Singing Transcription

Presented by:

Stephen Sinclair

MUMT 611 Winter 2006, McGill University

Uses

- Query by humming (QBH)
- Automatic notation
- Musician's tool for quickly recording a melody for a MIDI system
- Embedded lyrics

Steps

- Voice separation (if necessary)
- End-point detection

  i.e., segmentation
- Island building

  The pitch value during a vowel is usually good, but this is not so for consonants and silence. So vowel “islands” must be built in the pitch envelope.

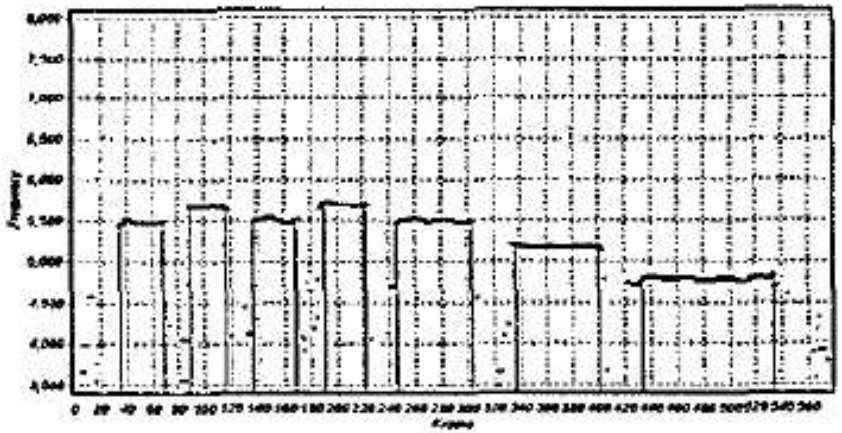


Figure 2. An instance of Island Building for pitch contour

from Wang et al. 2003

Steps (cont'd)

- Smoothing
- Melody tracking

  $N = 69 + 12 \times log_2 \frac{f}{440}$

  Quantized into 100 *cents* per N.

  However singers pitch is often unstable, changing with mood, etc. 1 in 10000 people claim tone-absolute pitch.

  Approaches: Round MIDI, relative adjustment (McNab), absolute adjustment (Haus)
- Music grammar constraints

  Emphasing note-number "guesses" that correspond to the expected musical scale can improve results. Grammatical rules can be applied.
- Estimation of beats and bars
- Transcription into music notation

Result

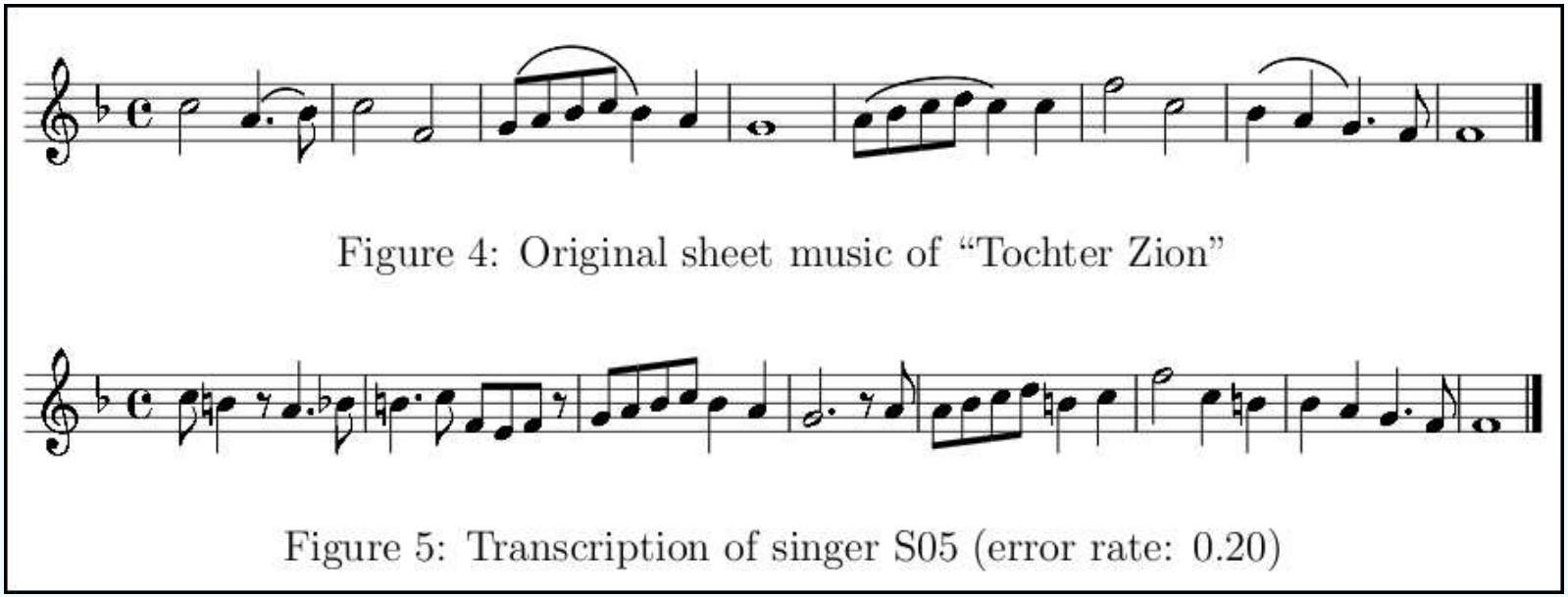


Figure 4: Original sheet music of “Tochter Zion”

Figure 5: Transcription of singer S05 (error rate: 0.20)

from Weihs and Ligges 2003

Sources of error

- Vibrato
- Erroneous absolute pitch
- Erroneous relative pitch
- Erroneous timing

System Scores

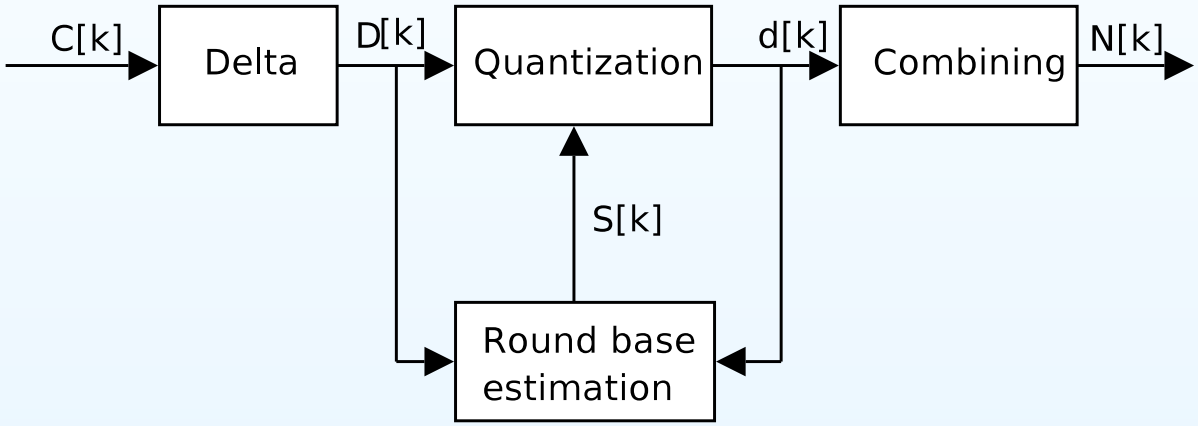
	Akoff	Autoscore	Meldex	Widi	Pollastri
Singing without lyrics					
notes deleted	6.72 %	7.26 %	37.31 %	5.22 %	4.76 %
notes inserted	11.19 %	14.29 %	4.48 %	64.18 %	7.94 %
notes deleted + inserted	17.91 %	21.55 %	41.79 %	69.40 %	12.70 %
exact note recognition error	40.71 %	54.26 %	53.73 %	31.15 %	48.31 %
note recognition error $>$ 1 semitone	4.42 %	10.64 %	28.36 %	1.64 %	10.37 %
Singing with lyrics					
notes deleted	18.50 %	22.95 %	52.46 %	18.50 %	13.66 %
notes inserted	30.00 %	12.02 %	3.28 %	60.50 %	5.46 %
notes deleted + inserted	48.50 %	34.97 %	55.74 %	79.00 %	19.13 %
exact note recognition error	48.34 %	44.27 %	66.23 %	34.72 %	58.39 %
note recognition error $>$ 1 semitone	13.91 %	15.27 %	31.17 %	6.25 %	16.79 %

from Clarisse et al. 2002

Adaptive Round Semitones

- ## Adaptive round semitones

  Proposed by Wang et al. 2003. Implements an adaptive autoregressive model to dynamically changing the tuning scale.



$S[k]$ is the tuning scale, automatically adjusted based on previous note differences.

References

Clarisse, L. P., J. P. Martens, M. Lesaffre, B. D. Baets, H. D. Meyer, and M. Leman (2002). An auditory model based transcriber of singing sequences. In M. Fingerhut (Ed.), *Proceedings of the Third International Conference on Music Information Retrieval: ISMIR 2002*, Paris, France, pp. 116–123. IRCAM - Centre Pompidou.

Wang, C.-K., R.-Y. Lyu, and Y.-C. Chiang (2003, Sept). A robust singing melody tracker using adaptive round semitones (ARS). In *Proceedings of 3rd International Symposium on Image and Signal Processing and Analysis (ISPA03)*, pp. 18–20.

Weihs, C. and U. Ligges (2003). Automatic transcription of singing performances. In *Bulletin of the International Statistical Institute, 54th Session*, Volume LX, pp. 507–510.