parison methods, leading to about 91%, 88%, and 85% F-measure for polyphony 2, 3, and 4, respectively. This performance is very promising and may be accurate enough for many other applications.

4. CONCLUSIONS AND DISCUSSIONS

In this paper, we built a note-level music transcription system based on an existing frame-level transcription approach. The system first performs multi-pitch estimation in each time frame. It then employs a preliminary note tracking to connect pitch estimates into notes. The key step of the system is to perform note sampling to generate a number of subsets of the notes, where each subset is viewed as a transcription candidate. The sampling was based on the note length and note likelihood, which was calculated using the single-pitch likelihood of pitches in the note. Then the transcription candidates are evaluated using the multi-pitch likelihood of simultaneous pitches in all the frames. Finally the candidate with the highest likelihood is returned as the system output. The system is simple and effective. Transcription performance was significantly improved due to the note sampling and likelihood evaluation step. The system also significantly outperforms two other state-of-the-art systems on both note-level and frame-level measures on music pieces with polyphony from 2 to 4.

The technique proposed in this paper is very simple, but the performance improvement is unexpectedly significant. We think the main reason is twofold. First, the note sampling step lets us explore the transcription space, especially the good regions of the transcription space. The single-pitch likelihood of each estimated pitch plays an important role in sampling the notes. In fact, we think that probably any kind of single-pitch salience function that have been proposed in the literature can be used to perform note sampling. The second reason is that we use the multi-pitch likelihood, which considers interactions between simultaneous pitches, to evaluate these sampled transcriptions. This is important because notes contained in a sampled transcription must have high salience, however, when considered as a whole, they may not fit with the audio as well as another sampled transcription. One limitation of the proposed sampling technique is that it can only remove false alarm notes in the preliminary transcription but not adding back missing notes. Therefore, it is important to make the preliminary transcription have a high recall rate before sampling.

5. ACKNOWLEDGEMENT

We thank Emmanouil Benetos and Anssi Klapuri for providing the source code or executable program of their transcription systems for comparison.

6. REFERENCES

[1] M. Bay, A.F. Ehmann, and J.S. Downie, "Evaluation of multiple-F0 estimation and tracking systems," in *Proc. ISMIR*, 2009, pp. 315-320.

[2] J.P. Bello, L. Daudet, M.B., Sandler, "Automatic piano transcription using frequency and time-domain information," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2242-2251, 2006.

[3] E. Benetos, S. Cherla, and T. Weyde, "An efficient shift-invariant model for polyphonic music transcription," in *Proc. 6th Int. Workshop on Machine Learning and Music*, 2013.

[4] E. Benetos and S. Dixon, "A shift-invariant latent variable model for automatic music transcription," *Computer Music J.*, vol. 36, no. 4, pp. 81-94, 2012.

[5] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: challenges and future directions," *J. Intelligent Information Systems*, vol. 41, no. 3, pp. 407-434, 2013.

[6] A. Dessein, A. Cont, G. Lemaitre, "Real-time polyphonic music transcription with nonnegative matrix factorization and beta-divergence," in *Proc. ISMIR*, 2010, pp. 489-494.

[7] Z. Duan, J. Han, and B. Pardo, "Multi-pitch streaming of harmonic sound mixtures," *IEEE Trans. Audio Speech Language Processing*, vol. 22, no. 1, pp. 1-13, 2014.

[8] Z. Duan, B. Pardo, and C. Zhang, "Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions," *IEEE Trans. Audio Speech Language Processing*, vol. 18, no. 8, pp. 2121-2133, 2010.

[9] G. Grindlay and D. Ellis, "Transcribing multi-instrument polyphonic music with hierarchical eigeninstruments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1159-1169, 2011.

[10] P. Grosche, B. Schuller, M. Mller, and G. Rigoll, "Automatic transcription of recorded music," *Acta Acustica United with Acustica*, vol. 98, no. 2, pp. 199-215, 2012.

[11] A. Klapuri, "Multiple fundamental frequency estimation by summing harmonic amplitudes," in *Proc. ISMIR*, 2006, pp. 216-221.

[12] G. Poliner, and D. Ellis, "A discriminative model for polyphonic piano transcription," in *EURASIP J. Advances in Signal Processing*, vol. 8, pp. 154-162, 2007.

[13] S.A. Raczyǹski, N. Ono, and S. Sagayama. "Note detection with dynamic bayesian networks as a post-analysis step for NMF-based multiple pitch estimation techniques," in *Proc. WASPAA*, 2009, pp. 49-52.

[14] M. Ryynänen and A. Klapuri, "Polyphonic music transcription using note event modeling," in *Proc. WASPAA*, 2005, pp. 319-322.

[15] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 528-537, 2010.