EVALUATING A WAVELET-BASED ANALYSIS OF SENSOR REED SIGNALS FOR PERFORMANCE RESEARCH

Alex Hofmann
Institute of Music Acoustics (IWK),
University of Music
and Performing Arts Vienna,
Austria
hofmann-alex@mdw.ac.at

Werner Goebl
Institute of Music Acoustics (IWK),
University of Music
and Performing Arts Vienna,
Austria
goebl@mdw.ac.at

Michael Weilguni
Institute of Sensor and Actuator Systems,
Vienna University
of Technology,
Austria
michael.weilguni@tuwien.ac.at

ABSTRACT

Empirical investigations of saxophone and clarinet performance are important for a thorough understanding of the human motor skills required to musically operate woodwind instruments. In this paper, we discuss two methods of detecting tonguing related landmarks in a sensor saxophone reed signal. We give detail about the reed signal's characteristics under three typical playing instructions (legato, portato and staccato articulation) and define detection tasks for physical tone onsets and tone offsets. When the player's tongue contacts the reed, the oscillations are dampened and the reed is bent towards the mouthpiece (tongue-reed contact). Removing the tongue from the reed returns it to its equilibrium position (tongue-reed release). From these observations we derive two landmark detection functions: a heuristic for peak detection, based on thresholding the smoothed reed signal, and a wavelet-based analysis, operating on specific sub-bands of the reed signal. When evaluating both methods, the wavelet analysis revealed better results using our current test dataset.

1. INTRODUCTION

To investigate how humans interact with musical instruments during skilled musical performance, empirical studies require the analysis of large data sets [1]. The raw data captured during studies of music performance may include audio recordings [2], video capture [3] or signals retrieved from sensors attached to musical instruments [4, 5].

For the analysis of specific sensor signals, such as our sensor saxophone reed [6], landmark detection functions (LDF) have to be developed for each new measurement setup.

For instruments that do not provide symbolic data (i.e., MIDI), state-of-the-art algorithms from music information retrieval can be used to roughly transcribe performances from audio material [7]. However, these algorithms were developed to track perceptual note onset times and may not define physical note onsets and offsets with the precision that is required for performance analysis.

Specifically developed detection functions demand careful evaluation of the results to ensure their reliability when making statements about the underlying performance. In this paper we describe two different approaches of detecting tongue-reed interaction in sensor saxophone reed signals: A peak detection function, based on thresholding the smoothed reed signal, and a wavelet-based analysis. Finally we evaluate the analysis results of both methods under different levels of stringency.

Articulation on single-reed woodwind instruments

When playing on single-reed woodwind instruments, the sound of consecutive played tones depends on the articulation technique used [8–10]. In contrast to legato playing, when the player blows constantly and only varies the fingerings, portato articulation and staccato articulation require tongue–reed interaction in order to control note beginnings and note endings.

A reliable, automated detection of these physical note onsets and note offsets from the sensor reed signal is essential, to investigate the motor control mechanisms in expressive woodwind performance.

2. METHOD

Sensor Reed Signal Properties

Strain-gauge sensors attached to woodwind single-reeds have been shown to capture the bending of the reed during performance. These signals contain precise information about tongue–reed interactions without environmental interference [6]. We observed different reed signals for different articulation techniques (see Figure 1). Whereas in legato playing (top panel) no tongue-reed contact (TRC) occurred, tonguing in portato articulation and staccato articulation was clearly visible (middle and bottom panel). During TRC the tongue pressed the reed towards the mouthpiece and thereby dampened the reed vibrations. To start the next tone, the tongue released the reed (tongue-reed release, TRR) [10].

2.1 Signal smoothing method

In our previous study on finger and tongue coordination in saxophone performance, we analysed the sensor reed

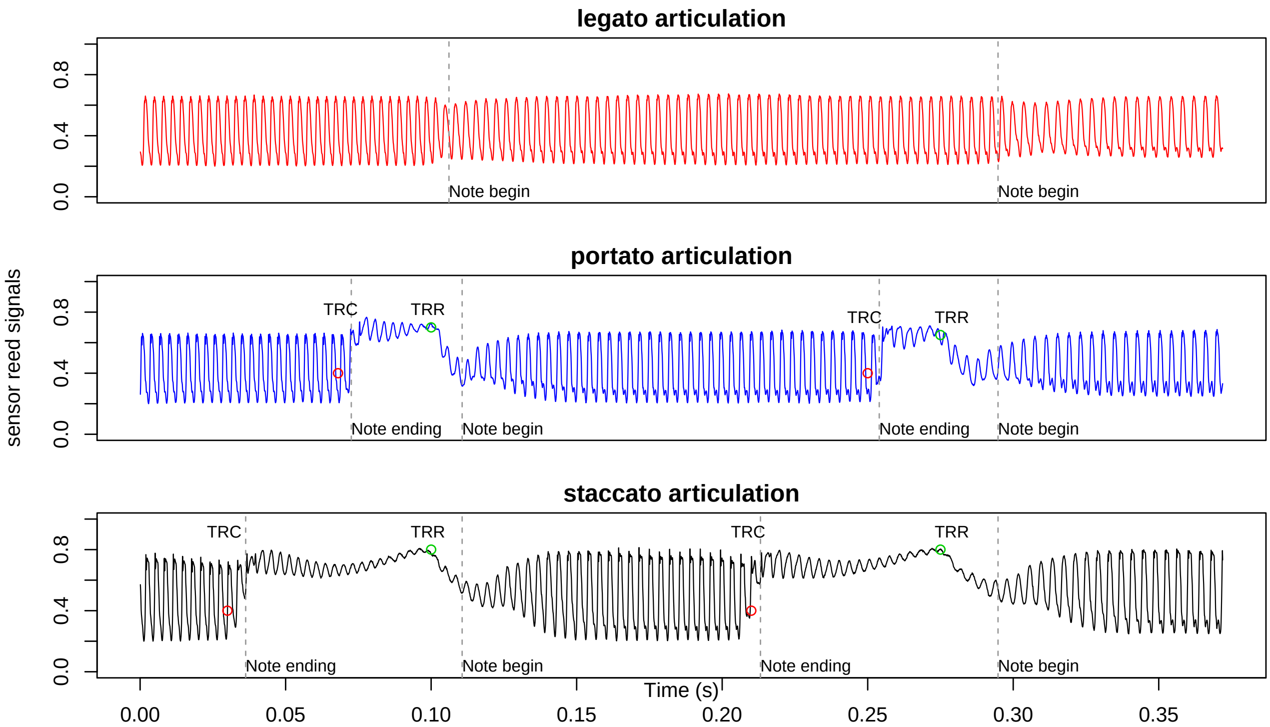

Figure 1. Alto-saxophone sensor reed signals showing the note transitions (e4–d4–c4) under a tempo instruction of 179 ms inter-onset interval (audio sampling rate 11 kHz). Legato articulation without tonguing (top); portato articulation with tongued note onsets (tongue reed release, TRR) and note offsets (tongue reed contact, TRC)(middle); staccato articulation, with extended tongue reed contact time (bottom)

signal with a rather simple approach, following two steps [11]:

Preprocessing

A smoothing function (Butterworth, low pass filter, 15 Hz) removed the high frequency oscillations, which occurred during sound production. Artefacts of the filter (i.e., phase shift) were ignored.

Landmark detection function

Local maxima above a certain threshold (40 % quantile of all maximum levels) were marked as potential TRRs and minima were marked as TRCs. Saddle points (pairs located on the same level of the smoothed signal) were automatically removed. For the calculations, external libraries (signal, msProcess) were used in the R-statistics software package [12].

The high error rate (multiple detections of one event) required a manual second step, in which the landmarks were verified by visual inspection before further processing was allowed. This time consuming procedure was possible for a relatively small data set containing 8700 tones, but is not a suitable method for further empirical studies with larger datasets.

2.2 Wavelet-based method

In the following section, we will discuss a LDF based on wavelet decomposition of the reed signal. A multiresolution analysis (MRA) is considered to be a suitable tool for time critical signal analyses (i.e., drum pattern transcription in audio material) [13–15]. The computational efficient pyramid algorithm calculates the wavelet coefficients, which represent the energy distribution over time and frequency, with $O\left(N \log_2 N\right)$ [16]. The external libraries (wmtsa, msProcess) were used in the R-statistics software package [12].

Preprocessing

The reed signal was decomposed into eleven sub-bands using the Maximal Overlap Discrete Wavelet Transform (MODWT, $J_0 = 11$). The advantage of the MODWT over a Discrete Wavelet Transform (DWT) is that the MODWT is well defined for any sample size and the resulting sub-bands (details $\widetilde{D}_J$ and smooth $\widetilde{S}_{J_0}$) are associated with zero phase filters. The Daubechies least asymmetric 8-tap filter LA(8) was chosen, because its phase properties allow direct reference from the MODWT coefficients to actual times in the original signal [16].

The choice of level $J_0 = 11$ allows investigations of relevant sub-bands, starting from the smooth with a horizontal spacing between individual coefficients with $\widetilde{S}_{J_0}$: $\lambda_{11} \triangle t = 2^{11} \cdot \frac{1000\, ms}{11025\, Hz} = 185.76\, ms\ (5.38\, Hz)$. Large

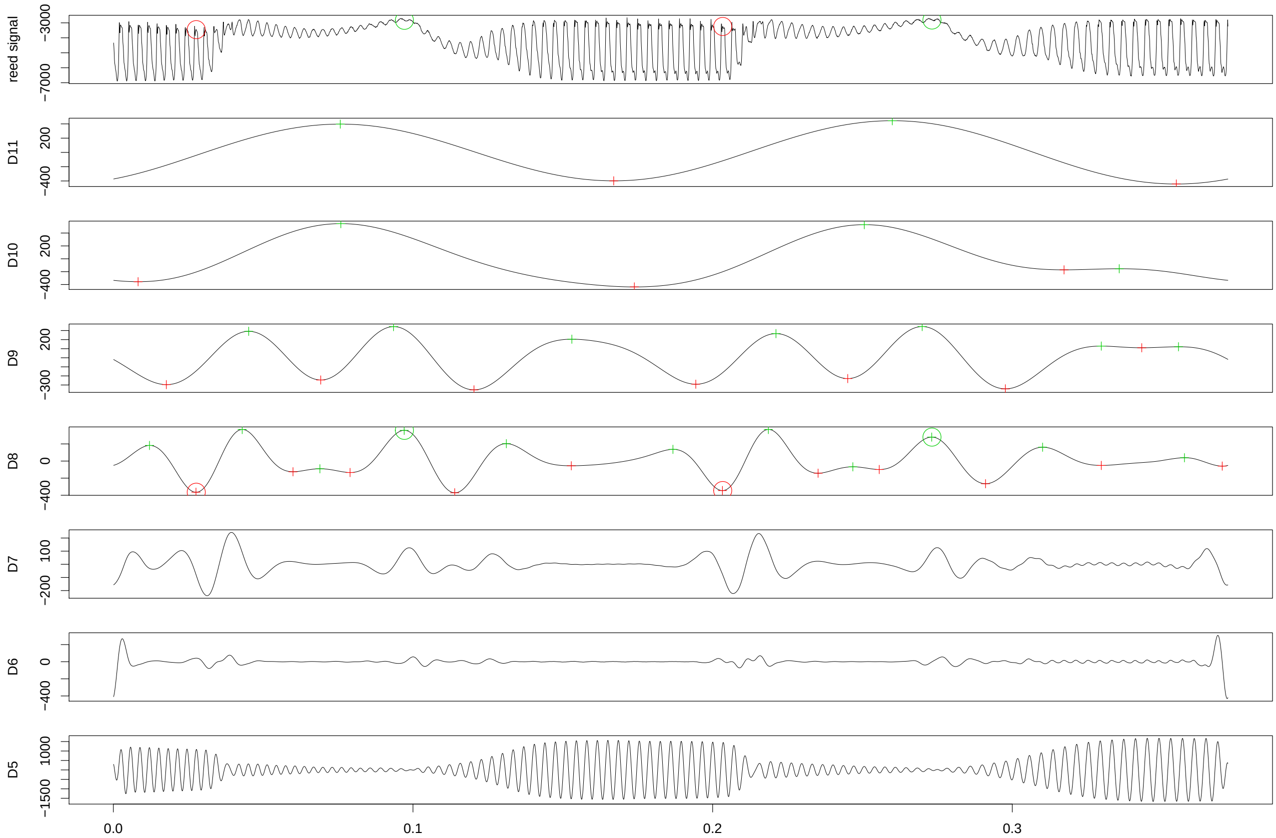

Figure 2. Maximal Overlap Discret Wavelet Transform of sensor reed signal showing staccato articulation: The figure shows the input signal (top) with detected landmarks (TRC: red circle, TRR: green circle) and the details $\widetilde{D}_{11-5}$ (below). The D11–D8 landmark detection function labeled maxima (green) and minima (red) in detail $\widetilde{D}_{11}$. These positions were refined to extrema of $\widetilde{D}_{10}$, $\widetilde{D}_9$ and $\widetilde{D}_8$.

scale fluctuations, caused by temperature effects to the strain gauge [17], were thereby isolated in $\widetilde{S}_{J_0}$ and do not further affect the analysis of the details $\widetilde{D}_J$.

The horizontal time spacing of the multiresolution analysis (MRA) sub-bands can be calculated, by seeing the MODWT as a constant-Q filter bank with octave spaced centres of the filters. Each higher sub-band contains double the time resolution as the previous sub-band ($\widetilde{D}_{11}$:$\tau_{11}\triangle t = 92.88\,ms$; $\widetilde{D}_{10}$: $\tau_{10}\triangle t = 46.44\,ms$; $\widetilde{D}_9$:$\tau_9\triangle t = 23.22\,ms$; $\widetilde{D}_8$: $\tau_8\triangle t = 11.61\,ms$).

Landmark detection function

$\widetilde{D}_{11}$ was used to detect the reed displacement caused by the tongue. Maxima of $\widetilde{D}_{11}$ (maximum displacement of the reed towards the mouthpiece) were marked as TRR, because the following signal decrease is an indicator that the player released the tongue. As a logical consequence, a TRC must have happened before a TRR. Consequently, minima were labelled as TRC and maxima as TRR. To precise the position, these labels were shifted to the extrema in sub-bands with a better time resolution ($\widetilde{D}_{10}$, $\widetilde{D}_9$ and $\widetilde{D}_8$: $\tau_8\triangle t = 11.61\,ms$). This LDF will be abbreviated as D11–D8.

Figure 2 depicts the wavelet-based landmark detection for staccato tone transitions: First, maxima and minima of $\widetilde{D}_{11}$ were labelled. These rough landmarks were then refined to extrema of $\widetilde{D}_{10}$, $\widetilde{D}_9$ and afterwards to those of $\widetilde{D}_8$.

3. EVALUATION

3.1 Dataset

Our test dataset contained 1744 visually annotated landmarks, based on sensor reed signals similar to the material from our previous study [11]. For this evaluation, we used eighth-note melodies, played with portato and staccato articulation, in three tempo conditions (IOI = $250\,ms$, 178.6 ms, $144.2\,ms$), performed by six alto-saxophone players in our laboratory.

3.2 Measures

To compare both LDFs, the standard measures *precision*, *recall* and *F-measure* were used. *Recall* describes the completeness of the search and *precision* gives status about the quality of the search results. *F-measure* combines the two previous measures by the following equation:

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

Starting from the annotated ground truth, the existence and number of detected landmarks around the annotated

Landmark Det. Func. WindowSize	% F-meas. ±25 ms	% Prec.	% Rec.	% F-m. ±15 ms	% Prec.	% Rec.	% F-m. ±10 ms	% Prec.	% Rec.
TRC–Sig. smoothing	78.7	74.5	83.3	43.3	41.1	45.9	3.3	3.1	3.4
TRR–Sig. smoothing	85.5	81.0	90.5	72.3	68.5	76.5	33.3	31.5	35.2
TRC–Wavelet analysis	95.2	95.5	94.8	90.1	90.5	89.8	55.0	55.2	54.8
TRR–Wavelet analysis	95.4	95.7	95.1	76.6	76.9	76.4	51.2	51.4	51.0

Table 1. *F-measure*, *precision* and *recall* for both landmark detection functions proposed in Section 2. Results for both tasks: TRC (tongue-reed contact, note offset) and TRR (tongue-reed release, note onset) detection within a $\pm 25\,ms$, $\pm 15\,ms$ and $\pm 10\,ms$ evaluation window are given and explained in Section 4.

events was checked. A single detection was counted as one true positive, whereas double detections were considered as one true positive and one false positive. A missing landmark was counted as one false negative. Remaining landmarks, not matched to annotated events, were counted as false positives. The strictness of such an evaluation is defined by the size of the evaluation window. Usually a $\pm 25\,ms$ evaluation window is used to check onset-detection functions working on percussive material [7]. For articulation detection, a higher accuracy of the detected events might be necessary, so $\pm 15\,ms$ and $\pm 10\,ms$ evaluation windows were additionally considered.

4. RESULTS

Comparison of both LDFs

Table 1 shows the results for the two LDFs described in Section 2. For the smoothing method (see Section 2.1.), only the automated detection of potential landmark positions was applied to the dataset, without manual correction steps.

Overall, the wavelet-based analysis gave better results in the *F-measure* of both detection tasks ($> 95\%$ within a $\pm 25\,ms$ evaluation window). The wavelet-based analysis also gave better results for precision ($> 95\%$ within a $\pm 25\,ms$ evaluation window) than the smoothing method ($> 74\%$). False positive detections, a main problem of the smoothing method, were reduced by using the wavelet LDF.

The choice of the evaluation window had a significant influence on the quality measure. Within the $\pm 25\,ms$ evaluation window both LDFs performed quite well (Smoothing LDF *F-measure*: $> 78\%$; Wavelet LDF *F-measure*: $> 95\%$). A reduction of the evaluation window to $\pm 15\,ms$ showed a significant difference in the behaviour of both methods. The wavelet LDF showed a better score for TRC detection (*F-measure*: 90.1%) compared to the smoothing LDF (*F-measure*: 43.3%). Both LDFs showed similar results for TRR detection (Smoothing LDF *F-measure*: 72.3%; Wavelet LDF *F-measure*: 76.6%). A further reduction of the evaluation window to $\pm 10\,ms$ showed clear superiority of the wavelet approach's accuracy, especially in TRC detection (Smoothing LDF *F-measure*: 3.3%; Wavelet LDF *F-measure*: 55.0%).

5. DISCUSSION

We developed two different methods to extract physical note onsets (and note offsets) from sensor saxophone reed signals, with the aim of enabling automated examination of large datasets of woodwind performances.

We compared the reliability of both proposed detection functions under different parameters of detection accuracy and found that the wavelet-based LDF outperformed the signal smoothing method.

Direct comparisons of our detection results with state-of-the-art onset detectors are not possible because these algorithms were evaluated on larger testset, containing various types of musical recordings from different genres and environments (*F-measure* between 70% and 87% [7, 18]).

The archived *F-measure* of $> 90\%$ (for TRC detection with wavelets, within a $\pm 15\,ms$ evaluation window) meets our criteria and will be used in future studies to analyse saxophone and clarinet performances' sensor reed signals. A detailed examination of the reasons for the decreasing *F-measure* of 76.6% for the TRR detection task is intended.

Finding the right balance between a sufficiently powerful, but not over-fitted detection function still remains a difficult task. The current wavelet LDF (D11–D8) is designed to accommodate different playing speeds and articulation techniques, but may be limited to sensor reed signals within a certain amplitude range. To further investigate the question of the detection quality, an evaluation with complex musical pieces, including varying dynamics, is required.

In the future, we aim to optimize our approach towards an online articulation detection function. This may enable the development of woodwind sensor instruments, which provide feedback about the actual performance in a learning situation or can be used as interfaces to physical modelling based sound synthesis in contemporary music performances.

Acknowledgement

This research was supported by the Austrian Science Fund (FWF): P23248-N24. The authors made extensive use of free software under GNU/Linux. Further, we would like to thank Kai Siedenburg for pointing us to wavelet analysis.

6. REFERENCES

[1] A. Gabrielsson, "Music performance research at the millennium," *Psychology of music*, vol. 31, no. 3, pp.

221–272, 2003.

[2] E. Scheirer, "Using musical knowledge to extract expressive performance information from audio recordings," in *Readings in Computational Auditory Scene Analysis*, H. Okuno and D. Rosenthal, Eds., Citeseer. Erlbaum, Mahwah, NJ, 1995, pp. 153–160.

[3] W. Goebl and C. Palmer, "Temporal control and hand movement efficiency in skilled music performance," *PLOS ONE*, vol. 8, no. 1, p. e50901, 2013, (open access).

[4] A. Almeida, R. Chow, J. Smith, and J. Wolfe, "The kinetics and acoustics of fingering and note transitions on the flute," *The Journal of the Acoustical Society of America*, vol. 126(3), pp. 1521–1529, 2009.

[5] H. Kinoshita and S. Obata, "Left hand finger force in violin playing: Tempo, loudness, and finger differences," *The Journal of the Acoustical Society of America*, vol. 126(1), pp. 388–395, 2009.

[6] A. Hofmann, V. Chatziioannou, M. Weilguni, W. Goebl, and W. Kausel, "Measurement setup for articulatory transient differences in woodwind performance," vol. 19, no. 1. ASA, 2013, p. 035060. [Online]. Available: http://link.aip.org/link/?PMA/19/035060/1

[7] S. Böck, F. Krebs, and M. Schedl, "Evaluating the online capabilities of onset detection methods," in *In Proceedings of the 13th International Society for Music Information Retrieval Conference, Porto, Portugal.*, 2012.

[8] K. Krautgartner, "Untersuchungen zur artikulation bei klarinetteninstrumenten im jazz," Ph.D. dissertation, Universität zu Köln, 1982.

[9] D. Liebman, *Developing a personal saxophone sound*. Medfield: Dorn Publications, 2006.

[10] A. Hofmann, V. Chatziioannou, W. Kausel, W. Goebl, M. Weilguni, and W. Smetana, "The influence of tonguing on tone production with single-reed instruments," in *5th Congress of the Alps Adria Acoustics Association*, Zadar, Croatia, 2012.

[11] A. Hofmann, W. Goebl, M. Weilguni, A. Mayer, and W. Smetana, "Measuring tongue and finger coordination in saxophone performance," in *12th International Conference on Music Perception and Cognition (ICMPC) and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM)*, Thessaloniki, Greece, 2012, pp. 442–445.

[12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012. [Online]. Available: http://www.R-project.org

[13] G. Tzanetakis, G. Essl, and P. Cook, "Audio analysis using the discrete wavelet transform," in *Proc. Conf. in Acoustics and Music Theory Applications*, 2001, pp. 318–323.

[14] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transforms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 1, no. 02, pp. 273–302, 1987.

[15] A. Paradzinets, H. Harb, and L. Chen, "Use of continuous wavelet-like transform in automated music transcription," in *Proceedings of EUSIPCO*, 2006.

[16] D. Percival and A. Walden, *Wavelet Methods for Time Series Analysis*, ser. Cambridge Series In Statistical And Probabilistic Mathematics. Cambridge: Cambridge University Press, 2006.

[17] A. Window, *Strain Gauge Technology*, 2nd ed., ser. Elsevier Applied Science. Essex: Elsevier Science Publishers, 1989.

[18] S. Böck, A. Arzt, F. Krebs, and M. Schedl, "Online realtime onset detection with recurrent neural networks," in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12), York, UK*, 2012.