Intonation Tendencies in Polyphonic Vocal Ensembles

Johanna Devaney (McGill University)
Ichiro Fujinaga (McGill University)
Daniel P.W. Ellis (Columbia University)

McGill
Schulich School of Music
École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Fonds de recherche sur la société et la culture Québec

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

Canada

Goals

- to reliably extract pitch information from recordings of polyphonic vocal ensembles in order to observe intonation tendencies
- to develop a model of the observed intonation tendencies in polyphonic vocal ensembles

Technical Challenges

- polyphonic pitch estimation
- further complicated by the presence of
  - vibrato
  - reverberation
  - glissandi between notes
  - homogeneity of timbre

Musical Challenges

- at any given point in a piece a vocal ensemble's tuning cannot be consistently related to a single reference point
- a combination of horizontal and vertical musical factors form the reference point for the tuning
- the weighting of these factors likely differs both within and across pieces

Horizontal

- the seventh of the home key functions differently in a piece that modulates
  - B is the leading tone of C Major and the mediant of G Major
    - in C Major it is an unstable pitch that generally resolves to the tonic (C)
    - in G major it is a stable pitch

Horizontal vs Vertical

- even within a single chord there are potential tuning conflicts
- in the context of a G Major chord in C major, B is both the leading-tone of the key and the third of the chord
  - it is commonly held that leading-tones are tuned sharp
  - theories of sensory consonance suggest that a vertical major third will be tuned flat

Vocal Intonation Studies

- Seashore and colleagues' work at the University of Iowa (1930s)
- "Speech, Music, and Hearing" group, Royal Institute of Technology, Stockholm (1980s-present)
- Prame's study of vibrato and intonation in solo singers (1997)

Applications

- singing pedagogy
  - it is not known if general tendencies exist
  - if they exist and can be generalized, the results of this work will be a useful baseline when training vocalists to sing in different tuning systems
- expressive performance
  - MIDI Renditions
  - theories of musical expression

Signal Processing

- need to
  - extract the pitch of multiple voices in polyphonic contexts
  - measure tuning differences that are far smaller than a semitone
  - estimate the perceived pitch over the duration of the note
- can exploit prior knowledge of the score

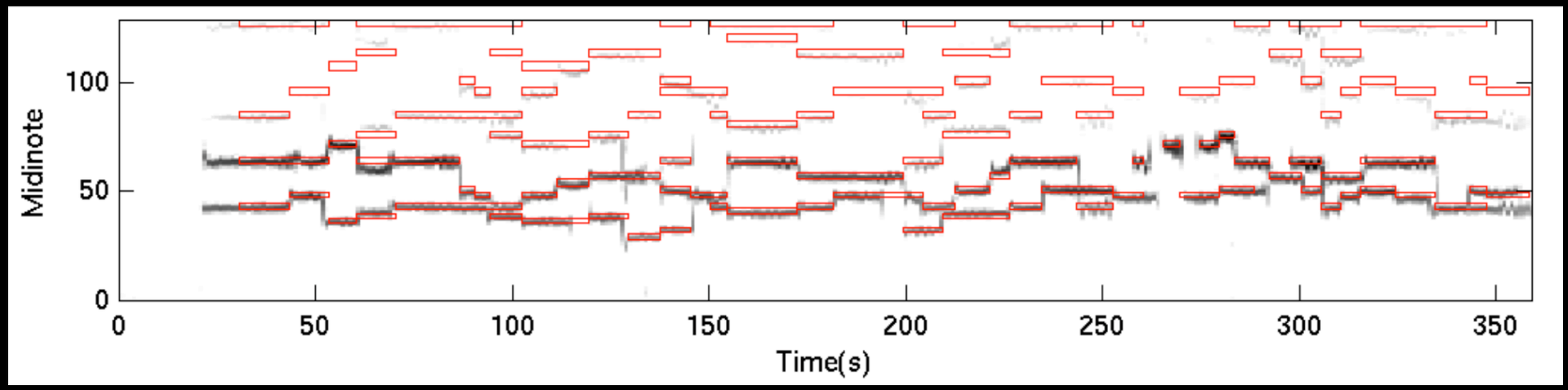
Test set

- multi-tracked recordings of a movement from Machaut's *Notre Dame Mass*
- note onsets and offsets manually annotated using Audacity
- test sets of one, two, three, and four voices
- addition of artificial reverb

Alignment

- dynamic time warping
- features
  - peak structural distance (Orio and Schwarz 2001)
  - chromagrams (Hu et al. 2003)
  - cosine difference of spectral power and first order difference in frequency (Turetsky and Ellis 2003)

Demonstration of Alignment



Instantaneous Frequency (IF)

- calculate a phase derivative within each time-frequency cell of a conventional short-time Fourier transform (Abe et al. 1996)
- the IF in three spectrally-adjacent cells must differ by less than 25% to be considered stable

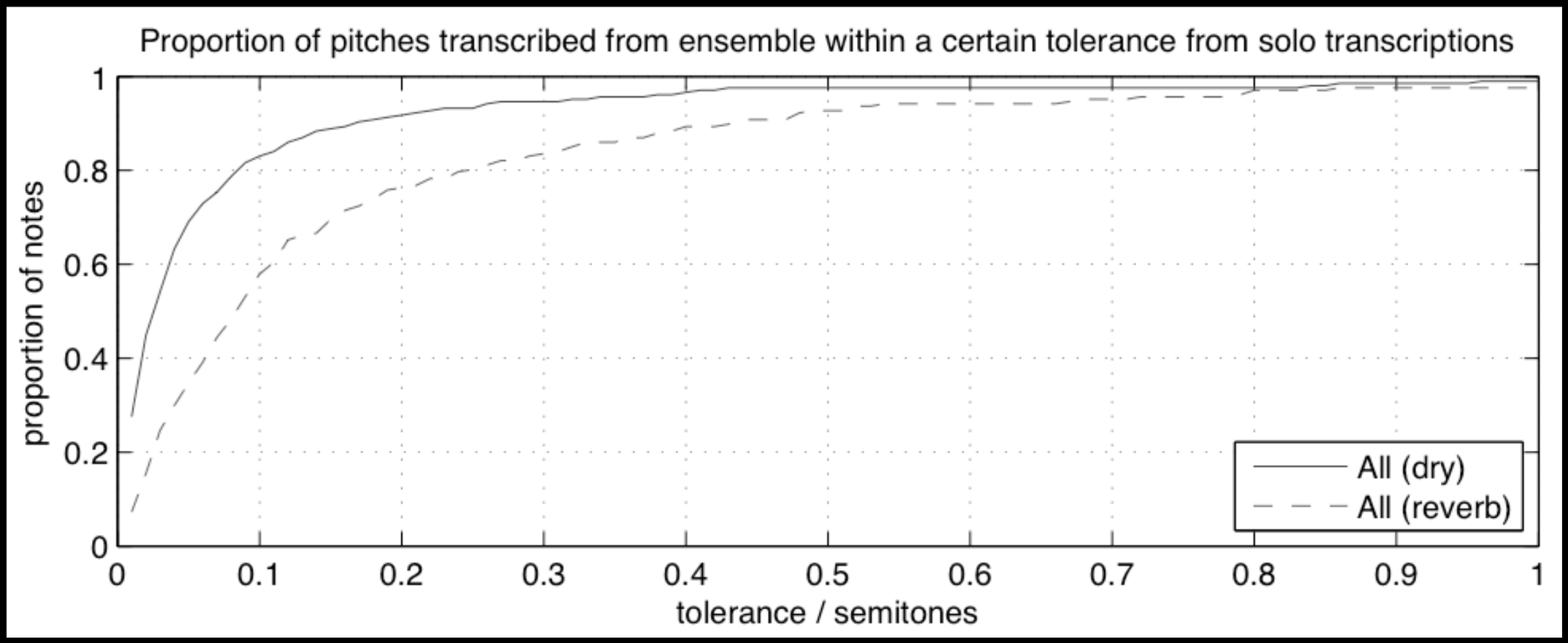
Instantaneous Frequency (IF)

- the generated IF spectrogram recovers the estimated energy and frequency of sinusoids at every time-frequency cell
- aligned MIDI file indicates the time-span for each expected note
- currently only using fundamentals - we are working with Christine Smit on expanding this to use harmonics

Calculating Perceived F0

- perceived F0 is assumed to be the mean frequency over the duration of the note (Brown and Vaughn 1993)
- we intuit this should only be the steady-state portion of the note
  - currently working on perceptual tests to confirm this

Accuracy of IF technique



Future Work

- once a sufficient number of recordings have been processed we will start to model the data
- short-term goal is to find if any generalities exist
- longer-term goal is to develop a theory of vocal intonation practices (Devaney and Ellis 2008)

Final Thoughts

- vocal intonation is a complex phenomenon
- empirical measurements of recorded performances are a useful way to observe intonation practices
  - this presents several non-trivial signal processing challenges for which workarounds are possible
- some of the code is available at
  - www.ee.columbia.edu/~dpwe/resources/matlab/

Abe, T., T. Kobayashi, and, S. Imai. 1996. Robust pitch estimation with harmonics enhancement in noisy environments based on instantaneous frequency. In *Proceedings of the International Conference on Spoken Language*. 1277-80.

Brown, J. C., and K. V. Vaughn. 1996. Pitch center of stringed instrument vibrato tones. *Journal of the Acoustical Society of America*. 100:3, 1728-35.

Devaney, J. and D. P. W. Ellis. 2008. An empirical approach to studying intonation tendencies in polyphonic vocal performances. *Journal of Interdisciplinary Music Studies*. 2: 141–56.

Hu, N., Dannenberg, R., and G. Tzanetakis. 2003. Polyphonic Audio Matching and Alignment for Music Retrieval. In *Proceedings of the IEEE Workshop on Audio and Signal Processing to Audio and Acoustics*. 185-8.

Orio, N., and Schwarz, D. 2001. Alignment of Monophonic and Polyphonic Music to a Score. In *Proceedings of the 2001 International Computer Music Conference*. 129-32.

Prame, E. 1997. Vibrato extent and intonation in professional western lyric singing. *Journal of the Acoustical Society of America*. 102(1): 616–21.

Seashore, C. E. 1938. *Psychology of music*. New York: Dover Publications.

Ternstrom, S. and J. Sundberg.1988. Intonation precision of choir singers. *Journal of the Acoustical Society of America*. 84(1): 59–69.

Turetsky, R., and D. Ellis. 2003. Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses. In *Proceedings of the 2003 International Conference on Music Information Retrieval*. 135-41.