

Exploratory analysis on the binaural characteristics of the rowing sound

Pablo F. Hoffmann* Dorte Hammershøi* Emanuele Ruffaldi† Alessandro Fillipeschi†

(*) *Acoustics, Department of Electronic Systems, Aalborg University, Denmark*

(†) *PERCRO Lab, Scuola Superiore S. Anna, Pisa, Italy*

E-mail: pfh@es.aau.dk, dh@es.aau.dk, e.ruffaldi@sssup.it, a.filippeschi@sssup.it

Abstract

Differences between the signals captured at the entrances to the left and right ears are generally quantified by interaural time differences (ITDs), interaural level differences (ILDs) and interaural cross-correlation (IC). Here, we attempt to gain knowledge on the changes of these interaural parameters during rowing and to evaluate their potential as source of information to determine rowing expertise exclusively from acoustic signals.

1 Introduction

It has been suggested by rowing experts that hearing the boat is among the possible techniques for rowing [7]. The motivation of this work is to investigate the potential of rowing sound as a source of information to improve the analysis of rowing performance. This is part of the more general goal of exploring the possibilities to develop a machine-hearing system that can perform the following operations: (1) capture binaural sound during rowing, (2) segment the sound according to the stroke cycle, or according to the different stroke phases, (3) extract and analyze relevant acoustic features, (4) classify sounds according to level of expertise and provide an assessment of performance.

In the present paper we address points 1 to 3. Point 4 is evaluated in an accompanying paper [5]. Specific to point 1, binaural sounds were recorded on three rowers. Regarding point 2, a segmentation procedure based on onset detection is presented. With respect to point 3 we look for features in the patterns of binaural activity. That is, we examine changes in the two major binaural parameters, i.e., the interaural level difference (ILD), and interaural time difference (ITD) and relations between them. To this purpose, binaural parameters are computed from the binaural recordings using

two models of binaural hearing: the interaural cross-correlation model proposed by Lindemann [6] and extended by Gaik [4], and the model proposed by Faller and Merimaa [3] based on interaural coherence.

2 Capture of binaural sound

A wearable binaural recorder was developed that is capable of capturing high-quality audio without disrupting the hearing and performance of the user. This aspect was critical because it was not desired that the recording equipment had any effect on the actions being performed while rowing. Figure 1 shows a rower wearing the binaural recording with zooming to the relevant components. Miniature microphones (Knowles Acoustics FG23629) fixed to the ear holders of an adjustable headband (DPA 4088) were placed at the entrance to the left and right ear canals, and connected to a custom-built microphone amplifier placed on the back of the rower. The output of the amplifier was connected to the stereo input of a digital recorder (Edirol R-09) that was placed into a bag tightly fastened around the rower's waist. This was done to minimize the risk of the recorder falling into the water. All recordings were made with a sampling frequency of 48 kHz and stored in WAV format with a 16-bit resolution.

Binaural audio was recorded for three rowers: R1, R2, and R3. Rowers R1 and R2 had 5 years of experience and rower R3 had 11 years of experience. For each rower, recordings were made for several velocities spanning a range between 18 and 40 strokes per minute (SPM). During the recordings, rowers were required to call out the velocity every time before they started rowing. In this way, the sound portion corresponding to that specific velocity could be readily identified during the analysis.

Figure 2 shows a representative spectrogram of a binaural audio excerpt from R3 rowing at 20 SPM. The cyclic component of rowing is clearly visualized in the

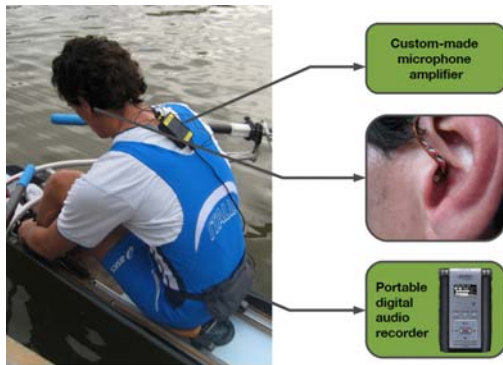


Figure 1: Illustration of a rower wearing the binaural recorder. The microphones are placed at the entrance to the ear canal.

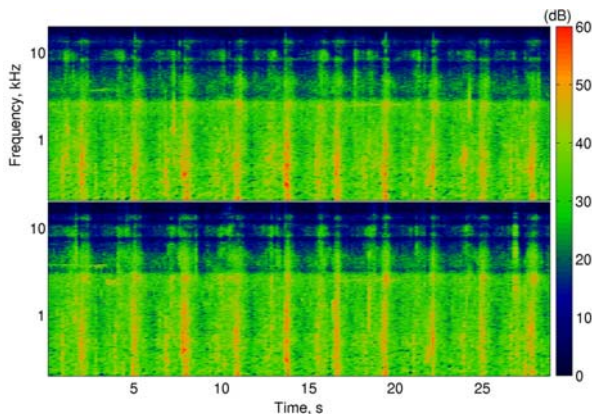


Figure 2: Spectrogram of binaural audio captured at the left (top panel) and right (bottom panel) ears during a 20-strokes-per-minute session.

spectrogram by the periodic increase in acoustic energy and bandwidth that occurs roughly every 3 seconds. Each of the periods represents a rowing stroke. According to the international rowing federation, a rowing stroke is divided into four phases: catch or entry, drive, finish and recovery. The time instance of the energy in the spectrogram corresponds to the finish phase, and, as further discussed below, served to identify the boundaries between stroke cycles in the audio segmentation procedure.

3 Audio Segmentation

The energy increment by about 20-30 dB over a frequency range of 200–8000 Hz that is observed roughly every 3 s in the spectrogram of Figure 2 is the sound event corresponding to the finish phase. From a purely acoustic perspective the finish phase appears to be the only reliable sound event for the segmentation of rowing sound based exclusively on acoustic information.

Audio was segmented in a semi-automatic way, namely, by visually inspecting spectrograms together with an onset detection algorithm. A suitable choice for the segmentation of the rowing sound is the detection of the onset times of the finish phase. Many onset detection algorithms have been proposed for music signals [2]. Among these algorithms, a simple and general approach is based on tracking changes in the spectrum of the sound by computing the difference between successive short-term Fourier spectra on a frame-based processing. This difference is typically referred to as the *spectral flux*, and can be expressed by the equation:

$$SF(n) = \sum_{k=0}^{K-1} H(|X(n, k)| - |X(n-1, k)|) \quad (1)$$

where k and n are the frequency and time indexes respectively, X is the short-term Fourier signal, and $H(x) = (x + |x|)/2$ is the half-way rectifier function used to count only those frequencies where there is an increase in energy, thus keeping onsets and disregarding offsets. The frame-based processing was performed as follows. A Hamming window of 30-ms duration (1440 samples) was employed and a window shift equal to 10 ms was used (480 samples), meaning an analysis rate of 100 Hz. For each frame a Fast Fourier Transform was applied using a size of 2048 points (frames were zero-padded). The resulting spectral flux signal was smoothed using a Savitsky-Golay filter (2nd order polynomial degree) and a sliding window of 32 points. Signals were normalized to have zero mean and a standard deviation equal to 1. This process was applied to the left and right audio signals separately and thresholds were computed as the root-mean square of the onset signals. Figure 3 shows an example of the computed onset function $f(n)$ with the selected onsets corresponding to the peaks of the function. The selection of a peak in $f(n)$ was based on the following criteria:

$$f(n) > f(k) \text{ for all } k \text{ such that } n - w \leq k \leq n + w$$

where $w = 3$ is the size of the window used to find a local maximum. As a first approximation to evaluate the accuracy of the segmentation procedure an autocorrelation of the onset function was computed. Figure 4 shows an autocorrelation function along with an indication of the time location of the first peak and how this peak relates to the rowing speed. The time index corresponding to the first peak of the autocorrelation function multiplied by the size of the window shift in milliseconds can provide an estimate of the speed. Since the speed of the rower was known in advance it is possible to compare the estimated speed with the real one and therefore evaluate the quality of the segmentation procedure. The smaller the error between the known speed

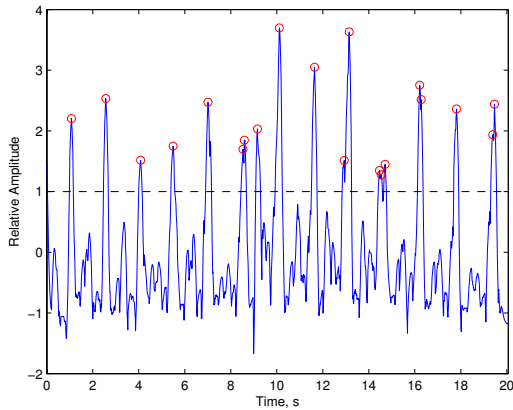


Figure 3: Onset function obtained for a rowing sound with a speed of 40 SPMs. Detected peaks are shown by circles. The dashed line indicates the selection threshold (see text).

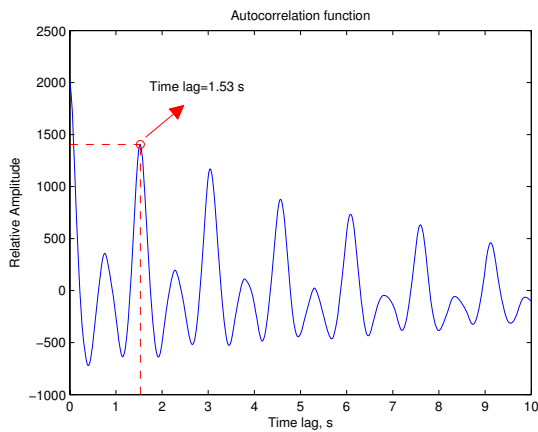


Figure 4: Autocorrelation of the onset function shown in Figure 3. The time lag of the first (actually second) peak is shown to be 1.53 s.

and the estimate the better the segmentation procedure. For the particular example in Figure 4 the index of the first peak is 153, which, multiplied by 10 ms (window shift) results in a time lag of 1.53 s. This value would correspond to a speed of about 39 SPM. The expected velocity was 40 SPM with an associated stroke duration of 1.5 s, and thus an error of about 2% was the result of the speed estimation. This error margin was acceptable considering that the procedure was complemented with visual inspections of the spectrograms. In this work the segmentation procedure was applied to binaural signals for only three velocities. Table 1 provides a summary of the number of segments obtained for each rower and velocity.

4 Binaural activity pattern

The Lindemann-Gaik binaural model [6, 4] implemented in the Auditory Modeling Toolbox (AM-Toolbox) [1] was used to compute binaural patterns. This binaural model is based on a running interaural cross correlation process for the estimation of interaural time difference (ITD), and a contralateral-inhibition mechanism responsible for processing interaural level differences (ILD). This mechanism sharpens the peaks in the cross-correlation helping in resolving ambiguities caused when two sources have similar ITDs. The model proposed by Faller and Merimaa computes ILDs and ITDs based on the interaural coherence (IC). Both models consisted of similar processing stages before the binaural processor: cochlear filtering by a filter bank, and inner hair cell transduction by half-wave rectification and low-pass filtering. The output of the Lindemann-Gaik model is a three-dimensional matrix with time, time lag (Tau), and frequency channel as their dimensions. The output of the Faller-Merimaa model is a set of ILDs, ITDs, and ICs as a function of time.

Figures 5(a)–(c) show computed binaural patterns for the three velocities: 20, 31, and 40 SPM. Each panel displays a 2D representation computed by integrating the cross-correlation output across frequency channel (averaged across audio segments). Darker areas represent higher interaural correlations. The dark region observed at about 1 s in most panels corresponds to the finish phase, and indicates that the sound event associated to the finish phase appears to be formed by a distributed sound source. From a qualitative viewpoint, there are no obvious or systematic differences between the binaural patterns across rowers or velocities. For instance, distinctive differences between rower R3 (the most experienced) and rowers R1 and R2 could have served as features to differentiate performance. In spite of the lack of qualitative differences, common patterns across rowers and velocities can also be advantageous. For example, it is interesting to note that if a parametric representation of the rowing sound

Table 1: *Number of audio segments per rower and velocity.*

Rower ID	Velocity (SPM)		
	20	31	40
R1	22	28	18
R2	19	30	20
R3	22	21	13

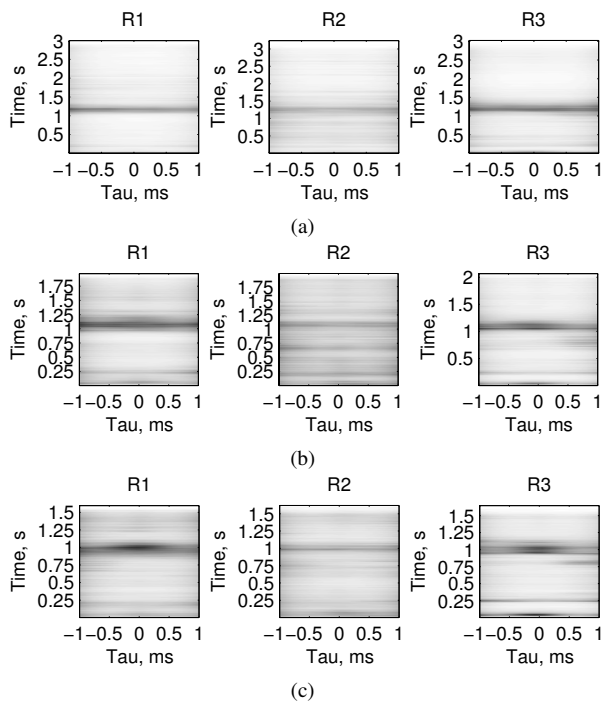


Figure 5: Binaural activity patterns of the rowing stroke for velocities (a) 20 SPM, (b) 31 SPM, and (c) 40 SPM. Each column represents data from a single rower. Gray shades indicate amount of interaural correlation with darker shades indicating higher correlations. The time lag (τ) is an estimation of ITD.

is desired for rendering purposes, besides modeling of its spectro-temporal characteristics, the output of the binaural model suggests that the sound produced during the finish phase may be modeled as a distributed sound source.

In addition, the probability density functions of ITD and ILD computed using the Faller-Merimaa model (Figure 6), suggest that for signals highly coherent the spatial auditory scene differs across frequencies. For an auditory filter centered at 500 Hz the sound appears to be slightly extended from the frontal direction, whereas for an auditory filter centered at 2000 Hz more lateral sources are observed as shown by the additional activity in more lateral directions along the ITD axis (τ) in the right panel of Figure 6. Again, the same pattern is observed across rowers and velocities.

In summary, qualitative observations of the changes of interaural parameters as a function of expertise suggest that they do not seem to differ considerable across different rowers. Rather, interaural parameters showed comparable trends in their changes across rowers and velocities. It is therefore suggested to further

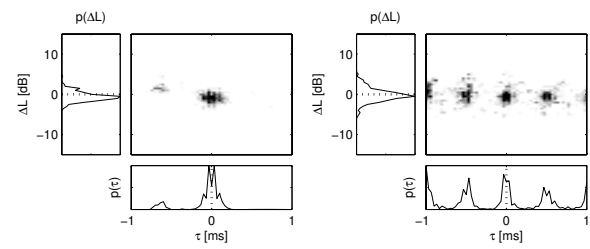


Figure 6: Probability density functions of ITD and ILD for the rowing sound. Data are shown for rower R3 and 40 SPM. Left, data for auditory filter centered at 500 Hz; Right, data auditory filter centered at 2000 Hz.

investigate their use in the design of a generic parametric representation for the rendering of rowing sound.

This work was supported by the SKILLS Integrated Project (IST-FP6 #035005, <http://www.skills-ip.eu>) funded by the European Commission.

References

- [1] Auditory Modeling Toolbox. <http://amtoolbox.sourceforge.net/>, 2011.
- [2] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. A tutorial on onset detection in music signals. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 13(5):1035–1047, 2005.
- [3] C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *J. Acoust. Soc. Am.*, 116(5):3075–3089, 2004.
- [4] W. Gaik. Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. *J. Acoust. Soc. Am.*, 94(1):98–110, 1993.
- [5] L. Johard, E. Ruffaldi, P. Hoffmann, and A. Filipposchi. Machine learning analysis of binaural rowing sounds. In *International Conference of the European SKILLS Project*, Montpellier, France, 15–16 December 2011.
- [6] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.*, 80(6):1608–1622, 1986.
- [7] V. Lippens. *Rowing faster*, chapter Inside the rower’s mind, pages 185–195. Human Kinetics Publishers, Inc, Champaign, IL, USA, 2005.