Analysis of Ultrasonic Components in Voices of Chosen Bird Species

Analysis of Ultrasonic Components in Voices of Chosen Bird Species

Conference Paper · June 2011

DOI: 10.13140/RG.2.1.1621.8720

4 authors, including:

Daniel Krol
Panstwowa Wyzsza Szkola Zawodowa w Ta…

Robert Wielgat
Państwowa Wyższa Szkoła Zawodowa w T…

T. Potempa
Panstwowa Wyzsza Szkola Zawodowa w Ta…

Analysis of Ultrasonic Components in Voices of Chosen Bird Species

Daniel Król, Robert Wielgat, Tomasz Potempa, Paweł Świętojański
Department of Technology, Higher State Vocational School in Tarnów

Summary

Abstract – Ultrasonic components of bird voices are rarely analyzed in bioacoustical papers. The main reason behind this situation is an equipment used for digital recording of bird voices in which: 1) sampling frequencies do not usually exceed 48 kHz and 2) microphones that limits bandwidth of recorded bird voices are also commonly used. The experiments described in this paper were carried out with 96 kHz sampling frequency and high quality broadband microphones. Moreover, because the commonly used 1-bit sigma-delta A/D converters introduce some distortion to the signal – they particularly affect its high frequency components. In the research we propose using the SAR converter instead. Preliminary automatic recognition research comparing performance of SAR and 1-bit sigma-delta A/D is also presented in the paper. Investigation concerned analysis of several voices of passerine birds e.g. *Parus major*, *Passer montanus* in ultrasonic range which unveiled existence of strong ultrasonic tonal frequency components and formants in the analyzed bird voices. Methodology that allowed us to record bird voices in an ultrasound range was described. Additionally, the general conclusion outlining the best bird voices recording strategy has been drawn. Biological meaning of the ultrasound components in bird voices still remains an open research question.

PACS no. 43.80.Ev, 43.80.Ka

1. Introduction

Most often encountered methodology of bird voices recordings is based on using digital recorders sampling analog signal with frequency up to 48 kHz and digitizing it with 1-bit sigma-delta A/D converter. Hipercardioidal and unidirectional microphones are commonly used. Some new concepts have been based on microphone arrays. Nevertheless all these recording methodologies does not allow proper signal acquisition containing ultrasound components. Solution presented in the paper is four-channel digital recorder with 96 kHz sampling rate and SAR (Successive Approximation Register) A/D converters. Signal is recorded with broadband microphones.

Certain justification of using SAR converters are research results from comparison of Successive Approximation Register (SAR) and 1-bit sigma-delta A/D converter which have been presented in [1], [2] and [3]. Superior performance of SAR analog-to-digital converters over the sigma-delta ones in audio measurement system has been shown. This paper presents some new research results for two types of analog-to-digital converters applied in bird voices recognition system.

Analysis of Ultrasonic components in bird voices is presented as well.

This paper is organised as follows: the second section contains general information about research methodology and presents some introductory assumptions. In the section three detailed dataset and experiments description are given. Fourth section is intended for results presentation which is directly followed by discussion and conclusions.

2. Methods

This chapter describes construction of digital audio recorder and its capabilities to record ultrasound components particularly these which are present in bird voices. Moreover comparison of SAR and sigma-delta A/D converters is also presented. There are also described feature extraction and classification methods used in the preliminary experiments comparing SAR and sigma-delta performance in bird voices recognition system.

2.1. Recorder

Digital audio recorder (Fig.1) was originally designed as a four channel sound recording device. A/D converters used in the recorder are 16 bits SAR converters. Device has been equipped with 32-bit DSP with ARM7 core which processes the signals from 4 A/D converters. Block diagram of the audio recorder is depicted on fig.2. Applied 96 kHz sampling frequency allows to record ultrasound from 20 kHz up to 48 kHz frequency range.

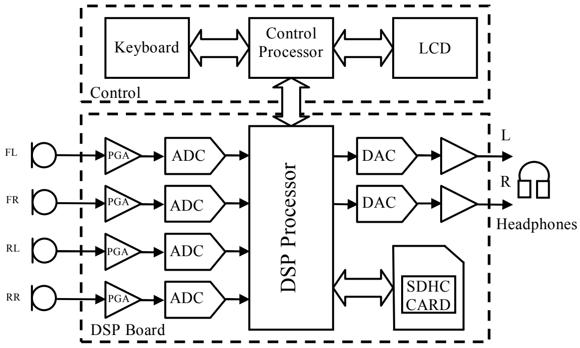

Figure 1. Block diagram of digital audio recorder

2.2. Microphones

Microphones used for bird voices recordings are usually microphones with less or more strong directivity pattern. Strong directivity pattern allows limiting noises from side signals but at the same time limits signal bandwidth and attenuates high frequencies. In the presented solution, four Panasonic Electret Capsules WM-61A which have linear frequency response in broad band have been used. Therefore these micropohones are capable to record ultrasound components without their significant attenuation. Using four microphones allows achieving more directional pattern by beam forming techniques, however these techniques were not the subject of the research described in the paper.

2.3. Comparison of SAR and sigma-delta A/D converters

Most voice recognition systems are based on sigma-delta A/D converters. Main disadvantage of SD-ADC technology is a ringing effect which is clearly visible in conversion of signals with fast rising and falling edges like square wave or in signals that contain usable information near Nyquist frequency e.g. ultrasounds. The sigma-delta ADC gives an average of oversampled bitstream [4], [5] and in consequence resolution for a fast rising signal is smaller than for low frequency signals. In case of 64-fold oversampling ADC there are 64 adds or subtracts. On the other hand volume of bird voices is diverse and distance-dependent. Influence of the ringing effect is relatively significant in case of silent bird voices.

In order to compare results of analog-to-digital conversion of two types of ADCs, a special audio recorder has been built. The hardware is based on ARM7 core processor LPC2468 and two A/D converters, PCM1804 with 1 bit sigma-delta modulator and AD7652 with Successive Approximation Register. Digital data from both A/D converters is streamed and writes as 16 bit stereo WAVE file on SDHC Card (Fig.1).

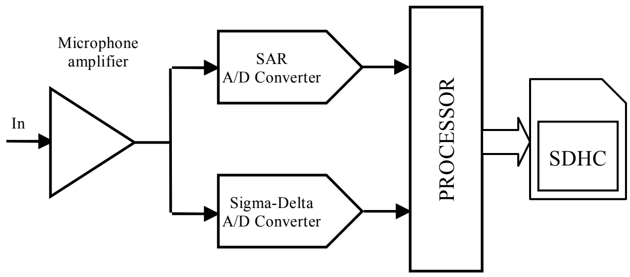

Figure 2. Block diagram of dual A/D converters audio recorder.

2.3.1. Sigma-delta A/D converter

The PCM1804 is a high-performance stereo A/D converter with full differential analog voltage input. The PCM1804 uses a precision delta-sigma modulator and includes a linear phase antialias digital filter as well as low-cut filter that removes DC (Direct Current) offset of the input signal. The PCM1804 is suitable for a wide variety of mid-to-high grade consumer and professional applications, where excellent performance are required. The PCM1804 can achieve both PCM audio and DSD format due to precision delta-sigma modulator. The input signal is sampled at x128, x64, and x32 oversampling rate for oversampling ratio. The single rate, dual rate, and quad rate eliminate the external sample-hold amp. Figure 2 illustrates how the PCM1804 for each oversampling ratio decimates the modulator output down to PCM data when the modulator is running at 6.144 MHz. The delta-sigma modulation randomizes the modulator outputs and reduces the idle tone level. The oversampled data stream from the delta-sigma modulator is converted to a 1 fS, 24-bit digital signal, while removing high-frequency noise components by a decimation filter. The DC components of the signal are removed by the low-pass filter.

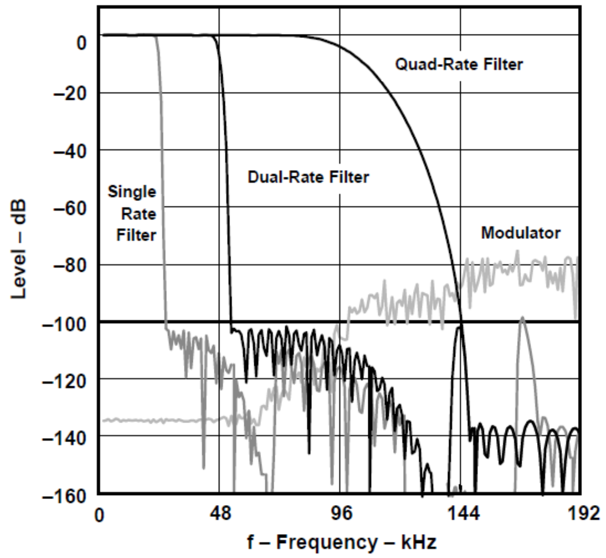

Figure 3. Spectrum of modulator output and decimation filter of PCM1804 [6]

2.3.2. Successive Approximation Register A/D converter

The AD7652 is a 16-bit, up to 500 kSPS, charge redistribution SAR analog-to-digital converter. The part contains a high speed 16-bit sampling A/D converter an internal conversion clock, internal reference and error correction circuits.

3. Used Algorithms

In order to examine influence of the high frequency ultrasonic noise of Sigma-Delta ADC several automatic signal recognition experiments comparing SAR and Sigma-Delta ADCs performance have been carried out. Recordings have been obtained from the circuit from Figure 1. It is clearly seen, that the only factor influencing recognition algorithms is used type of AD converter.

Algorithms used in the recognition experiments included basically hidden Markov models (HMM) and dynamic time warping (DTW). As the signal features commonly used robust Mel-frequency cepstral coefficients (MFCC) have been used.

Hidden Markov Models method (HMM) is presently the most popular and universal recognition strategy. Here whole word - hidden Markov models have been used. Basic concept of HMM assumes calculating overall probability of the model providing we have a signal. Before using HMMs for recognition training procedure has to be run first. A standard Baum-Welch procedure has been used in training HMMs for Bird voices Signal. As a result of training procedure a set of k HMMs is obtained were k is the number of species (types of signal) to be recognized. It is assumed that each HMM gives the highest probability for signals of bird species for which the HMM was trained. In order to simplify training procedure simple left-right model of the HMMs was accepted Fig. 4.

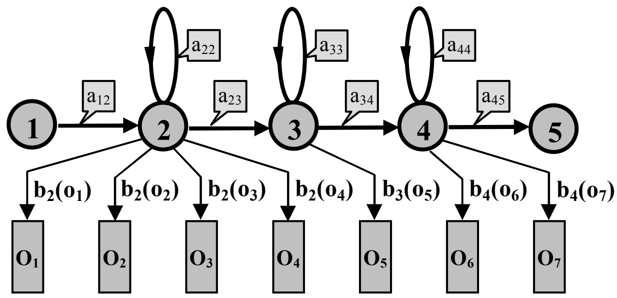

Figure 4. HMM of bird species voice used in experiments.

Number of states (circles on Figure 4) in model was different for different bird species. Model in figure 4 besides of topology is described by transition matrix:

$$A_T = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1}$$

Element a_{ij} of the matrix denotes probability of passing from state i to state j. The model is described also by observation probabilities described by Gaussina Mixture Densities:

$$b_j(o_t) = \sum_{m=1}^{M} c_{jm} N(o_t, \mu_{jm}, \Sigma_{jm}) \tag{2}$$

Where: j-state number, M – number of mixture components, c_{jm} – weight of m-th component, $N(o_t, \mu_{jm}, \Sigma_{jm})$ - is a multivariate Gaussian with mean vector μ and covariance matrix Σ given by:

$$N(o_t, \mu_{jm}, \Sigma_{jm}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-\frac{1}{2}(o_t - \mu_j)' \Sigma_j^{-1} (o_t - \mu_j)} \quad (3)$$

Where: n – dimensionality of the observation vector o_t.
Training procedure estimate parameter A_T, Σ_j, μ_j in order to maximize probability generating by model for the observation sequences from training set representing the class (bird voice type) the model is trained for.
While recognizing unknown bird voice type, all trained HMMs calculate probability for the observation sequence representing recognized bird voice type. The model giving the highest probability is chosen and the class (bird voice type) the model is trained for is assumed to be recognized class. At the recognition stage Viterbi algorithm is used.
Another recognition method with accuracy comparable to HMM classifier is dynamic time warping (DTW) method. The main idea in DTW algorithm used in speech recognition is finding an optimal path with minimal cost from lower left corner to upper right corner of the local distance array. A single element d_{mn} of the array equals to distance between m-th feature vector (MFCC) of recognized utterance and n-th feature vector of the reference pattern. An Euclidean distance was used as a distance measure in the reported research. An accumulated distance at each point of the search path was calculated according to recursive procedure given by the equation:

$$g(i,j) = d(i,j) + \min \begin{bmatrix} (g(i,j-1) \\ g(i-1,j-1) \\ g(i-1,j) \end{bmatrix} \quad (4)$$

In order to normalize obtained result the accumulated cost was divided by factor D:

$$D = \sqrt{N_P^{\,2} + N_T^{\,2}} \quad (5)$$

where: N_P – number of feature vectors of the reference pattern, N_S – number of feature vectors of test pattern.
The search path was limited by two parallel lines shifted by the coefficient Q=4. A detailed description of DTW algorithm can be found in literature.
Observation of feature vectors were calculated as mel-frequency cepstral coefficients (MFCC). Subsequent steps of the MFCC method are as follows:

1) blocking signal into frames and windowing by Hamming window;
2) application of the Fast Fourier Transform (FFT) to the windowed frames;
3) packing of the FFT power by Hamming window into the uniform, overlapping by 50% Mel frequency bands with equally spaced center Mel frequencies (number of bands is a parameter of the algorithm); conversion from linear-frequency scale to the Mel frequency scale is given by equation:

$$f_{mel} = 2095 \log_{10}(1 + f_{Hz} / 700) \quad (6)$$

4) calculation of log spectral power coefficients in mel bands;
5) performing DCT on log spectral power coefficients vectors.

4. Experiments and results

In order to evaluate performance of SAR and Sigma-Delta converters there were recorded above 5 hours of bird voices belonging to 5 common bird species: *Corvus frugilegus*, *Corvus monedula*, *Parus major*, *Passer montanus, Turdus merula*. Spectrogram and waveforms of the recordings have been manually analysed first. In case of silent voices the influence of ringing effect of sigma-delta A/D converter is clearly observed.

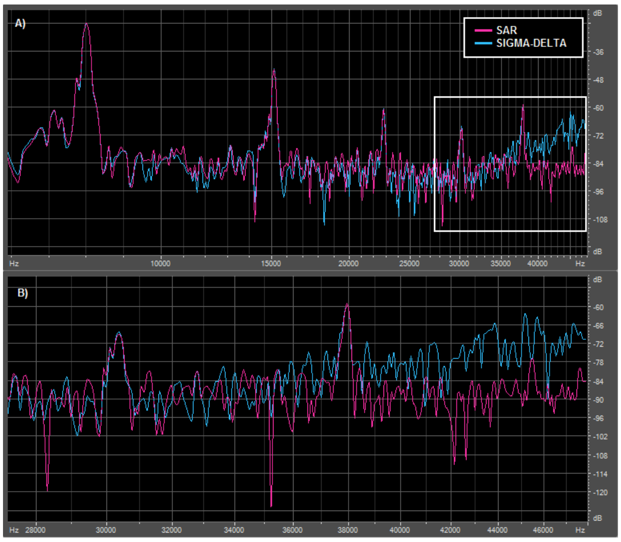

Figure 5. Comparing of frequency of bird voices by two types of A/D converters

Analysis revealed existence of significant ultrasound components in bird voices of *Parus major* and *Passer montanus* (Fig. 5 and Fig. 6). Certain signal types of these species indicated ultrasounds up to half sampling frequency (48 kHz) limit but the real limit is probably higher. These components of course can not be noticed using standard recording methodology with 48 kHz sampling rate. Moreover increase of high frequency noise above 35 kHz coming from Sigma-Delta ADC was also visible (Fig. 6.a). At the same time there was no increase of noise in case of SAR ADC (Fig. 6.b).

After this manual analysis preliminary experiment with bird voices recognition has been carried out. In order to accomplish the experiment a set of 153 signals representing bird voices has been extracted from recordings. Whole set has been divided into the training and testing sets. The structure of the sets has been shown in table 1.

As can be seen from table 1, training set and testing set are relatively small. Therefore experiments with 20 randomly chosen train and testing sets have been carried out in order to enhance reliability of the research.

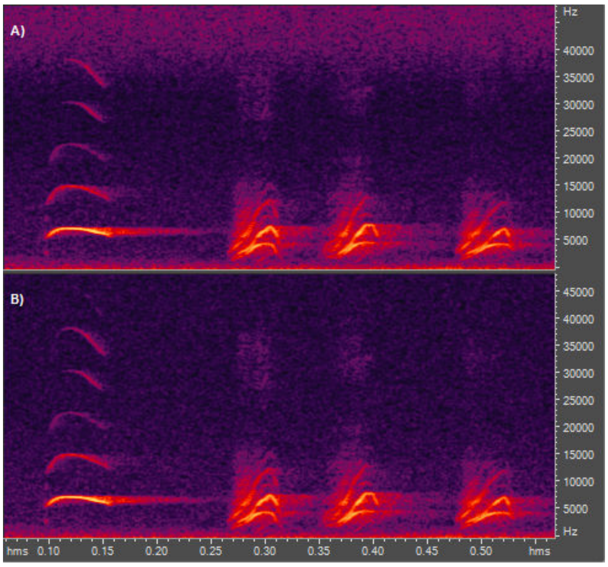

Figure 6. Comparing of spectral frequency of bird voices, a) sigma-delta A/D converter, b) SAR A/D converter.

Table 1. Specification of training and testing sets, and number of states for trained HMMs

Species	Signal type	Number of Examples		No of HMMs states
		Train. set	Test. set	
Corvus frugilegus	A	25	38	5
Corvus monedula	A	13	20	5
Parus major	I	5	9	6
Passer domesticus	C	5	8	9
Turdus merula	A	12	18	3

One group of experiments included unfiltered bird voices and the second one included bird voices filtered by high-pass 10-th order Butterworth filter with cutt-off frequency 1000 Hz. Due to high-pass filtration low-frequency background noise has been eliminated in order to examine more precisely influence of high frequency noise of sigma-delta converter on recognition accuracy.

Mel frequency cepstral coefficients (MFCC) have been chosen as signal features. Experiments included bird voices recognition with MFCC features. Frames of the length 15 ms and spacing 1 ms have been used. Relatively short frame times are motivated by short sounds emitted by birds and fast changes in bird voices. In order to achieve sufficient frequency resolution relatively big number of bands has been used. In the research 100 frequency bands and 20 (including 0th coefficient) MFCC coefficients have been accepted.

Table 2. Recognition accuracies for accomplished experiments. Abbreviations not explained in the text: Filtr. - filtration type, N –none, HP – high pass, Class. – classifier, Ex. No. – experiment number, Avg. accu. – average recognition accuracy, CF- *Corvus frugilegus*, CM – *Corvus monedula*, PMa – *Parus major*, PMo – *Passer montanus, TM – Turdus Merula*..

ADC	Filtr.	Class.	Accuracy [%]					
			Species					Avg. accu.
			CF	CM	PMa	PMo	TM	
SAR	N	HMM	100	95,25	100	15,18	56,11	**77,71**
		DTW	99,87	96,75	100	100	100	**99,32**
	HP	HMM	100	97,25	100	90,63	91,11	**95,80**
		DTW	100	96,75	100	100	100	**99,35**
SD	N	HMM	100	94,25	100	23,61	51,39	**75,06**
		DTW	100	97,25	100	99,38	100	**99,32**
	HP	HMM	100	97,25	100	92,50	89,17	**95,78**
		DTW	100	96,5	100	100	100	**99,30**

Table 3. Error rate for accomplished experiments. Abbreviations not explained in the text: Avg. error – average error rate., the rest of abbreviations like in Table 2.

ADC	Filtr.	Class.	Error rate [%]					
			Species					Avg. error
			CF	CM	PMa	PMo	TM	
SAR	N	HMM	0	4,75	0	85,63	43,89	**26,85**
		DTW	0,13	3,25	0	0	0	**0,68**
	HP	HMM	0	2,75	0	4,69	8,89	**3,10**
		DTW	0	3,25	0	0	0	**0,65**
SD	N	HMM	0	5,75	0	77,50	48,61	**26,37**
		DTW	0	0	0	2,75	0	**0,55**
	HP	HMM	0	2,75	0	4,81	10,83	**3,63**
		DTW	0	3,5	0	0	0	**0,7**

Experiments were carried out using two classifiers: hidden Markov models and Dynamic Time Warping. Accepted Hidden Markov models had

simple left-right structure (fig. 4). Number of states for every species and signal types are given in table 1. Mixture of 5 Gaussian probability density functions have been used per one state of the HMM.
Distances between feature vectors in DTW method were calculated as Euclidean distances. Tolerance of choosing start and end point was set to 4 feature vectors. Totally 8 experiments have been carried out. Results of experiments are presented in table 2 and table 3.

5. Disscussion

Analysis of the obtained results is not straightforward. It can be seen from Figure 6 that distinct ultrasound component are present in bird voices. This issue was so far omitted in majority of the research concerning bird voices analysis.
Another phenomena visible on spectrogram from Figure 6 is strong rise of power in high frequency ultrasonic noise above 35000 Hz. This high frequency noise causes also a little-bit worse recognition accuracy for sigma-delta converter. Probable explanation of this worse recognition accuracy is that feature vectors for different bird voices are more similar because of high frequency noise of SD converter than in case of SAR converter where the noises are not present.
Error rate of HMM without filtration is higher for SAR and lower for SD converter. HMM error rate for filtered signal is in turn lower for SAR converter and lower for SD converter.
Results of experiments for DTW method seems to be a little bit surprising because recognition accuracies are higher and error rates are lower in comparison with HMM method. Moreover influence of filtration seems to be accidential however in case fo SAR converter it caused slight improvement of recognition performance. In case of SD converter performance is a little bit worse.
Results of experiments give the chance that recognition by DTW method and HMM method with SAR converter will be more robust than using SD converter.
Looking at the results from table 2 there are also visible two somehow side effects of the research. The firs effect is higher recognition accuracy in case of DTW method in comparison with HMM one. The second effect is significant increase of recognition accuracy of HMM method after low pas filtration with 1000 Hz cut off frequency.
Results of the experiments of course are not statistically significant and should be verified on larger bird voice set including more bird species.

6. Conclusion

Experiments revealed that:
in order to properly analyse bird voices in range of near ultrasounds (20 kHz – 48 kHz) SAR AD converter with sampling 96 kHz should be used.
Using SAR converter instead of SD one in signal acquisition module gave slightly better results of bird voices recognition in the experiments described in the paper.
The best recognition accuracy has been achieved using SAR converter together with DTW method and low-pas filtration with cut off frequency 1000 Hz.
Results obtained in the research are not statistically significant but encourage to further research on larger sets of data.
-

Acknowledgements

The presented research was supported by the Polish Ministry of Science and Higher Education, grant no. N N519 402934.

References

[1] D. Król, "Choice of analog-to-digital converters for audio measurements using MLS algorithm", in Proc. EUSIPCO 2007, Poznań, Poland, September 3-7, 2007, pp. 2346–2349.

[2] D. Król, "Generowanie i akwizycja sygnałów testowych w metodach pomiarowych elementów toru audio" Measurement Automation and Monitoring, vol. 53 bis, pp. 209-212, Sep. 2007.

[3] D. Król, "On superiority of Successive Approximation Register over Sigma Delta AD converter in standard audio measurements using Maximum Length Sequences", in Proc. ICSES 2008, 14-17 September 2008, Kraków, Poland. pp. 51- 54.

[4] R.M. Gray "Quantization noise in DeltaSigma A/D converters", Chapter 2 of Delta-Sigma Data Converters, edited by S. Norsworthy, R. Schreier, G. Temes, IEEE Press, 1997, pp. 44-74.

[5] N. T. Thao, "Overview on a new approach to one-bit n-th order sigma-delta modulation", in Proc. ISCAS, 2001.

[7] Burr-Brown Products from Texas Instruments "Full Differential Analog Input 24-bit 192-kHz Stereo A/D Converter", Burr-Brown Data Sheet.

[8] Analog Devices "AD7552 16-Bit 500 kSPS PulSAR Unipolar ADC with Reference",Analog Devices Data Sheet