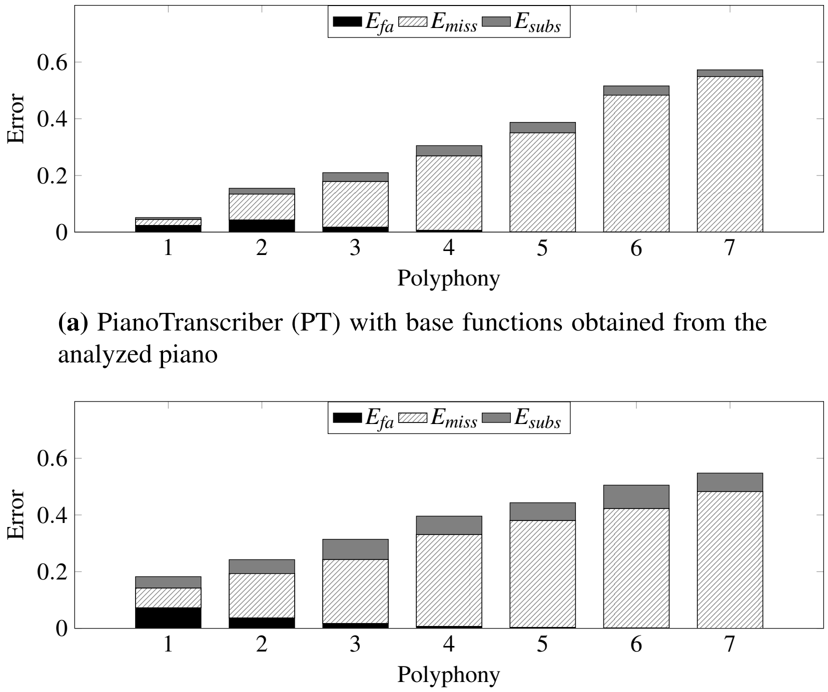
PIANOTRANSCRIBER – A NOTE-BASED APPROACH FOR MULTI-PITCH-TRACKING

Peter Steiner, Simon Stone, Peter Birkholz
Institute of Acoustics and Speech Communication, Technische Universität Dresden
peter.steiner@mailbox.tu-dresden.de

Abstract: Music is an important part of human communication and culture and present in almost every part of the daily life. Just like the acoustic speech sounds can be transcribed in written form, musical sounds can be transcribed in a muscial notation, also called a score. Transcription by listening to music is difficult, because the ability to identify the correct f_0 of a musical note (the so called "absolute pitch") is necessary. According to [1], only 1 out of 10 000 people in North America and Europe have this ability, which makes an automatic system very desirable. Thus, in recent years many approaches for Multi-Pitch-Tracking (identify the individal notes played at any given time) have been proposed. The classification schemes can be grouped into signal-based approaches [2, 3, 4], matrix factorization [5], and neural-network-based approaches [6, 7]. Many signal-based approaches such as [2] use an auditory filterbank that mimicks the human perception of music to obtain a mid-level representation. In this shaped mid-level-representation or directly in the unscaled magnitude spectrum [3, 4], the most pronounced frequency component is estimated and used to model an artificial musical note of the perceived pitch. This note is iteratively subtracted from the entire signal representation. Approaches based on matrix factorization use linear basis transforms to decompose a complex music signal into its basic components, which are supposed to represent notes. Approaches based on neural networks use the time-domain or several mid-level representations as input for the neural networks, which are responsible for the classification step. In the signal-based approaches, no information about music has been used so far for the identification.

In this paper, we therefore propose an algorithm called *PianoTranscriber* (PT) that takes advantage of the limited set of 88 musical notes on a standard piano. Similar to how the FOURIER-transform correlates an input signal with sinusodials, we used the 88 notes on the standard piano in 2 different loudnesses as $2 \cdot 88 = 176$ base functions for the signal decomposition. For each frame, a magnitude spectrum was modeled using those base functions. The resulting time-series for every possible note were smoothed by using knowledge about minimum tone duration and cross-correlation of the entire time-signal. Finally, the smoothed time-series were used to resynthesize the signal. Original and resynthesized signals were compared using onset-detection to remove notes erroneously detected in previous steps. Base functions were calculated using isolated notes of the public available MAPS-database [8]. For parameterization and evaluation, we used further subsets of the database including isolated notes, monophonic excerpts and chords. Because the onset and offset times were extracted from MIDI-data, these were manually adjusted to the original signal. We compared the identification results of PT in three analysis settings to the state-of-the-art algorithm SONIC [6]: using pure sinusodial functions, single notes obtained from the analyzed piano and single notes obtained from other pianos as base functions. PT outperformed SONIC up to a polyphony 5 and achieved similar results for a polyphony of 6 and 7. The error scores were taken from [9].



(a) PianoTranscriber (PT) with base functions obtained from the analyzed piano

(b) The reference algorithm SONIC

Figure 1 – Transcription results of isolated notes and chords with increasing polyphony of PT with base functions obtained from the analyzed piano and the reference algorithm SONIC. E_{fa} counts outputs that cannot be paired with any ground truth, E_{miss} counts missing outputs and E_{subs} reports substitutions [9].

References

[1] Deutsch, D., K. Dooley, T. Henthorn, and B. Head: *Absolute pitch among students in an American music conservatory: Association with tone language fluency. The Journal of the Acoustical Society of America*, 125(4), pp. 2398–2403, 2009. doi:10.1121/1.3081389.

[2] Klapuri, A. P.: *A perceptually motivated multiple-F0 estimation method.* In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*, pp. 291–294. 2005. doi:10.1109/ASPAA.2005.1540227.

[3] Yeh, C., A. Robel, and X. Rodet: *Multiple fundamental frequency estimation of polyphonic music signals.* In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3, pp. iii/225–iii/228 Vol. 3. 2005. doi:10.1109/ICASSP.2005.1415687.

[4] Dressler, K.: *Pitch Estimation by the Pair-Wise Evaluation of Spectral Peaks.* In *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio.* 2011. URL `http://www.aes.org/e-lib/browse.cfm?elib=15960`.

[5] Smaragdis, P. and J. C. Brown: *Non-negative matrix factorization for polyphonic music transcription.* In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pp. 177–180. 2003. doi:10.1109/ASPAA.2003.1285860.

[6] Marolt, M.: *A connectionist approach to automatic transcription of polyphonic piano music. IEEE Transactions on Multimedia*, 6(3), pp. 439–449, 2004. doi:10.1109/TMM.2004.827507.

[7] Thickstun, J., Z. Harchaoui, and S. M. Kakade: *Learning features of music from scratch. ArXiv e-prints*, 2016. `arXiv:1611.09827`.

[8] Emiya, V., R. Badeau, and B. David: *Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. IEEE Transactions on Audio, Speech, and Language Processing*, 18(6), pp. 1643–1654, 2010. doi:10.1109/TASL.2009.2038819.

[9] Poliner, G. E. and D. P. W. Ellis: *A Discriminative Model for Polyphonic Piano Transcription. EURASIP Journal on Advances in Signal Processing*, 2007(1), p. 048317, 2006. doi:10.1155/2007/48317.