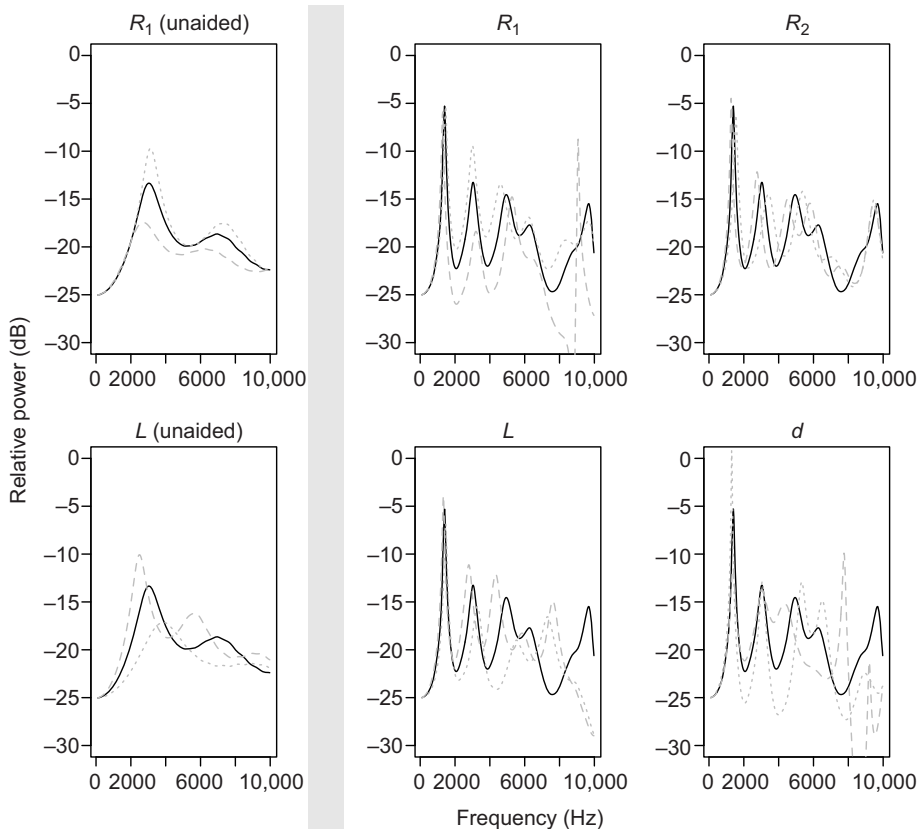


Fig. 4. Comparison of radiated power for different variations on the two-dimensional kiss-squeak model. The left column shows curves for the unaided system, the middle and right columns show curves for the system with hand. Black lines indicate the default parameters, dotted gray lines indicate variation to a lower value and dashed gray lines indicate variation to a higher value. The parameters that were varied are printed above the graphs. Values were $R_1=1.5, 2, 3$ cm; $R_2=6, 7, 8$ cm; $L=2, 3, 4$ cm; $d=0.5, 1, 2$ cm.

was not significant according to Wilcoxon's rank sum test ($P=0.40$, $W=103.5$). This is in line with earlier findings (Hardus et al., 2009a).

For power, the median for hand kiss-squeaks with was 0.13 mW m^{-2} (0.12 and 0.16 mW m^{-2}). The median for unaided kiss-squeaks was 0.15 mW m^{-2} (0.14 and 0.18 mW m^{-2}) (it should be noted that these values can only be compared within the recording, as we have no accurate reference signal to gauge the sound pressure level). This difference was not significant according to Wilcoxon's rank sum test ($P=0.052$, $W=74.5$) and this is again in line with earlier findings (Hardus et al., 2009a). In any case, the effect size is very small and is therefore not expected to influence the spectral analysis. It should be noted that the comparison of power between field recordings can potentially be problematic, as power depends of course on the distance between the sound source and the microphone. Although the distance between the microphone and orangutan probably changed slightly during the recording, the potential influence of this on the analyses was negated because both hand and unaided kiss-squeaks were produced throughout the recording, so there were no systematic differences in recording distance between the two variants.

Average frequency spectra (over all samples per category) for hand ($N=14$) and unaided ($N=18$) kiss-squeaks are shown in Fig. 5. Fig. 5A shows the unmodified spectra, while Fig. 5B shows the same average spectra with a trend line of 6 dB/octave removed. This

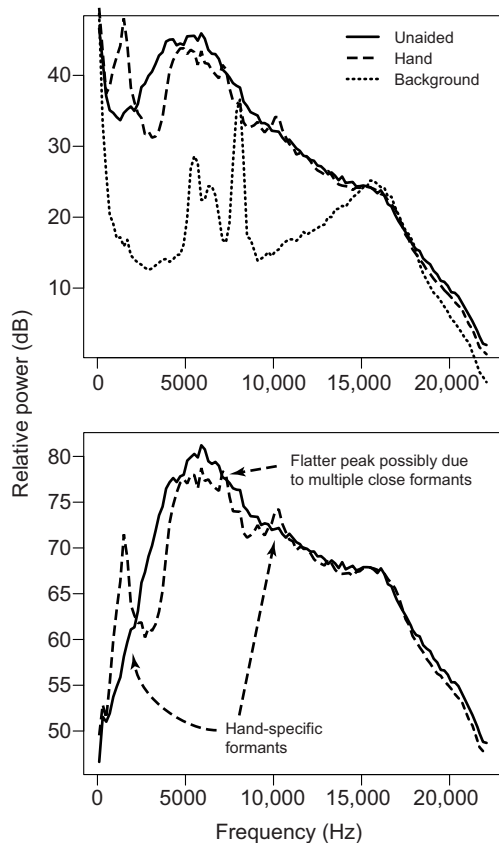


Fig. 5. Average spectra of unaided kiss-squeaks, hand kiss-squeaks and background noise. These were calculated as the average over all the long-term average spectra of the kiss-squeaks. (A) The raw signal. (B) The signal with a 6 dB/octave linear trend added. This was done in order to remove the effect of the acoustic characteristics of the source of acoustic energy (lip vibration). This usually has power that falls with 6 dB/octave. As can be seen, the peaks at low frequencies and around 16,000 Hz are part of the background noise, and may even be artifacts of the recording setup.

line is the average slope of the spectra, and is due to the fact that the source (lip vibration) generates more energy at low frequencies than at high frequencies. Removing this trend thus gives a better impression of the filter characteristics. The unaided kiss-squeak tended to have a single peak that occurred at an average frequency of 5820 Hz (with a s.d. of 1075 Hz). Sometimes this peak would be split in two with a peak around 4000 Hz and a second peak around 6500 Hz. The rest of the spectrum was relatively flat.

The hand kiss-squeaks invariably showed an important peak around 1470 ± 170 Hz (mean \pm s.d.). Further peaks occurred around 5160 ± 520 , 7040 ± 810 and $10,510 \pm 750$ Hz. Not all peaks would be equally clear in all recordings, and the peaks at 5160 and 7040 Hz are merged in the average spectrum in Fig. 5. The mean (\pm s.d.) number of peaks below 10 kHz for unaided kiss squeaks was 1.3 ± 0.36 , and for hand kiss-squeaks it was 3.9 ± 0.84 ($P < 0.01$, $t_{24,3} = 11.4$). For comparison, the spectra of two individual kiss-squeaks are given in Fig. 6 to illustrate the sharper peaks of the hand kiss-squeak more clearly. For this figure, the hand and unaided kiss-squeaks that had the highest calculated power were selected.

The measurements compare well qualitatively with the spectra that were predicted theoretically on the basis of the two-dimensional model (Fig. 4 should be compared with Fig. 5B, as this best represents the filter without the source). It should be noted that no attempt was made to fine-tune the dimensions of the model to get precise correspondence with the measured values. The dimensions of the model were estimated from what would be realistic for a real orangutan. The extra low-frequency peak caused by the hand is the most prominent correspondence between theory and observation, while the sharper peak at around 10,000 Hz is also predicted by the model. It should be noted that in the theoretical predictions, neither characteristics of the sound source (which would add their own characteristic peaks to the spectrum) nor effects of background noise (which would tend to obscure peaks) were taken into account.

A more quantitative comparison can be obtained by doing a principal components analysis of the measured spectra and checking whether the observed spectra cluster together with the theoretically predicted spectra (shown in Fig. 4). This is illustrated in Fig. 7. It can be observed that the hand kiss squeaks form one cluster, and the unaided kiss squeaks form another, except for one data point. In addition, signals produced by the same individual tend to cluster together.

The probability of the permutation of theoretical predictions (nine out of nine theoretical hand kiss-squeak spectra cluster with the observed hand kiss-squeaks and five out of five theoretical unaided kiss-squeaks cluster with the observed unaided kiss squeaks) occurring by chance if we assign them randomly to a cluster is $p = 5! \times 9! / 14! \approx 5 \times 10^{-4}$. This confirms that the agreement between theory and observation is real.

It should be stressed that the difference between the two variants of kiss-squeaks cannot be explained by selective damping due to the hand, as the low-frequency peak of the hand kiss-squeak exceeds the energy of the unaided kiss-squeak at this frequency. The difference between the spectra must therefore be mostly due to the selective amplification and attenuation of frequencies as a result of the different resonances of the two acoustic systems. This is corroborated by the qualitatively good correspondence between the theoretically predicted spectra and the observed spectra, as the theory did not take damping into account at all.

DISCUSSION

Kiss-squeaks that are produced with a hand have more and somewhat better-defined resonances than kiss-squeaks that are

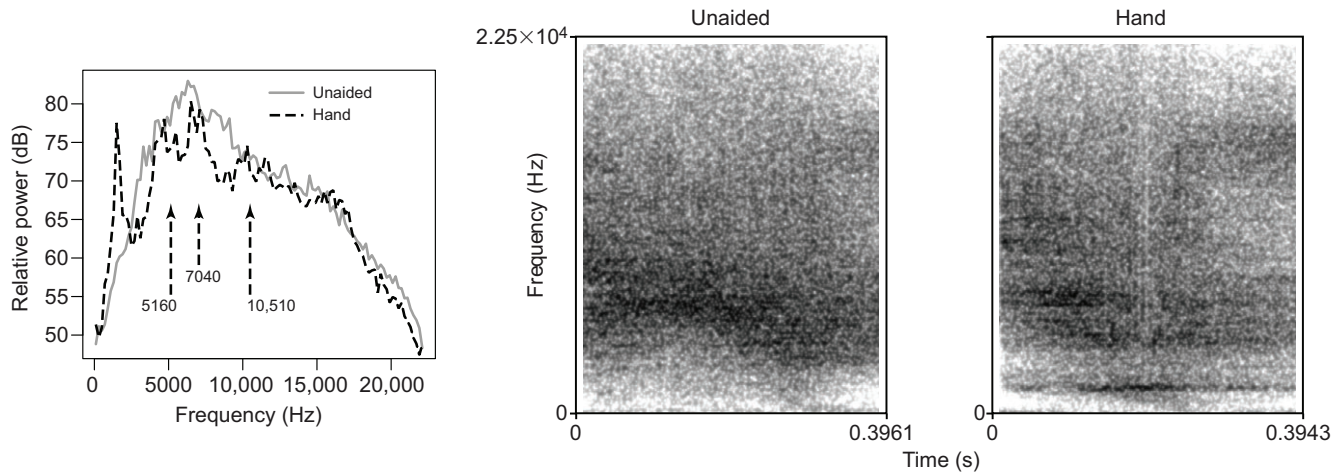


Fig. 6. Illustration of long-term average spectra (with 6 dB/octave added) and spectrograms of two individual kiss-squeaks: one unaided and one with hand. These were the loudest kiss-squeaks in our sample (and therefore the ones with the best signal to noise ratio). Note, there were more and sharper peaks for the hand kiss-squeak. The three arrows indicate the position of the average frequencies of resonances as reported in the Results on observational work. Their positions should be compared with those of the peaks in the right panels in Fig. 4.

produced without a hand. Moreover, hand kiss-squeaks have a clear resonance at a much lower frequency than unaided kiss-squeaks. This cannot be explained by selective damping, but it is predicted by the one-dimensional theory that considers the acoustic system as consisting of a tubular section followed by a cylindrical flange. The cylindrical flange is predicted to have a similar effect to lengthening the acoustic system (and therefore to create more and lower resonances), and this prediction is borne out by the observations. A more accurate two-dimensional approximation of the acoustic system predicts, moreover, that the hand kiss-squeak will have sharper resonances than the unaided kiss-squeak (because acoustic energy will be amplified more selectively in the system with hand). Again, this is borne out by the observations. It can therefore be concluded that the hand serves as an extension of the lips in which a cylindrical acoustic wave propagates. To the best of our knowledge, acoustic systems that make use of cylindrical waves have not been described before in the experimental biology literature. Other

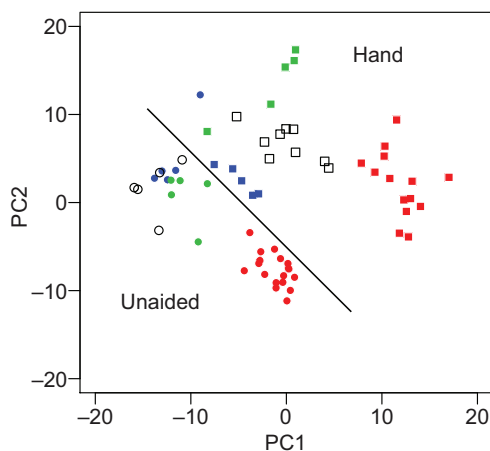


Fig. 7. Scatter plot using the first and second principal components (PC) of the measured long-term average spectra. Squares indicate hand kiss-squeaks, and circles indicate unaided kiss-squeaks. Filled symbols are measurements and open symbols are theoretically predicted spectra. Different individuals have differently colored points (red, unknown male; blue, Sultan; green, Rambo). The line illustrates the near-linear separability of hand and unaided kiss-squeaks.

factors appear to play much less of a role in producing the two kinds of kiss-squeak. It should be noted that the data that were used to check the models were based on noisy field data of one individual. However, even with these limited data, a good qualitative correspondence between theory and observation was found.

The difference between hand and unaided kiss-squeaks is congruent with the hypothesis that the use of the hand serves to exaggerate size as suggested in earlier work (Hardus et al., 2009a). Even though establishing directly whether the hand is indeed used for size exaggeration may be very hard, if not impossible (Hardus et al., 2009a), our results do support this hypothesis. It is therefore worthwhile to invest more effort in the precise behavioral correlates of the use of the hand in kiss-squeak production – for example, into whether hand kiss-squeaks are more effective in deterring predators or other animals (wild boar or deer or goat species) than ordinary kiss-squeaks whenever the occasion arises.

Another interesting observation is that kiss-squeak modified variants appear to be socially learned and represent local traditions (Krützen et al., 2011; van Schaik et al., 2003, 2006). The presence of hand and leaf kiss-squeaks across orangutan populations is inadequately explained by genetic or ecological differentiation between the populations (Krützen et al., 2011). Moreover, the way orangutans use modified kiss-squeaks and the potential function of these calls can differ between populations of the same sub-species sharing partly overlapping habitat types (Lameira et al., 2013b). If kiss-squeaks are indeed more effective in deterring predators, we could learn something about orangutan cognition by investigating the details of the learning process. After all, we can expect that both orangutans and potential predators, like humans and tigers, can observe the relation between formant dispersion and perceived size, as this is something many mammals appear to be able to do (Fitch and Hauser, 1995, 2002). Use of this ability to modify calls could then be an indication of sophisticated cognition in the call domain. Similar examples of call manipulation and tool-assisted acoustic effects are possibly limited to humans, well exemplified, for instance, by the use of musical instruments.

We suspect that, given that cylindrical waves have the same effect as lengthening an acoustic tube, there will be other examples in the animal world that use obstacles to create cylindrical waves, similar to the way certain cricket species use excavated holes and hollows to amplify the low frequencies of their songs (Bailey et al., 2001;

Bennet-Clark, 1987). The technique of singing facing a wall or another hard obstacle (W. T. Fitch, personal communication) may also rely on the effect of cylindrical waves to lengthen the acoustic system. The technique of ‘corner loading’ (playing into the corner of a room) allegedly used by legendary blues player Robert Johnson (Obrecht, 1990) and observed in male orangutan long calling in captivity (A.R.L., personal observation) may also partly depend on the effect of a cylindrical wave.

Another message from this work is the importance of the interaction between research into the physics of sound production of animal signals and of the function of these signals. When working with signals that are recorded in field circumstances, there is often a lot of noise and variation in the recorded signals. This makes it difficult to apply the sensitive analysis techniques that exist for human speech, which have been developed to deal with the relatively constant properties of human speech and the high uniformity of recordings made in often highly idealized circumstances. In addition, animal sounds may be produced in ways that are acoustically fundamentally different from human speech production (based on small-diameter acoustic tubes) – for example, by using large volumes such as air sacs, or (as in our study) by using obstacles, resulting in cylindrical wave propagation. This may make analysis techniques based on human speech less applicable.

There is therefore a potential pitfall for bioacoustics research to focus on properties of the signal that are relatively well behaved in the available recordings and that are easy to analyze with tools developed for human speech; these properties may have relatively little relevance for assessing the ecological function of the signals. In the case of kiss-squeaks, the spectral center of gravity was a relatively reliable feature, and this was measured in previous work (Lameira and Wich, 2008), but in reality it has little relation to the potential ecological function of kiss-squeaks in size exaggeration. This is because the spectral center of gravity is influenced not only by size but also by source characteristics and selective damping. We would therefore like to stress, and hope to have illustrated in this paper, that understanding the basic physics of the production of signals greatly helps in accurately measuring their acoustic properties and in establishing their function.

MATERIALS AND METHODS

The finite element model

In order to calculate the acoustic properties of the more realistic model, a finite element approximation was used, based on the GMSH and GETDP software packages (Dular et al., 1998). The model was implemented as a two-dimensional axisymmetric mesh with the distance between mesh points ranging from 0.1 cm at the lips and the cylindrical plate to 0.5 cm at the radiation boundary (resulting in approximately 13,000 mesh elements for each mesh). In order to model radiation, a spherical Bayliss–Turkel radiation boundary (Bayliss and Turkel, 1980; Givoli, 1991) was implemented at 20 cm from the lip opening. In order to model the sound source, the base of the cone was modeled as a Neumann boundary with value one (modeling a constant finite value for acoustic volume flow, which is probably a reasonable approximation for the acoustic source involved in the production of the kiss-squeak). The finite element software was then used to calculate the pressure field on the mesh. It was then calculated how much power was radiated for frequencies between 100 and 10,000 Hz.

Data collection and analysis

For comparing the theory with real orangutan calls, an existing recording of kiss-squeaks was used. The recording was made of adult male orangutans (*P. pygmaeus wurmbii*) at the Tuanan Orangutan Research station in Kalimantan (2°09'S; 114°26'E), Indonesia, at a distance of approximately 20 m (M.E.H., unpublished data). The recorded orangutans produced the kiss-squeaks in response to the presence of the human observer. The

recording was made with a sampling frequency of 44,100 Hz, a 16-bit resolution using a Marantz Analog Recorder PMD222 in combination with a Sennheiser Microphone ME 64. For the present analyses, we used a recording containing a total of 37 kiss-squeaks, five of which could not be used because of excessive background noise. Kiss-squeaks were identified via direct observation as being made with a hand or without a hand.

It should be noted that the recordings were made under field conditions and that there is therefore considerable background noise, especially in bands around 5500, 6500 and 8000 Hz, most likely due to cicadas. In addition, there is considerable noise at low frequencies (<300 Hz), probably due to movement of the observer or the animal, while there is also a background noise peak at 16 kHz, which is possibly due to the recording setup. Average background noise power was approximately $5 \mu\text{W m}^{-2}$, which is less than 5% of the power of the analyzed kiss-squeaks. The spectrum of background noise is shown as the dotted line in Fig. 5A.

The recordings were analyzed using the PRAAT software package version 5.3.49 (Boersma and Weenink, 2013) as follows: first the kiss-squeaks were isolated by selecting the part of the signal where the kiss-squeak was clearly visible in the spectrogram (i.e. where the signal strength rose above the noise level). For each kiss-squeak, the long-term average spectrum over the whole kiss-squeak (the range of durations is reported in the Results) was calculated using a bin size of 200 Hz. From this long-term average spectrum, the trend line (that is, the straight line with the least squared error) between 0 and 22,050 Hz (the Nyquist frequency) was subtracted. In the resulting spectrum, the prominent peaks (more than 6 dB higher than the neighboring peaks) were selected. As the bin size of the spectrum was 200 Hz, the maximum accuracy of the position of the peaks was therefore 200 Hz. For the principal components analysis, a maximal frequency of 10,000 Hz and a bin size of 25 Hz were used in order to ensure the same resolution and coverage as the theoretically calculated spectra. Power was measured using PRAAT's ‘power in air’ function. This calculates the power of the signal assuming that an amplitude of 1 in the recording corresponds to a sound pressure level of 1 Pa. Although this may not correspond to the precise sound pressure in reality, we only used it to compare kiss-squeaks in the same recording, so only the relative values are important.

APPENDIX 1

Derivation of the acoustic theory

The general equation for sound propagation (Morse and Ingard, 1968) is as follows:

$$c^2 \nabla^2 P - \frac{\partial^2 P}{\partial t^2} = 0, \quad (\text{A1})$$

where t is time, P is the differential pressure (that is, the difference between the pressure in the wave and the ambient atmospheric pressure) and c is the speed of sound.

In the case of harmonic waves, the pressure can be expressed as the product of a place-dependent term and a time-dependent term:

$$P(\mathbf{x}, t) = p(\mathbf{x})e^{i\omega t}, \quad (\text{A2})$$

where \mathbf{x} is the three-dimensional position and ω is the angular frequency (2π times the frequency). Note that uppercase P refers to the time-dependent pressure and lowercase p refers to the spatial component. In this case, Eqn A1 simplifies to the Helmholtz equation (e.g. Martin, 2004):

$$\nabla^2 p + k^2 p = 0, \quad (\text{A3})$$

where $k = \omega/c$ is the wave number.

If the diameter of the tube and the distance between the planes representing the orangutan's face and hand are small compared with the wavelengths of the sound, one-dimensional approximations of the acoustic wave can be used. In the tube, this means that only pressure differences along the length of the tube need to be taken

into account, while in the flange, only pressure differences along the radius of the cylinder need to be taken into account. Thus, Eqn A3 becomes:

$$\frac{\partial^2 p}{\partial x^2} + k^2 p = 0 \quad (\text{A4})$$

in the tube (where x is the position along the length of the tube) and:

$$\frac{\partial^2 p}{\partial r^2} + \frac{1}{r} \frac{\partial p}{\partial r} + k^2 p = 0 \quad (\text{A5})$$

in the cylindrical flange (where r is the position along the radius of the cylinder).

Boundary conditions are necessary in order to solve these equations. It is a reasonable approximation to assume that the tube is closed at the beginning (as there is probably only a very small opening that generates the acoustic energy of the kiss-squeak), while at the edge of the cylinder, it can be assumed to be ideally open. This leads to the following boundary conditions $\partial p/\partial x=0$ at $x=0$ (the sound source) and $p(R_2)=0$ at the exterior end of the cylinder (R_1 , R_2 and d are defined in Fig. 1).

This leads to the solution:

$$p(x) = C_1 \cos kx \quad (\text{A6})$$

in the tubular section and

$$p(r) = C_2 \left(J_0(kr) - \frac{J_0(kR_2)}{Y_0(kR_2)} Y_0(kr) \right) \quad (\text{A7})$$

in the cylindrical flange, where J_0 is a Bessel function of the first kind and Y_0 a Bessel function of the second kind. C_1 and C_2 are arbitrary constants that determine the amplitude of the sound.

At the connection between the tube and cylindrical flange, pressure and air flow are continuous, so that the boundary conditions are:

$$\begin{aligned} p(L) &= p(R_1) \\ U(L) &= U(R_1), \end{aligned} \quad (\text{A8})$$

where

$$U(x, t) = -\frac{A(x)}{\rho} \int \frac{\partial P}{\partial x} dt \quad (\text{A9})$$

is the flow along the x -direction, with $A(x)$ the area perpendicular to the x -direction, and ρ is the density of air. An entirely analogous expression holds for the r -direction. Because P is harmonic in time (Eqn A2), U is harmonic in time as well:

$$U(x, t) = u(x)e^{i\omega t}, \quad (\text{A10})$$

where $u(x)$ is the place-dependent component of volume velocity.

Using Eqns A9 and A10 for u and Eqns A2 and A6 for p , one obtains:

$$u(x) = iC_1 \frac{A(x)}{\rho c} \sin kx \quad (\text{A11})$$

for the tubular section, where $A(x)=\pi R_1^2$.

Using Eqn A7, one obtains:

$$u(r) = iC_2 k \frac{A(r)}{\rho c} \left(J_1(kr) - \frac{J_0(kR_2)}{Y_0(kR_2)} Y_1(kr) \right) \quad (\text{A12})$$

for the cylindrical flange, where $A(r)=2\pi r \cdot d$ and d is the width of the cylindrical flange.

The ratio between p and u is the acoustic impedance z , which is independent of the constants C_1 and C_2 . The boundary conditions (Eqn A8) are met if $z(L)=z(R_1)$.

APPENDIX 2

Peak numbers of all individuals

Table A1 gives the mean number of peaks below 10 kHz for all three individuals. Overall means are also given. An analysis of variance on these data showed significant effects of call type ($F_{1,46}=114$, $P<10^{-6}$) and individual ($F_{2,46}=6.24$, $P=0.004$), but no significant interaction ($F_{2,46}=1.9$, $P=0.161$).

Table A1. Mean peak number

| | With hand | Unaided |
|--------------|----------------|----------------|
| Unknown male | 3.9±0.36, N=14 | 1.3±0.84, N=18 |
| Rambo | 4.2±0.84, N=5 | 2.6±1.3, N=5 |
| Sultan | 3.2±0.45, N=5 | 1.4±0.55, N=5 |
| Total | 3.8±0.59, N=24 | 1.6±1.0, N=28 |

Data are means±s.d. and the number of data points (N).

Acknowledgements

We thank RISTEK (Ministry of State for Research and Technology, Indonesia), PHKA (Direktorat Jenderal Perlindungan Hutan dan Konservasi Alam), TNGL (Taman Nasional Gunung Leuser) and BPKEL (Badan Pengelola Kawasan Ekosistem Leuser) for research permission in Indonesia, and UNAS (Universitas Nasional) and SOCP (Sumatran Orangutan Conservation Programme) for supporting the project. We thank David Weenink for assistance with the arcana of the PRAAT user interface.

Competing interests

The authors declare no competing or financial interests.

Author contributions

B.d.B. carried out the mathematical and data analysis, A.R.L. and M.E.H. collected the data and S.A.W. coordinated the field work. All authors contributed to writing the manuscript.

Funding

B.d.B. was funded by the European research council starting grant ABACUS project and the Innoviris 'Brains back to Brussels' programme. S.A.W. was funded by the Netherlands Organisation for Scientific Research NWO. A.R.L. was funded by the Menken Funds (University of Amsterdam). Data collection in the wild by ARL was financially supported by: Fundação para a Ciência e Tecnologia (SFRH/BD/44437/2008); Wenner-Gren Foundation for Anthropological Research; Dr J. L. Dobberke voor Vergelijkende Psychologie; Lucie Burgers Foundation for Comparative Behaviour Research, Arnhem, The Netherlands; Schure-Beijerinck-Popping Fonds; Ruggles-Gates Fund for Anthropological Scholarship of the Royal Anthropological Institute of Great Britain and Ireland; Netherlands Organisation for Scientific Research (NWO); Primate Conservation, Inc.; and Fundação Calouste Gulbenkian.

References

- Bailey, W. J., Bennet-Clark, H. C. and Fletcher, N. H. (2001). Acoustics of a small Australian burrowing cricket: the control of low-frequency pure-tone songs. *J. Exp. Biol.* **204**, 2827-2841.
- Bayliss, A. and Turkel, E. (1980). Radiation boundary conditions for wave-like equations. *Commun. Pure Appl. Math.* **33**, 707-725.
- Bennet-Clark, H. C. (1987). The tuned singing burrow of mole crickets. *J. Exp. Biol.* **128**, 383-409.
- Boersma, P. and Weenink, D. (2013). *PRAAT: Doing Phonetics by Computer*. Amsterdam: Universiteit van Amsterdam.
- de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *J. Acoust. Soc. Am.* **126**, 3329-3343.
- Dular, P., Geuzaine, C., Henrotte, F. and Legros, W. (1998). A general environment for the treatment of discrete problems and its application to the finite element method. *IEEE Trans. Magn.* **34**, 3395-3398.
- Fitch, W. T. and Hauser, M. D. (1995). Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.* **37**, 191-219.
- Fitch, W. T. and Hauser, M. D. (2002). Unpacking "honesty": vertebrate vocal production and the evolution of acoustic signals. In *Acoustic Communication* (ed. A. M. Simmons, R. R. Fay and A. N. Popper), pp. 65-137. New York: Springer.

- Fitch, W. T. and Reby, D.** (2001). The descended larynx is not uniquely human. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 1669-1675.
- Flanagan, J. L.** (1965). *Speech Analysis, Synthesis and Perception*. Berlin: Springer.
- Givoli, D.** (1991). Non-reflecting boundary conditions. *J. Comput. Phys.* **94**, 1-29.
- Hardus, M. E., Lameira, A. R., van Schaik, C. P. and Wich, S. A.** (2009a). Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? *Proc. R. Soc. Lond. B Biol. Sci.* **276**, 3689-3694.
- Hardus, M. E., Lameira, A. R., van Schaik, C. P. and Wich, S. A.** (2009b). A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In *Orangutans* (ed. S. A. Wich, T. Mitra Setia, S. S. Utami and C. P. van Schaik), pp. 49-60. New York: Oxford University Press.
- Krützen, M., Willems, E. P. and van Schaik, C. P.** (2011). Culture and Geographic Variation in Orangutan Behavior. *Curr. Biol.* **21**, 1808-1812.
- Lameira, A. R. and Wich, S. A.** (2008). Orangutan long call degradation and individuality over distance: a playback approach. *Int. J. Primatol.* **29**, 615-625.
- Lameira, A. R., de Vries, H., Hardus, M. E., Hall, C. P. A., Mitra-Setia, T., Spruijt, B. M., Kershenbaum, A., Sterck, E. H. M., van Noordwijk, M., van Schaik, C. P. et al.** (2013a). Predator guild does not influence orangutan alarm call rates and combinations. *Behav. Ecol. Sociobiol.* **67**, 519-528.
- Lameira, A. R., Hardus, M. E., Nouwen, K. J. J. M., Topelberg, E., Delgado, R. A., Spruijt, B. M., Sterck, E. H. M., Knott, C. D. and Wich, S. A.** (2013b). Population-specific use of the same tool-assisted alarm call between two wild orangutan populations (*Pongo pygmaeus wurmbii*) indicates functional arbitrariness. *PLoS ONE* **8**, e69749.
- Martin, P. A.** (2004). On Webster's horn equation and some generalizations. *J. Acoust. Soc. Am.* **116**, 1381-1388.
- Matrosova, V. A., Volodin, I. A., Volodina, E. V. and Babitsky, A. F.** (2007). Pups crying bass: vocal adaptation for avoidance of age-dependent predation risk in ground squirrels? *Behav. Ecol. Sociobiol.* **62**, 181-191.
- Morse, P. M. and Ingard, K. U.** (1968). *Theoretical Acoustics*. Princeton, NJ: Princeton University Press.
- Obrecht, J.** (1990). Ry Cooder – talking country blues and gospel. *Guitar Play. Magn.* **24**.
- Ohala, J. J.** (1984). An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica* **41**, 1-16.
- Rijksen, H. D.** (1978). A field study on Sumatran orangutans (*Pongo pygmaeus abelii*, Lesson 1827): Ecology, behavior, and conservation. PhD thesis Landbouwhogeschool, Wageningen University, The Netherlands.
- van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Utami, S. S. and Merrill, M.** (2003). Orangutan cultures and the evolution of material culture. *Science* **299**, 102-105.
- van Schaik, C. P., van Noordwijk, M. A. and Wich, S. A.** (2006). Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). *Behaviour* **143**, 839-876.