ЕВРОАЗИАТСКАЯ РЕГИОНАЛЬНАЯ
АССОЦИАЦИЯ ЗООПАРКОВ И АКВАРИУМОВ
EUROASIAN REGIONAL ASSOCIATION
OF ZOOS AND AQUARIUMS

ПРАВИТЕЛЬСТВО МОСКВЫ
GOVERNMENT OF MOSCOW

МОСКОВСКИЙ ЗООЛОГИЧЕСКИЙ ПАРК
MOSCOW ZOO

Научные исследования в зоологических парках
Scientific Research in Zoological Parks

Выпуск 28
Volume 28

Москва
Moscow

2012

УДК [597.6/599:639.1.04]:59.006
ББК 20.18:28.6
Н34

Под редакцией первого заместителя генерального директора Московского зоопарка, члена-корреспондента РАЕН, д. б. н. С.В. Попова

Н34 Научные исследования в зоологических парках. Scientific Research in Zoological Parks. Вып. 28 / Под ред. С.В. Попова. – М.: ГБУ «Московский зоопарк» – Тверь: ООО «Издательство «Триада», 2012. – 184 с.
ISBN 978-5-904012-30-4

Статьи настоящего выпуска рассматривают различные аспекты зоопарковской деятельности. Продолжены публикации по морской аквариумистике. Сотрудники Тульского экзотариума представили ряд статей, посвященных содержанию и разведению в неволе редких видов амфибий и рептилий. Рассмотрено влияние различных условий содержания на состояние и поведение копытных Московского зоопарка. В разделе «Проблемные статьи» затронуты вопросы асимметричного использования передних конечностей у домового опоссума, взаимодействия аборигенных видов мелких млекопитающих со сложной средой современного зоопарка, а также посещаемости Московского зоопарка. Статья, посвященная применению программного обеспечения при автоматизированном анализе человеческой речи для измерения основной частоты в звуках животных, публикуется на английском языке. Представлены обзоры по гнездованию журавлей в искусственно созданных условиях и истории охраны щетинохвостых кенгуру в Австралии. Теперь этот вымирающий вид содержится и в Московском зоопарке. Дан анализ структуры и проведена оценка успеха основных зарубежных программ по спасению редких видов млекопитающих. Две статьи посвящены слоновым прыгунчикам. В них рассмотрены вопросы систематики, экологии, поведения и содержания в неволе этих малоизученных животных.

Сборник рассчитан на специалистов зоопарков, зоотехников, зоологов и студентов биологических специальностей.

The current issue of Scientific Research in Zoological Parks is dedicated to different aspects of zoos activities. Traditionally we continue publishing articles on marine aquariumistics and husbandry and breeding of rare amphibians and reptyles. Ungulates' behaviour depending on different keeping conditions is discussed. Problematic articles consider asymmetrical forelimb-use in grey short-tailed opossums, interactions of aboriginal small mammals with complex zoo environment and attendance of Moscow zoo by different groups of visitors. Article devoted to application of software for automatic analysis of human speech to measuring the fundamental frequency in calls of nonhuman mammals is published in English. Special reviews discuss nesting of cranes in captive conditions, and conservation history of brush-tailed bettongs in Australia. Now this critically endangered species is represented in collection of Moscow zoo. Structure and rating of success of foreign conservation programms are analyzed. Two articles are devoted to elephant shrews describing different aspects of taxonomy, ecology, ethology and captive maintenance of these poorly studied animals.

The issue is destined for zoo specialists and keepers.

ББК 20.18:28.6

ISBN 978-5-904012-30-4

APPLICATION OF SOFTWARE FOR AUTOMATIC ANALYSIS OF HUMAN SPEECH TO MEASURING THE FUNDAMENTAL FREQUENCY IN CALLS OF A NONHUMAN MAMMAL

I.A. Volodin[1,2], I.S. Makarov[3], K.O. Efremova[2], E.V. Volodina[1], N.V. Soldatova[4]
[1] Scientific Research Department, Moscow Zoo;
[2] Faculty of Biology, Lomonosov Moscow State University;
[3] Joint Stock Company «Research-and-production association «EKRAN»;
[4] Ecocenter «Djeiran», Uzbekistan

For development of speech applications such as automatic speech/speaker recognition and speech synthesis, it is important to know how acoustic parameters vary as a function of age and gender. Thus, softwares for automatic speech analysis should be tested for a wide range of acoustic stimuli, to ensure that it performs correctly in all the range of variation, including voices of all age and sex classes as well as aroused or patient voices (Lee et al., 1999; Palethorpe et al., 1996; Rao, Kumaresan, 2000; Mustafa, Bruce, 2006; Sorokin, Makarov, 2008; Natour, Saleem, 2009). Advanced testing may involve calls of primates and ruminants with similar to humans voice acoustic patterns (Rendall et al., 2005; McComb et al., 2009; Gogoleva et al., 2011; Volodin et al., 2011), to estimate the robustness of the automatic tracing on the expanded range of acoustic stimuli.

At the same time, automatic recognition of animal species by vocalization has proved to be useful for conservation and for censuses (Sturtivant, Datta, 1997; Deecke, Janik, 2006; Selin et al., 2007). Further, the automatic measuring of acoustic parameters can be used for preliminary sorting of acoustic recordings in sound libraries (Baptista, Gaunt, 1997) and welfare monitoring of farmed animals (Schön et al., 2001, 2004; Jahns, 2008; Moura et al., 2008). For these reasons, acoustic software, primarily designed for analysis of human speech, is now widely applied for analyses of calls of nonhuman mammals (e.g., McComb et al., 2003; Owren, 2008; Rendall et al., 2009).

Potentially, software designed for automatic speech analysis should be appropriate for analysis of calls with acoustic characteristics close to those of human voice. For calls, differing somehow from the human voice pattern, the analysis settings should be adjusted accordingly. However, software with default characteristics established for human voice or even for one gender/age class of humans, allows adjustments only within a rather narrow range of variation, which restrains the applicability of the software designed for analysis of human speech from a wide circus of species and vocalizations with strongly distinctive acoustic patterns. The general focus of this study was to test the TF32 software (Milenkovic, 2001), designed originally for automatic analysis of human speech, to measuring the fundamental frequency of mammalian calls that are very close in their acoustic characteristics to those of human voice. In particular, (1) we examine nasal calls of goitred gazelles (*Gazella subgutturosa*) along ontogenesis from birth to 6 months of age for the possibility to automatically trace their fundamental frequencies with the TF32 software. (2) Concurrently, we analyze the same set of calls manually using Avisoft SasLab Pro software (Avisoft Bioacoustics, Berlin, Germany), designed for analysis of animal sounds, to receive independent control measurements of the fundamental frequency. (3) Finally, we compare the results of the automatic and manual analyses and determine the frames where the TF32 performs correctly.

METHODS

Site, Subjects, and Data Collection

Study subjects were 23 (10 male, 13 female) goitred gazelles, human-raised from birth onwards in breeding centre Ecocenter «Djeiran» (Uzbekistan, Bukhara region, 39°41'N, 64°35'E). The animals were individually dye-marked with p-phenylenediamine (Rhodia, Paris, France). We recorded calls between May 8 and August 28, 2008 from the 23 study subjects and between October 17 and October 28, 2008 from 18 (7 male and 11 female) of the 23 subjects, because 5 animals had been transferred from the Ecocenter to another place. We recorded calls before the morning or evening feeding (for keeping and recording details, see Efremova et al., 2011 and Volodin et al., 2011). For acoustic recordings (48 kHz, 16 bit), we used a Marantz PMD660 (D&M Professional, Kanagawa, Japan) and Zoom-H4 (Zoom Corp., Tokyo, Japan) digital recorders with Sennheiser K6-ME64 and Sennheiser K6-ME66 cardioid electret condenser microphones (Sennheiser electronic, Wedemark, Germany). In total, we collected 100 hours of audio recordings.

Call Sample and Analysis

We chronologically segmented each individual longitudinal set of acoustic recordings into 9 age classes: of 1–2, 3–4, 5–6, 7–8, 9–10, 11–12, 13–14, 15–16 and 23–24 weeks of age. Only calls produced through the nose (nasal calls) were selected for the current analysis, because only this call type occurred from birth to 6 mo of age (Efremova et al., 2011). For acoustic analyses, we selected calls not disrupted by wind and non-overlapped by noise or human voice. For each animal we randomly selected five calls per age class, however some animals provided fewer calls in some ages classes (mean ± SE = 4.70 ± 0.06 per animal per age). In total, we analyzed 931 nasal calls (418 from males and 513 from females).

Before analyzing, calls were downsampled from 48 to 16 kHz. The fundamental frequency period (i.e. the distance from the previous pulse to the following pulse) was measured from the screen with the standard marker cursor in the main window of Avisoft, showing the spectrogram and the waveform, with the following settings: Hamming window, FFT-length 512, frame 50%. Frequency resolution of the spectrographic analysis was 31 Hz, time resolution varied between 0.3–0.5 ms. Then we calculated the mean fundamental frequency of each call as the inversed value to the mean fundamental frequency period of the call *(Figure 1)*.

The algorithm used to compute fundamental frequency employs zero crossings and signal amplitude to make voice/unvoiced decisions on the original waveform. Then it performs linear predictive coding (LPC) inverse filtering prior to cross-correlation analysis to reduce formant artifact on fundamental frequency tracking. It uses an initial 8 ms window that correlates at lags ranging from 0.5 to 25 ms (fundamental frequency range is set to 40–1900 Hz). And on finding an initial fundamental frequency period, it uses a window that is adapted to the fundamental frequency period but limited to a minimum of 4 ms. The window is advanced with each fundamental frequency period, and the fundamental frequency trace changes on fundamental frequency period boundaries. Automatic fundamental frequency measurements with TF32 software has been conducted in the same sound files as manual measurements with Avisoft software. Point

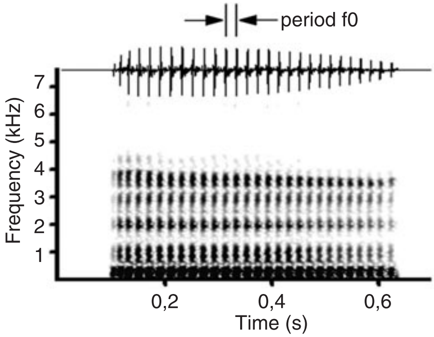

Figure 1. The nasal call of a 15–16 weeks-old male goitred gazelle: waveform (above) and spectrogram (below). Fundamental frequency period (period f0) was shown. The spectrogram was created with Hamming window; 16 kHz sampling rate; FFT 512 points; frame 50%; overlap 93.75%

fundamental frequency values were exported automatically to Matlab (MathWork, Natick, MA, USA) where the median fundamental frequency value has been calculated for each call.

Statistical analyses

All statistical analyses were made with STATISTICA, v. 6.0 (StatSoft, Tulsa, OK, USA), all tests were two-tailed, and differences were considered significant where $p < 0.05$. Since distributions of parameter values departed from normality, we log-transformed them to apply parametric tests. We used one-way repeated measures ANOVA to compare the manually and automatically measured values within each age class. We used χ^2 test to compare call proportions at different ages.

RESULTS

The manually measured fundamental frequency values showed a steady decrease from 2 to 24 weeks of age *(Figure 2)*. Unlikely, the automatically measured fundamental frequency values showed a steady decrease only up to 12 weeks of age, followed plateau up to 16 weeks of age and then abrupt increase to the age of 24 weeks. Repeated measures ANOVA showed that starting with 12 weeks of age, the automatically measured fundamental frequency values were significantly higher than the measured manually values *(Figure 2)*.

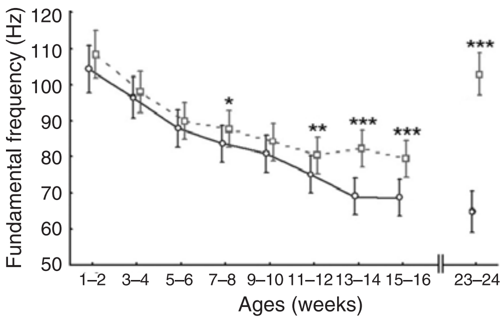

*Figure 2. Age-related changes in fundamental frequency (f0) of nasal calls in goitred gazelles, measured automatically with TF32 software (dashed line) and manually with Avisoft software (solid line). N = 23 animals (10 males, 13 females) at ages from 1–2 to 15–16 weeks and N = 18 animals (9 males, 11 females) at 23 – 24 weeks of age. Central points show means of means, whiskers show 0.95 confidence intervals. Significant differences between measurements: * – p < 0.05, ** – p < 0.01, *** – p < 0.001; repeated measures ANOVA*

Comparison of manual and automatic measurements of fundamental frequency in all the 931 calls showed that for 40.4% calls, the difference between the measurements did not exceed 1 Hz, for 61.5% – 2 Hz, and for 81.3% – 5 Hz. For the remaining 18.7% calls, the difference between manual and automatic measurements exceeded 5 Hz, and in most of these calls, the automatic values exceeded the manual ones *(Figure 2)*. Call percentages with incorrectly automatically taken fundamental frequency (with difference exceeding 5 Hz) ranged between 11.4–24.3% in all age classes from 2 to 16 weeks *(Figure 3)*, but to the age of 24 weeks increased significantly to as much as 43.8% ($\chi^2 = 29.55$, $df = 1$, $p < 0.001$).

DISCUSSION

In this study, the acoustic software designed primarily for automatic analysis of human speech, performed well also for analysis of animal calls. We found closely coinciding fundamental frequency values between the automatically taken and manually taken control measurements for most vocalizations produced by goitred gazelles along the period of their vocal development, when their vocal characteristics remained within the human speech sound range. Preceding ontogenetic studies of goitred gazelles (Efremova et al., 2011; Volodin et al., 2011) revealed very close vocal anatomy structures and the related acoustics to those of adult humans despite the considerable difference of body mass between juvenile gazelles and adult people. In

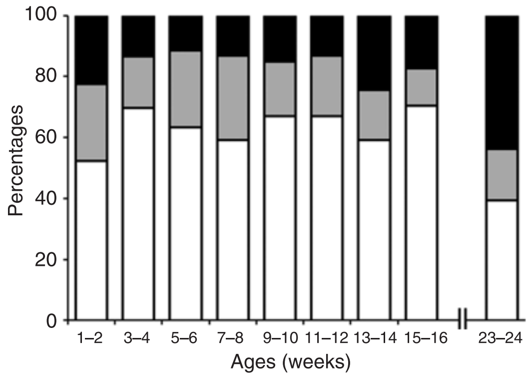

Figure 3. Call percentages of goitred gazelles at age classes from 1–2 to 23–24 weeks, where differences between fundamental frequency values taken automatically (with TF32 software) and manually (with Avisoft software) were less than 2 Hz (white bars), 2 to 5 Hz (grey bars) and exceeded 5 Hz (black bars)

juvenile male and female goitred gazelles, the lengths of vocal folds (15 and 13.5 mm respectively) were comparable to sizes reported for adult men (17–25 mm) and women (12.5–19 mm) (Titze, 1994; Roers et al., 2009).

Consistently, very low fundamental frequency values, found in nasal calls of goitred gazelles since the age of 1–2 weeks (mean ± SD = 104 ± 24 Hz), closely approach the fundamental frequency ranges of 107–129 Hz, reported for adult men and of 189–210 Hz, reported for adult women (Monsen, Engebretson, 1977; Lass, Brown, 1978; Rendall et al., 2005; Evans et al., 2006; Apicella et al., 2007). Therefore, in addition to existing primate models of voice production in humans (Owren et al., 1997; Rendall, 2003; Rendall et al., 2005), the similarity in vocal anatomy and acoustic features of growing goitred gazelles to those of humans makes the growing goitred gazelle a convenient model for application and testing acoustic software, initially developed for humans.

However, following the increase of body size and mass, the fundamental frequency of calls of goitred gazelles decreased from 104 ± 24 Hz at 1–2 weeks of age to as low as 65±13 Hz at 23–24 weeks. In older ages, the fundamental frequency values fall below the lower edge of the human speech fundamental frequency range, which resulted in a systematic overestimation of the automatically taken fundamental frequency values, that was especially prominent at the age of 23–24 weeks *(Figure 2)*. Nevertheless, for most calls, whose values were still within the human speech range of fundamental frequency, the error of automatic analysis was rather low *(Figure 3)*. Given that, we infer that the spectrographic software TF32 has proved to be useful tool for the automatic measurements of fundamental frequency in goitred gazelles and may be recommended for analysis of fundamental frequency in calls of other nonhuman mammals, whose fundamental frequencies coincide approximately with those of human speech. Changing the recommended for humans settings within TF32 should make automatic measuring more accurate in cases where fundamental frequency of nonhuman mammals decline strongly from those of speech.

ACKNOWLEDGEMENTS

We acknowledge the staff of the Ecocenter «Djeiran» for their help and support, and Gregory Francis for valuable comments and corrections. The study was supported by the Russian Foundation for Basic Research grant 12-04-00260.

REFERENCES

Apicella C.L., Feinberg D.R., Marlowe F.W. **Voice pitch predicts reproductive success in male hunter-gatherers** // Biology Letters. 3. 2007. P. 682–684.

Baptista L.F., Gaunt S.L.L. **Bioacoustics as a tool in conservation studies** // In: Clemmons J.R., Buchholtz R. (Eds.), Behavioral Approaches to Conservation in the Wild, Cambridge: Cambridge Univ. Press. 1997. P. 212–242.

Deecke V.B., Janik V.M. **Automated categorization of bioacoustic signals: avoiding perceptual pitfalls** // Journal of the Acoustical Society of America, 119. 2006. P. 645–653.

Efremova K.O., Volodin I.A., Volodina E.V., Frey R., Lapshina E.N., Soldatova N.V. **Developmental changes of nasal and oral calls in the goitred gazelle** Gazella subgutturosa**, a nonhuman mammal with a sexually dimorphic and descended larynx** // Naturwissenschaften, 98. 2011. P. 919–931.

Evans S., Neave N., Wakelin D. **Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice** // Biological Psychology, 72. 2006. P. 160–163.

Gogoleva S.S., Volodin I.A., Volodina E.V., Kharlamova A.V., Trut L.N. **Explosive vocal activity for attracting human attention is related to domestication in silver fox** // Behavioural Processes, 86. 2011. P. 216–221.

Jahns G. **Call recognition to identify cow conditions – A call-recogniser translating calls to text** // Computers and Electronics in Agriculture, 62. 2008. P. 54–58.

Lass N.J., Brown W.S. **Correlational study of speakers heights, weights, body surface areas and speaking fundamental frequencies** // Journal of the Acoustical Society of America, 63. 1978. P. 1218–1220.

Lee S., Potamianos A., Narayanan S. **Acoustics of children's speech: Developmental changes of temporal and spectral parameters** // Journal of the Acoustical Society of America, 105. 1999. P. 1455–1468.

McComb K., Reby D., Baker L., Moss C., Sayialel S. **Long-distance communication of acoustic cues to social identity in African elephants** // Animal Behaviour, 65. 2003. P. 317–329.

McComb K., Taylor A.M., Wilson C., Charlton B.D. **The cry embedded within the purr** // Current Biology, 19. 2009. R507–R508.

Milenkovic P. **TF32 [computer program]**. Madison: Univ. of Wisconsin-Madison. 2001.

Monsen R.B., Engebretson A.M. **Studies of variations in the male and female glottal wave** // Journal of the Acoustical Society of America, 62. 1977. P. 981–993.

Moura D.J., Silva W.T., Naas I.A., Tolón Y.A., Lima K.A.O., Vale M.M. **Real time computer stress monitoring of piglets using vocalization analysis** // Computers and Electronics in Agriculture, 64. 2008. P. 11–18.

Mustafa K., Bruce I.C. **Robust formant tracking for continuous speech with speaker variability** // IEEE Transact Audio, Speech, and Language Process, 14. 2006. P. 435–444.

Natour Y.S., Saleem A.F. **The performance of the time-frequency analysis software (TF32) in the acoustic analysis of the synthesized pathological voice** // Journal of Voice, 23. 2009. P. 414–424.

Owren M.J. **GSU Praat Tools: Scripts for modifying and analyzing sounds using Praat acoustics software** // Behavior Research Methods, 40. 2008. P. 822–829.

Owren M.J., Seyfarth R.M., Cheney D.L. **The acoustic features of vowel-like grunt calls in chacma baboons *(Papio cynocephalus ursinus)*: Implications for production processes and functions** // Journal of the Acoustical Society of America, 101. 1997. P. 2951–2963.

Palethorpe S., Wales R., Clark J.E., Senserrick T. **Vowel classification in children** // Journal of the Acoustical Society of America, 100. 1996. P. 3843–3851.

Rao A., Kumaresan R. **On decomposing speech into modulated components** // IEEE Transact Audio, Speech, and Lang. Process, 8. 2000. P. 240–254.

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

Rendall D., Kollias S., Ney C., Lloyd P. **Pitch (F0) and formant profiles of human vowels and vowel-like baboon grants: the role of vocalizer body size and voice-acoustic allometry** // Journal of the Acoustical Society of America, 117. 2005. P. 944–955.

Rendall D., Notman H., Owren M.J. **Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons** // Journal of the Acoustical Society of America, 125. 2009. P. 1792–1805.

Roers F., Murbe D., Sundberg J. **Predicted singers' vocal fold lengths and voice classification – a study of X-ray morphological measures** // Journal of Voice, 23. 2009. P. 408–413.

Schön P.C., Puppe B., Manteuffel G. **Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs *(Sus scrofa)*** // Journal of the Acoustical Society of America, 110. 2001. P. 1425–1431.

Schön P.C., Puppe B., Manteuffel G. **Automated recording of stress vocalisations as a tool to document impaired welfare in pigs** // Animal Welfare, 13. 2004. P. 105–110.

Selin A., Turunen J., Tanttu J.T. **Wavelets in recognition of bird sounds** // EURASIP Journal on Advances in Signal Processing, Article ID 51806, doi: 10.1155/2007/51806. 2007.

Sorokin V.N., Makarov I.S. **Gender recognition from vocal source** // Acoustical Physics, 54. 2008. P. 571–578.

Sturtivant C., Datta S. **Automatic dolphin whistle detection, extraction, encoding, and classification** // Proceedings of the Institute of Acoustics, 19. 1997. P. 259–266.

Titze I.R. **Principles of Voice Production**. Englewood Cliffs: Prentice Hall. 1994.

Volodin I.A., Lapshina E.N., Volodina E.V., Frey R., Soldatova N.V. **Nasal and oral calls in juvenile goitred gazelles *(Gazella subgutturosa)* and their potential to encode sex and identity** // Ethology, 117. 2011. P. 294–308.

РЕЗЮМЕ

И.А. Володин, И.С. Макаров, К.О. Ефремова, Е.В. Володина, Н.В. Солдатова. Применение программного обеспечения, предназначенного для автоматизированного анализа человеческой речи, для измерения основной частоты в звуках животных. В этом исследовании программное обеспечение TF32, предназначенное для автоматизированного анализа человеческой речи, было протестировано для измерения основной частоты в звуках детенышей копытных (джейранов *Gazella subgutturosa*), для того чтобы оценить возможности TF32 анализировать звуки, отклоняющиеся от нормальных характеристик человеческого голоса, и его потенциал как инструмента для автоматизированного анализа криков животных в целях сохранения видов, проведения учетов в природных популяциях, а также оценки благополучия животных по акустическим параметрам их звуков. Был проанализирован 931 носовой крик от 23 детенышей джейрана (10 самцов и 13 самок) на протяжении онтогенеза от 2 до 24 недель. Основную частоту одних и тех же криков анализировали автоматически с помощью TF32 и вручную. Для 81,3% звуков, в которых величина основной частоты звуков джейранов соответствовала диапазону основной частоты речи взрослых людей, различия между ручными и автоматическими измерениями были менее 5 Гц, для 61,5% звуков менее чем 2 Гц и для 40,4% звуков менее чем 1 Гц. Для оставшихся 18,7% звуков джейранов, которые были ниже по основной частоте, чем звуки речи взрослых людей, автоматическое измерение завышало основную частоту, так что различия между ручными и автоматическими измерениями превышали величину в 5 Гц. Эти результаты свидетельствуют о том, что корректные автоматические измерения основной частоты могут быть сделаны для звуков животных, диапазон основной частоты которых приблизительно соответствует диапазону основной частоты речи взрослых людей.