Bioacoustics
The International Journal of Animal Sound and its Recording

ISSN: 0952-4622 (Print) 2165-0586 (Online) Journal homepage: http://www.tandfonline.com/loi/tbio20

COMPUTER-AIDED ANALYSIS OF ACOUSTIC PARAMETERS IN ANIMAL VOCALISATIONS: A MULTI-PARAMETRIC APPROACH

LARS SCHRADER & KURT HAMMERSCHMIDT

Bioacoustics
The International Journal of Animal Sound and its Recording, 1997, Vol. 7, pp. 247–265
0952-4622/97 $10

COMPUTER-AIDED ANALYSIS OF ACOUSTIC PARAMETERS IN ANIMAL VOCALISATIONS: A MULTI-PARAMETRIC APPROACH

LARS SCHRADER[1] AND KURT HAMMERSCHMIDT[2]

[1]*Institut für Verhaltensbiologie, FU Berlin, Haderslebener Straße 9, D-12163 Berlin, Germany*
[2]*Laboratory of Comparative Ethology, NICHD, NIH, Elmer School Road, Box 529, Poolesville, MD 20837, USA*

ABSTRACT

The computer-aided analysis of acoustic signals of mammals is still a problem, as often (a) sound structures are complex, (b) vocal repertoires often comprise an enormous variety of vocalisations, (c) recordings are influenced by the acoustic conditions of the environment, and (d) the distance and spatial orientation of the sender to the microphone changes. In recent software packages for the analysis of acoustic signals, procedures are integrated which allow the calculation of a variety of signal features. However, these algorithms are often problematic under the conditions mentioned above. In this paper, we present a multi-parametric approach which reduces these problems and which allows a quantitative and reproducible analysis of complex animal vocalisations. Our approach comprises the following aspects: (1) reduction of influences of recording conditions, (2) determination of different sound features and (3) calculation of parameters to characterise these sound features. All calculations are done on the basis of the digitised spectrograms. Special attention is given to the use of smoothing algorithms and dynamic thresholds in order to estimate sound features and to reduce influences resulting from recording conditions. The suitability of our approach has been demonstrated successfully for vocalisations of different species.

Keywords: computer sound analysis, multiparametric approach, vocalisations, mammals, primates

INTRODUCTION

For a number of questions concerning acoustic communication it is necessary to consider subtle variations of signals. Estimation of covariation of vocalisations with internal state (Jürgens 1979), social and nonsocial context (Bauer 1987, Fischer et al. 1995, Geiss and Schrader in press, Goedeking and Immelmann 1986, Gouzoules and Gouzoules 1990, Schrader and Todt 1993) or individuality (Allenbacher et al. 1995, Hammerschmidt and Todt 1995, Symmes et al. 1979) have shown finely graded differences among signals. It is difficult to determine such variations, especially in vocalisations of mammals. Acoustic structure in this group can be rather complex, as energy is often spread over a wide frequency range, and frequency structures are modulated within a given signal. In addition, with regard to sound production, signals can be noisy or even atonal. Thus, it often is difficult to decide which parameters should be measured to characterise the properties of a signal.

Due to limited memory and calculation capacity, early computer programs for signal analysis primarily focused on specific tasks, such as on analysing single pure tones (e.g. MNKVOC; described in Biben et al. 1989). Computer programs developed more recently comprise a higher variety of procedures to calculate the physical structures of sounds. For example, a peak-detection function determines the frequency with the highest amplitude in the spectrogram (e.g. SIGNAL, Engineering Design; AVISOFT, R. Specht; CANARY, Bioacoustics Research Program), and a pitch function calculates the time course of the fundamental frequency (e.g. SIGNAL, Engineering Design). Especially in software for speech analysis (e.g. ESPS/waves+, entropic; CSL, Kay Elemetrics Corp.), algorithms of linear predicting coding (LPC) are integrated for calculating formants and fundamental frequency. Owren and Linker (1995) discussed the advantages of LPC for data reduction and resynthesis in non-human primate calls. However, they concluded that many problems still remain when LPC is used in the analysis of animal sounds. Other studies on differences in the vocal tract of non-human primates (Schön Ybarra 1995) also concluded that the use of most pitch extraction algorithms, which are based on the condition of human male speakers (Fitch and Hauser 1995), need to be treated with care.

Considering the differences between human and non-human primates it is obvious that even more problems will arise if vocalisations of other mammals are analysed. Here we present a multiparametric approach which was developed in order to analyse vocalisations especially with complex structures. Primarily, we applied this multiparametric approach to vocalisations of non-human primates (Hammerschmidt et al. 1994, Hammerschmidt and Todt 1995, Fischer et al. 1995), but this analysis has proven to be suitable also for vocalisations of other mammals (e.g. domestic pigs: Schrader and Todt 1996) or even for complex bird calls (e.g. Hooded Crow: Allenbacher et al. 1995, or Bald Ibis: Böhner and Hammerschmidt 1996).

To deal with the problems mentioned above, our method of signal analysis is based on the spectrograms of signals. Spectrograms are the most comprehensive way to represent the distribution of signal energy in the frequency as well as in the time domain. The transformed acoustic signal in the cochlea has a corresponding structure, and such spectra, similar to the fourier transformed signals, could be found at different neuronal levels (Picklers 1988, Zenner 1994). Furthermore, the inspection of spectrograms allows an overall view of the general signal characteristics and of the quality of recordings.

In general, our method is a two step analysis: First, certain physical structures are extracted from the digitised frequency spectra of the signals. Second, specific sets of parameters are calculated that describe these signal structures. One set of parameters characterises formant-like structures of signals. These structures represent the frequency ranges with highest energy within the spectrum and therefore may reflect the resonance- and filter-properties of the vocal tract. A second set of parameters is calculated to estimate dominant frequency peaks of signals, which correspond to the harmonics in tonal sounds. The third set of parameters describes the statistical distribution of the spectral energy. The latter set is especially helpful in characterising atonal or noisy vocalisations.

Before we extract the signal structures from the spectrograms, preliminary calculations can be done in order to reduce the influence of different recording conditions, of changes in distance and orientation of the vocalising animal to the microphone, and of the acoustic environment (Piercy et al. 1977, Wiley and Richards 1978). Different algorithms can be used to separate noise from signal amplitudes and to determine the start and the end of a signal. These calculations are required if the acoustic signals to be analysed are recorded in the field.

PROCEDURE AND ALGORITHMS

(1) Dealing with noise

For an automatic computer-aided analysis of signals it is necessary to make some preliminary decisions:

A first decision is to define a threshold which determines the start- and the end-point of the signal. Besides duration, other signal parameters strongly depend on the onset of a signal. Examples are the slope of a given frequency structure, the start frequency of a dominant frequency peak or the start frequency of the peak frequency (see below). Although a sudden and high amplitude onset often significantly marks the start of a signal, a fading amplitude or reverberation at the end makes it difficult to define the limits of a signal. In general, a threshold is necessary in order to reliably determine the start- and end-points. To deal with different recording conditions, we calculate a separate threshold for each signal. Instead of the amplitude or the envelope of the signal we use the frequency amplitudes of the spectrogram to determine the threshold. This allows the detection of certain signal structures even if the signal-to-noise ratio is low. In such cases the start and end of a signal could not be detected in the amplitude or the envelope.

One possibility is to calculate the threshold in relation to the maximum amplitude of the frequency amplitude found in the signal. However, this has the disadvantage that one single high amplitude within the spectrum would unproportionally affect the threshold. Therefore we use the mean of the ten highest frequency amplitudes found in the different time segments of a signal as a reference for the threshold calculation. The value of this threshold can be expressed in percent of the mean amplitudes of the ten highest frequencies. This percentage can be adjusted according to the quality of recordings.

Second, low frequency background noise is common in many recordings. Therefore, it is often necessary to use a cut-off frequency (high pass filter) to discard the lower frequencies. This is an easy way to improve the calculation when the signal is not within these lower frequencies. In this way the low frequency background noise in Figure 6a can be excluded with a cut-off frequency of 400 Hz and has no influence on the further calculation of the dominant frequency peaks (Figure 6b) or the distribution of the frequency amplitudes (Figure 6c).

In a third decision, signal and background noise must be separated if the respective frequencies overlap. The easiest way is to determine a threshold with a fixed value for the signal-to-noise ratio. However, the continuous change in distance and spatial orientation of the sender, which is quite common in field recordings, results in different signal-to-noise ratios between recordings, or even within one signal. Therefore a dynamic threshold is useful. One possibility to separate frequencies with high amplitudes from frequencies with low amplitudes is to calculate the central tendency (mean or median, depending on the distribution) and the variation of the frequency amplitudes of a signal. Using only the central tendency as a reference for the threshold would result in a threshold which is independent from the regularity of the course of the frequencies in the spectrum. Using the combination of the variation with the central tendency has the advantage that the threshold will be higher for signals with a highly irregular spectrum than for signals with a more regular spectrum. As an irregular spectrum often results from a poor signal-to-noise ratio, this kind of threshold will reduce effects of poor recording conditions. Figure 2b gives an impression of the effect of such a dynamic threshold.

An alternative method to separate noise from signals is to calculate the change point (Siegel and Castellan 1988) of the distribution of frequency amplitudes. If the background noise is distributed randomly, the frequency amplitudes of the noise and the frequency amplitudes of the signal belong to two different distributions. Then the value of the change point between both distributions can be used as a threshold to separate the signal from noise. The value of this threshold is shown in Figure 4b and 5b. It shows that the randomly distributed noise will be separated well from the spectrum.

(2) Formant-like structures

In spectrograms, the amplitude of a frequency is coded by colour or increasingly darker shades of grey (see Figure 1). The different amplitudes of certain frequency ranges are the result of resonance- and filter-properties of the vocal tract. In human speech these structures are labelled formants, and they are crucial for the quality and the meaning of signals (e.g. the vowels of human speech; Liebermann 1984). In general, the structure of the vocal tract in other mammal species corresponds to the vocal tract of humans. Species-specific are the length, the diameters and the physical properties of the tubes in the vocal tract. These anatomical differences result in different resonance- and filter-properties. Nevertheless, the properties of the vocal tract are important for the sound characteristics in humans as well as in other mammals as they will result in frequency ranges with different energy (Figure 1). To distinguish such frequency ranges with high energy in non-human vocalisations from the formants in human vocalisations, we name them formant-like structures.

To analyse which frequencies have predominant amplitudes and therefore may constitute formant-like structures, we determine an amplitude threshold. With this definition, frequencies with an amplitude higher than the threshold belong to a formant-like structure. Such a threshold, however, should depend on the actual amplitudes of the signal, as it is used to separate frequencies with high amplitudes from frequencies with lower amplitudes within a given signal. As mentioned above, it has proven to be useful to take the central tendency and the variation of the distribution of frequency amplitudes as the threshold.

Before we calculate the range of formant-like structures we smooth the course of frequency amplitudes to emphasise this feature. This is necessary as the course of frequency amplitudes is often rather irregular (Figure 2a). In addition, even within ranges of high amplitudes—for example between 2 and 5 kHz in Figure 2a—some frequencies with lower amplitudes may occur. Often it is difficult to decide if these amplitude gaps result from environmental influences or from sound production itself. Therefore we use a floating mean to remove amplitude irregularities from the frequency amplitudes. In the resulting modified signal (Figure 2b), it is much easier to estimate those frequencies which exceed the threshold. Consequently, the range of formant-like structures can be estimated unambiguously. Such an example is shown in Figure 2b. Here the frequency amplitudes are smoothed using a 500 Hz rectangular averaging window, and the threshold is set by the mean plus the standard deviation of the smoothed amplitude-values.

If formants are determined subsequently for each spectrum, the temporal course of the formant-structure can be identified clearly (compare Figure 1 and Figure 3).

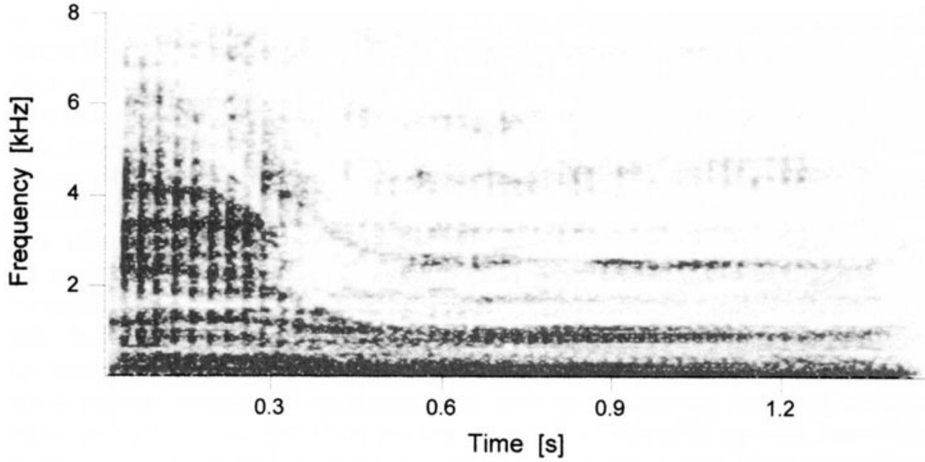


Figure 1. Spectrogram of a 'grunt' of a pig.

(3) Dominant frequency peaks

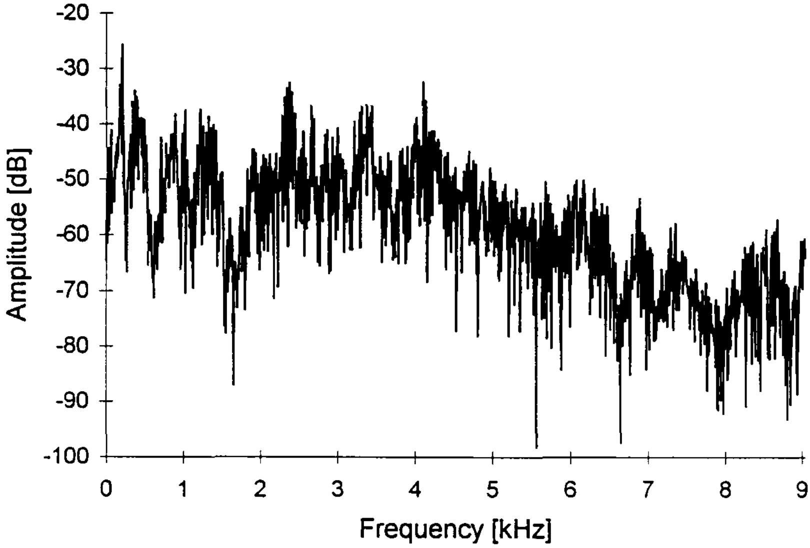
In the last section we presented approaches to calculate formant-like structures, or, in other words, frequency ranges with high amplitudes. Beside these ranges, certain frequencies with peak amplitudes are also crucial for the quality of sounds. This is especially obvious in tonal signals where energy is concentrated on the fundamental and the harmonics. In this section we describe algorithms to estimate such dominant frequency peaks.

The crucial step to determine dominant frequency peaks is to distinguish local, less pronounced peaks from the dominant peaks. First, two different averaging curves of the power spectra are calculated. The first curve (dotted line in Figure 4c and 5c) has nearly the same function as the floating mean for calculating the formant-like structures as it may reflect the filter properties of the vocal tract. The second curve (solid line in Figure 4c and 5c) results if irregular components are filtered from the spectrum and may reflect the frequency spectrum of the glottal in tonal sounds. These curves are calculated using equation (1).

$$\tilde{a}_i = \sum_{j=1}^{r} \left(\frac{1}{w} \sum_{i=\frac{1-w}{2}}^{\frac{w-1}{2}} a_i \right) \qquad \text{Equation (1)}$$

(a = value of the respective frequency amplitude, r = repetition factor, w = frequency window)

(a)



(b)

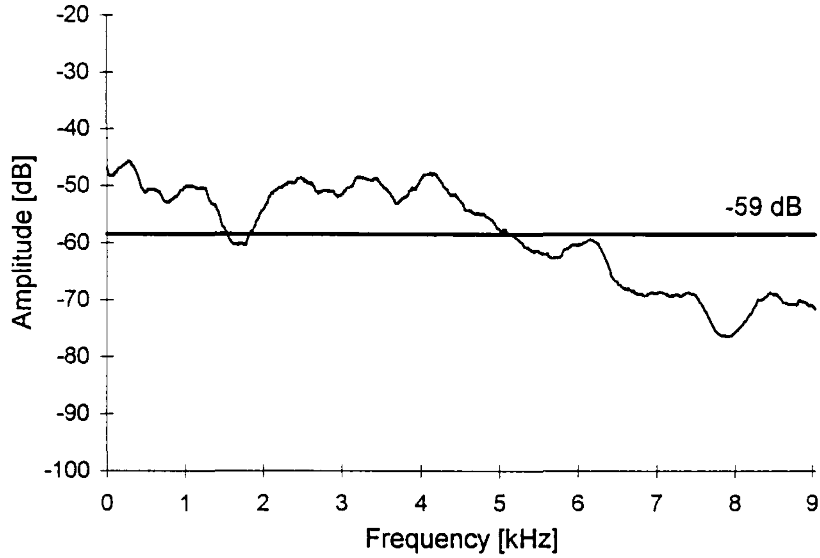


Figure 2. (a) Course of frequency amplitudes (derived from the first time segment of the spectrogram in Figure 1). (b) Frequency amplitudes of Figure 2(a) were smoothed using a 500 Hz rectangular averaging window. Bold line indicates the threshold at −59 dB for the determination of formant-like structures. Threshold is calculated using the mean plus the standard deviation of the smoothed frequency amplitudes.

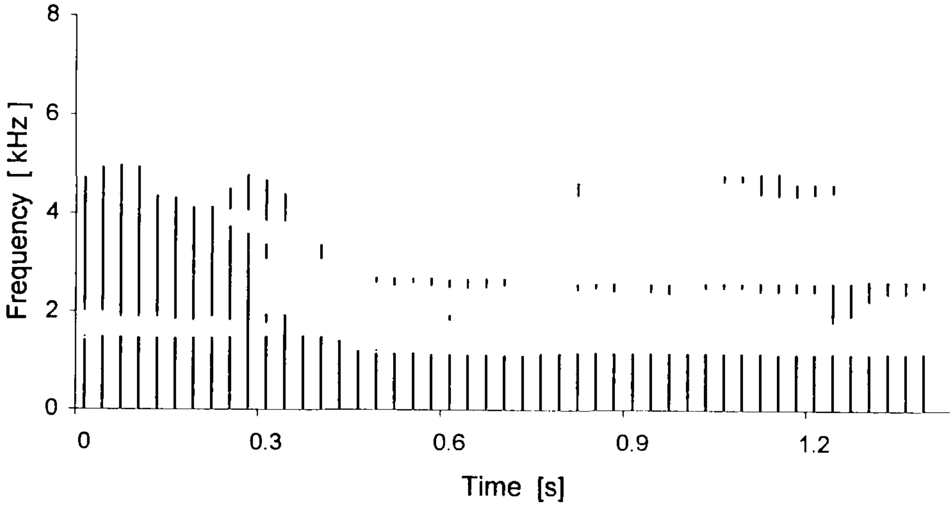


Figure 3. Temporal course of the formant-like structures of the pig's grunt in Figure 1. For each given time segment the lines indicate the ranges of frequency amplitudes which exceed the threshold.

The equation for calculating the averaging curves (Equation 1) comprises two variables: the width of the frequency window (w) and the number of repetitions of the smoothing algorithm (r). The width of frequency windows and the number of repetitions of the smoothing algorithm in equation (1) are empirical values. For calculating the first curve we used w = 11 and r = 10, for the second curve w = 5 and r = 3. These values have been proven to be appropriate for various vocal types of primate vocalisations (Hammerschmidt et al. 1994, Todt et al. 1995). In general, the values can change, as they depend on the chosen frequency resolution. Nevertheless, in our experience it is more appropriate to change the frequency resolution by varying the number of FFT-points or the frequency range instead of varying the values used above. Hence, if the relation between frequency resolution and sound structure is appropriate, the algorithm also fits with other vocalisations, as could be shown in bird calls (Allenbacher et al. 1995, Böhner and Hammerschmidt 1996).

In a last step, a peak detection algorithm searches for frequencies with maximum amplitudes within dominant frequency parts. In some cases the dominant frequency parts have more than one local maximum (see Figure 5c, two maxima at about 1.3 kHz and two maxima at about 2.6 kHz). Therefore, the peak detection algorithm tests if the difference between the amplitudes of the dominant frequency peaks and the amplitude of the frequency gap between the peaks has a higher value than the value of noise-threshold (cf. section 1). If the differences between the amplitude of the peaks and the

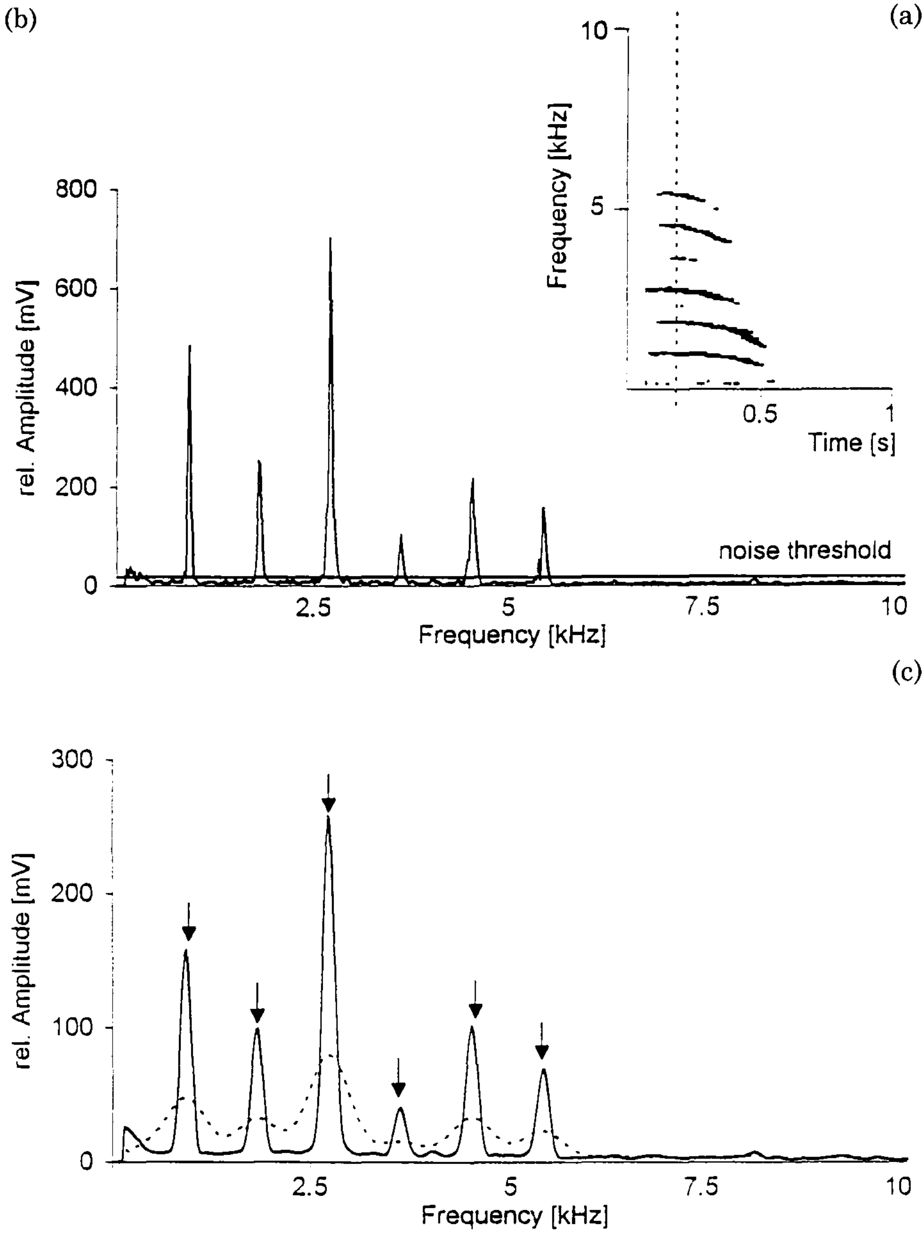


Figure 4. Examples of how the sliding average processes the frequency amplitudes (powerspectra) of single time segments. (a) Spectrogram of a tonal sound (Barbary macaques). The dotted line indicates the time segment which is shown in (b) and (c). (b) Course of frequency amplitudes of a time segment of the sound in (a). (c) The two sliding averages of the time segment in (b). The dotted line marks the first sliding average, the solid line the second (Equation 1). The arrows mark the dominant frequency peaks detected by the algorithm.

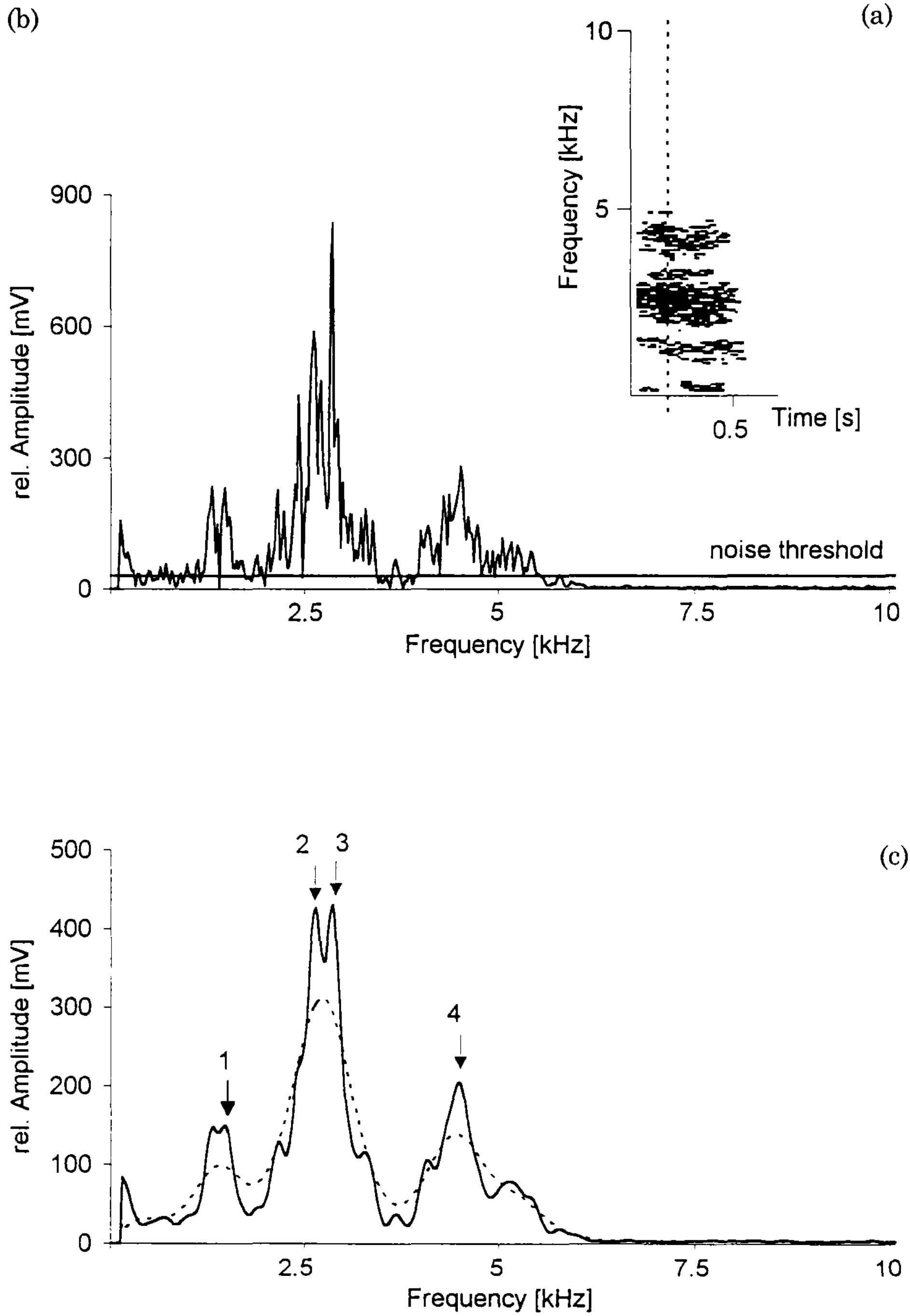


Figure 5. The same procedures as Figure 4 for an atonal vocalization of a Barbary macaque.

amplitude of the frequency gap is smaller than the noise threshold, the lower peak will be discarded and only the higher peak will be determined as a dominant frequency peak (see Figure 5c; arrow 1). Otherwise both peaks will be counted (see Figure 5c; arrows 2 and 3). Additionally, the difference between a local maximum and the second threshold (dotted line in Figure 5c) must exceed the value of the noise threshold which is not fulfilled by the local maximum at about 3.8 kHz in Figure 5c.

In tonal sounds, the dominant frequency peaks refer to the fundamental frequency and their harmonics or overtones (Figure 4). In contrast, in atonal signals, dominant frequency peaks will reflect the frequencies with highest energy within the formant-like structures (Figure 5). Again, this kind of analysis is done for each time segment of the signal and will result in the time course of the dominant frequency peaks (Figure 6b).

(4) Statistical distribution of special energy

Animal vocalisations often vary enormously in structure. Therefore, different calls may be described best by calculating different signal structures. The structure of tonal calls, for instance, may be best represented by dominant frequency peaks, whereas other vocalisations like pig grunts may be best reflected by formant-like structures. In addition, many sounds of animals are noisy or have noisy parts so that certain signal structures sometimes can not be sufficiently analysed. Therefore, additional algorithms are needed to characterise different signals with the same parameters. Such parameters address the statistical distribution of spectral energy. An appropriate measure is the quartiles of the distribution of frequency amplitudes which are calculated by the following equation:

$$f_q = q \cdot \frac{1}{4} \sum_{i=1}^{n} a_i \qquad \text{Equation (2)}$$

(fq = frequency of the quartile, q = lst, 2nd, 3rd quartile, a = value of the respective frequency amplitude)

Hence, quartiles are determined as follows: First, the frequency amplitudes are summed. Then the lst, 2nd, and 3rd quarters of the summed amplitude are calculated. Next, amplitudes will be summed again, starting with the amplitude value of the lowest frequency. The frequency value at which the sum reaches the limit of the lst quarter of the amplitudes is the first quartile. In the same way, the frequency for the 2nd and 3rd quartile will be calculated. If this calculation is done for every time segment of the sound, three curves will result

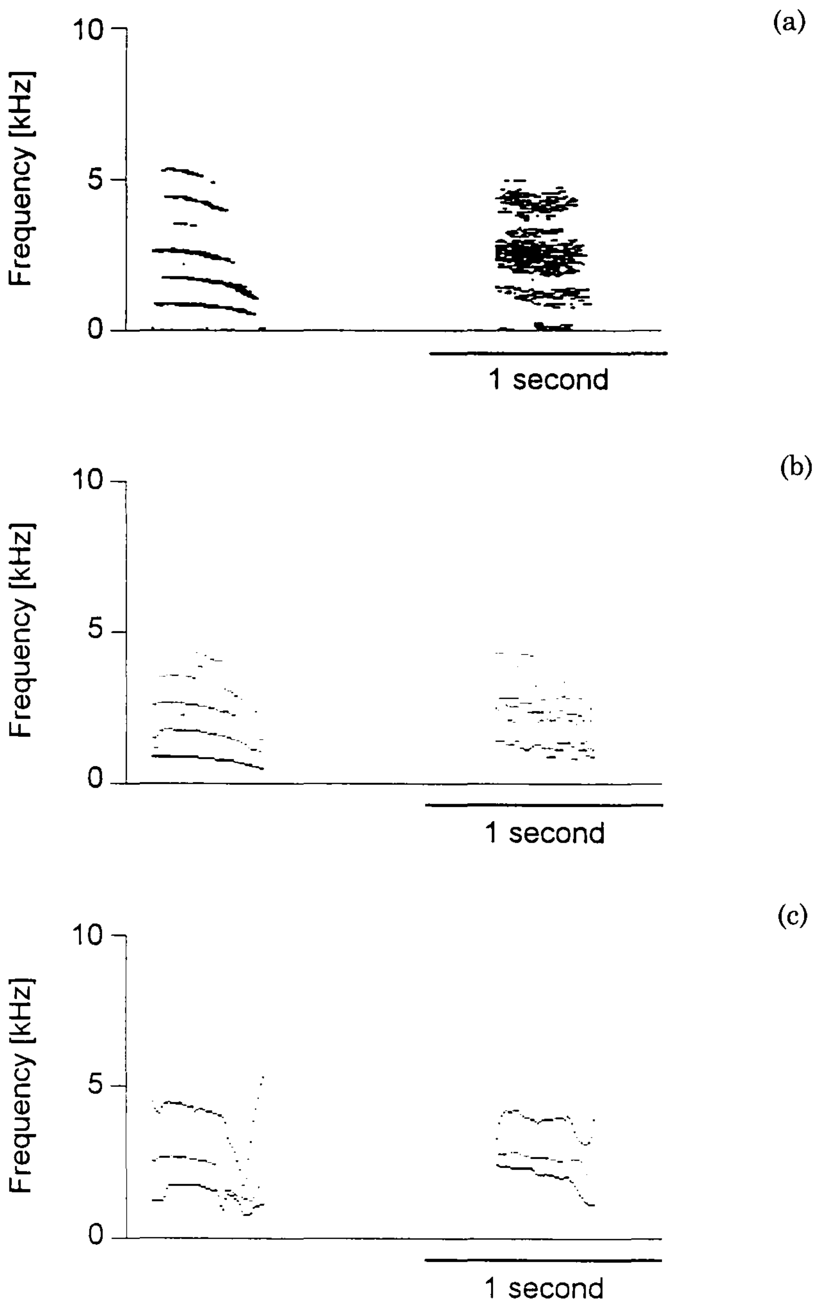


Figure 6. (a) Spectrograms of a tonal (left) and an atonal sound (right) of a Barbary macaque; (b) the respective curves of the first four dominant frequency peaks. The first dominant frequency peak of the tonal sound (left) shows the modulation of the fundamental frequency; (c) the respective curves of the 1st, 2nd and 3rd quartile of the distribution of the frequency amplitudes.

which show the modulation of the distribution of frequency amplitudes (Figure 6c).

(5) Parameters representing sound structure

In the preceding steps of analysis, we estimated specific signal structures on the basis of the digitised spectrograms. These structures represent the acoustic properties of signals in the frequency as well as in the time domain. Although this step already results in considerable data reduction, statistical analysis of acoustic properties is not yet possible. For example, it is not possible to make statistical calculations with the first dominant frequency peak, as the first dominant frequency peak consists of a series of frequency values for each time segment in a given signal. Therefore, in a last step, parameters are to be calculated which are suitable to characterise these structures. For instance, one can estimate the mean of the frequency values of the first dominant frequency peak. With this single parameter it is possible to characterise the central tendency of this structure.

For each signal structure (formant-like structures, dominant frequency peaks, quartiles of frequency amplitudes) a set of parameters can be calculated to represent the structure's (a) central tendency, (b) spectral extension, (c) temporal modulation and (d) the temporal distribution of certain parameters in the signal.

(a) *central tendency*: The central tendency can be described by the arithmetical mean of the structure, for example the mean of frequency values of the first dominant frequency peak. If data are not normally distributed, the median is more appropriate.

Start and end frequencies can comprise information about transient and fade properties of the acoustical system when they are related to the central tendency of a signal structure.

(b) *spectral range*: The spectral range is calculated from the lowest and highest frequency values of a given signal structure. For example, the frequency range of the median frequency (the 2nd quartile) will comprise information about the overall range of the spectral energy of a signal.

The formant-like structures are frequency ranges *per se*. Here, it is possible to estimate the difference between the lowest and highest limit of a given formant-like structure. Another possibility is to calculate the mean difference between the low and high limits of a formant-like structure in each time segment.

(c) *temporal modulation*: Statistical values like the variation or the standard deviation are suitable to characterise the overall modulation of a signal structure (e.g. the variation of frequency values in Figure 6c) (Goedeking and Immelmann 1986). Short-term modulation can be described by the following index:

$$m_f = \frac{\sum_{i=2}^{n-1} \left| \frac{1}{3} \sum_{i-1}^{i+1} f_i - f_i \right|}{\Delta t} \qquad \text{Equation (3)}$$

(f = value of the respective frequency, Δt = fixed time interval)

This index is calculated as the difference between the original values and the smoothed values. In Equation (3) f_i is the actual frequency value and Δt is a fixed time interval to which the difference (absolute value) will be related. Note that a floating mean for three values is used in Equation (3). However, other types of floating means are possible. It is important to point out that the value of this index directly depends on the chosen time resolution as well as the frequency resolution of the FFT.

The temporal trend (e.g. increase or decrease) of a given frequency structure can be characterised by its slope. The slope of the entire signal will describe the overall trend, whereas, for instance, the slope of the first third will describe the start-trend, and the slope of the last third will describe the end-trend, respectively. One possibility to calculate the slope is to use the method of least squares. Here a linear function is calculated which minimises the distances between the values and the approximated function (Fröhlich and Becker 1971).

(d) *temporal distribution of parameters*: The temporal distribution of parameters in the signal can bear important information. For example, it may be crucial for sound quality if the highest values of the median frequency (the 2nd quartile) are at the beginning or at the end of a signal.

DISCUSSION

The aim of this paper was to present a computer-aided method for extracting parameters of animal vocalisations with complex acoustic features or vocalisations recorded under varying acoustic conditions. Suggestions were given for the calculation of different signal structures.

The main advantage of computer-aided analyses in general is its speed and automation and the physically reliable calculation of sound representations such as spectrograms or powerspectra. The latter yields high comparability of calculations even if done with different analysis-systems, but is undermined by the influences of different recording conditions. These are mainly caused by environmental disturbances and continuous changes in distance and spatial orientation of the sender to the microphone (Dabelsteen *et al.* 1993, Wiley and Richards 1978).

One way to extract specific sound features regardless of these problems would be to calculate the fundamental frequency (energy source) and the formants (filter function of vocal tract). Although there exist powerful tools to calculate these acoustic features, namely pitch- and LPC-functions, these procedures sometimes fail on human vocalisations and more often fail on animal sounds (Owren and Linker 1995).

The algorithms for pitch calculation can be inappropriate, because the source harmonics may be too sparse to provide much information about the fundamental frequency, which is typical for the high-pitched calls of smaller primates (Fitch and Hauser 1995). In other cases, the lower harmonics may be filtered by passing the vocal tract, as in many young primates' calls and in the cries of human infants. One other reason concerns the sound production. For example, Schön Ybarra (1995) argued that non-human primates could be able to produce two independent vocal fold vibrations as a result of mechanical properties of their vocal lips. But the probably most frequent problem is that pitch analysis fails in noisy or disturbed sounds. Such sounds are common in primates and other mammals, in particular in field recordings.

Another way to estimate characteristic sound structures is to calculate the formants (source filter function) using a LPC function. Especially in speech analysis, several algorithms are used to determine the formants. These are based on the maximum-entropy method (LPC, e.g. autocorrelation, methods of covariance and Burg-methods) or on a cepstral-analysis (e.g. cepstrally smoothed-method)—for further discussion see Hess (1983), Owren and Linker (1995). These methods calculate the formant-maxima and parts of the bandwidth of the formants. However, besides general similarities between the vocal tract of humans and other mammals, there still are several differences which impair the use of these methods for many mammal sounds. In addition, the investigator must select the order of the linear predictor which has essential influence on the resulting number of formants, and requires a sufficient knowledge of the acoustic properties of the respective species. Nevertheless, some recent studies showed that LPC can be useful in certain cases. For example, Owren (1990) showed that vervet monkeys reacted to playbacks of resynthesized (by LPC) alarm calls to the same degree as to natural calls. In general, LPC-spectra are a useful tool for data reduction and the description of the time-frequency modulation of sounds. Nevertheless, it must be taken into account that the results of LPC do not necessarily reflect the real filter properties of the vocal tract.

The use of spectrograms as a basis to extract signal structures and to calculate signal parameters requires less knowledge about sound production. Spectrograms represent the distribution of signal energy in the frequency as well as the time domain in a physically exact way. Of course, it must be taken into account that the settings for the Fast Fourier Transformation have a strong influence on the resolution of the resulting spectrogram. The choice of frequency range, FFT-size, length of the time segment, and of the transformation window must be done with respect to the quality of the signals to be analysed. For example, to detect harmonics in tonal signals it is indispensable that the frequency resolution is clearly high enough to distinguish differences in frequency with lower values than the fundamental frequency. Hence, it is important to get a fundamental idea of the quality of signals to decide the settings of the FFT. However, it is an advantage of the suggested method that it is not necessary to know the exact way that the signal is generated.

The combination of parameters from multiple sound features, like the distribution of frequency amplitudes, dominant frequency peaks, and formant-like structures, has several advantages: First, vocal repertoires of species are usually composed of different types of sounds. The sounds can differ in crucial features, such as in their degree of tonality (cf. Figure 6). Hence, some parameters which might be useful to describe one type of a sound might be inappropriate for describing another.

Second, parameters can be estimated that potentially correlate with other parameters which can not be determined. For instance, the fundamental frequency in many vocalisations eludes exact calculation. Also, the absolute values of the signal amplitude often cannot be determined. This is due to changing spatial orientations and distances of the sender. In addition, the environmental conditions can mask the signal amplitude, especially in field recordings. With both sound structures (fundamental frequency and signal amplitude), the distribution of frequency amplitudes seem to correlate, as high-pitched and loud calls tend to have a maximum of energy in high frequency parts.

Third, it is not possible to know the decisive signal parameters in advance. A good example are the age graded changes in call parameters of Barbary macaques (Hammerschmidt et al. 1994). Normally, one would expect that the fundamental frequency is an appropriate measure for the age of a sender. However, the most decisive call parameters in that study were peak frequency (frequency with the highest amplitude in the call) and maximum frequency range. These parameters covaried strongly with the age of the sender. Another study on individual differences revealed that an increase in analysed call parameters leads to an increase in correct assignment of calls to the individuals (Hammerschmidt and Todt 1995). In addition, individuality was not encoded in one or a discrete set of specific parameters. Instead, the combination of different parameters represented the individual voice quality of each subject.

It has been demonstrated already that dynamic thresholds described in this paper are suitable to solve the problems of varying signal-to-noise ratios (Fischer et al. 1995, Hammerschmidt and Todt 1995). However, it is important to emphasise that the type of threshold depends on the respective sound features and the quality of recordings. If the aim of the analysis is the extraction of the prominent signal parts, the calculation of central tendency and the variation seems to be appropriate. This kind of threshold deals well with changing sound amplitudes. A minor disadvantage is the dependence of the mean on the relation of signal and noise amplitudes in a given spectrum. If this has a negative influence on the analysis the calculation of a change point could be more appropriate. The change point works reliably on sounds with a more or less random noisy background. Even if additional decisions about cut-off frequency and signal limits must be made, such statistically defined thresholds have the advantage that they are not dependent on ad hoc evaluations of the researcher. Therefore, the results are better reproducible and comparable.

In general, the presented approach has two advantages: the simultaneous analysis of different sound features, and the use of dynamic thresholds. Both take into account different sound productions and recording conditions. This flexibility allows a broad use for the analysis of diverse animal vocalisations, which are marked by complex structures.

ACKNOWLEDGEMENT

We would like to thank Jörg Böhner, Julia Fischer, Henrike Hultsch, Marc Naguib, John Newman and Dietmar Todt for helpful discussion on earlier drafts of this manuscript.

REFERENCES

Allenbacher, R., Böhner, J. & Hammerschmidt, K. (1995). Individuelle Merkmale im "krah"—Ruf der Nebelkrähe *Corvus corone cornix*. *J. Ornithol.*, **136**, 441–446.

Bauer, H. R. (1987). Frequency code orofacial correlates of fundamental frequency. *Phonetica*, **44**, 173–191.

Biben, M., Symmes, D. & Bernhards, D. (1989). Contour variables in vocal communication between squirrel monkey mothers and infants. *Development Psychology*, **22**, 617–631.

Böhner, J. & Hammerschmidt, K. (1996). Computer-aided acoustic analysis of complex bird calls. *Bioacoustics*, **6**, 313–314.

Dabelsteen, T., Larsen, O. N. & Pedersen, S. B. (1993). Habitat induced degradation of sound signals: Quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song. *J. Acoust. Soc. Am.*, **93**, 2206–2220.

Fischer, J., Hammerschmidt, K. & Todt, D. (1995). Factors affecting acoustic variation in Barbary-macaque (*Macaca sylvanus*) disturbance calls. *Ethology*, **101**, 51–66.

Fitch, W. T. & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustic, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.*, **37**, 191–219.

Fröhlich, W. D. & Becker, J. (1971). *Forschungsstatistik*. Bouvier Verlag; Bonn.

Geiss, S. & Schrader, L. (in press). Temporal and structural features of infant calls in relation to caregiving behaviour in common marmosets, *Callithrix j. jacchus. Behavioural Processes*.

Goedeking, P. & Immelmann, K. (1986). Vocal cues in cotton-top tamarin play vocalizations. *Ethology*, **73**, 219–224.

Gouzoules, H. & Gouzoules, S. (1990). Body size effects on the acoustic structure of pigtail macaques (*Macaca nemestrina*) screams. *Ethology*, **85**, 324–334.

Hammerschmidt, K., Ansorge, V. & Fischer, J. (1994). Age-related variations in the vocal repertoire of Barbary macaques. In *Current primatology. Volume II, Social development, learning and behaviour* (J. J. Roeder, B. Thierry, J. R. Anderson and N. Herrenschmidt, eds). Université Louis Pasteur; Strasbourg, pp. 287–295.

Hammerschmidt, K. & Todt, D. (1995). Individual differences in vocalizations of young Barbary macaques (*Macaca sylvanus*): a multi-parametric analysis to identify critical cues in acoustic signalling. *Behaviour*, **132**, 381–399.

Hess, W. (1983). *Pitch Determination of Speech Signals*. Springer-Verlag; Heidelberg.

Jürgens, U. (1979). Vocalizations as an emotional indicator. A neuroethological study in the squirrel monkey. *Behaviour*, **69**, 80–117.

Liebermann, P. (1984). *The Biology and Evolution of Language*. Harvard University Press; Cambridge, Mass.

Owren, M. J. (1990). Classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*): II. Synthetic calls. *J. Comp. Psychol.*, **104**, 29–41.

Owren, M. J. & Linker, C. (1995). Some analysis methods that may be useful to acoustic primatologists. In *Current topics in primate vocal communication* (E. Zimmermann, J. D. Newman, U. Jürgens, eds). Plenum Press; New York, pp. 1–27.

Picklers, J. O. (1988). *An Introduction to the Physiology of Hearing*. Academic Press; London.

Piercy, J. E., Embleton, T. F. W. & Sutherland, L. C. (1977). Review of noise propagation in the atmosphere. *J. Acoust. Soc. Am.*, **61**, 1403–1418.

Schön Ybarra, M. A. (1995). A comparative approach to the non-human primate vocal tract: implications for sound production. In *Current topics in primate vocal communication* (E. Zimmermann, J. D. Newman, U. Jürgens, eds). Plenum Press; New York, pp. 185–198.

Schrader, L. & Todt, D. (1993). Contact call parameters covary with social context in common marmosets, *Callithrix j. jacchus. Anim. Behav.*, **46**, 1026–1028.

Schrader, L. & Todt, D. (1996). Vocal cues reflect physiological stress response in domestic pigs (*Sus scrofa domestica*). Proceedings 30th International Congress ISAE, Guelph, 17.

Siegel, S. & Castellan, N. J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company; New York.

Symmes, D., Newman, J. D., Talmage-Riggs, G. & Katz-Lieblich, A. (1979). Individuality and stability of isolation peeps in squirrel monkeys. *Anim. Behav.*, **27**, 1142–1152.

Todt, D., Hammerschmidt, K., Ansorge, V. & Fischer, J. (1995). The vocal behavior of Barbary macaques (*Macaca sylvanus*): call features and their performance in infants and adults. In *Current topics in primate vocal communication* (E. Zimmermann, J. D. Newman, U. Jürgens, eds). Plenum Press; New York, pp. 141–160.

Wiley, R. H. & Richards, D. G. (1978). Physical constraints on acoustic communication

in the atmosphere: implications for the evolution of animal vocalizations. *Behav. Ecol. Sociobiol.*, **3**, 69–94.
Zenner, H. P. (1994). *Hören*. Georg Thieme Verlag; Stuttgart, New York.

Received 11 March 1996, revised 25 July 1996 and accepted 2 August 1996.