ISMIR 2015 Tutorial

Why singing is interesting

Simon Dixon Queen Mary University of London, UK

Masataka Goto AIST, Japan

Matthias Mauch Queen Mary University of London, UK

2015/10/26

Why Singing is Interesting

- All popular music cultures around the world use singing
- The singing voice is the most expressive of all musical instruments
- "Of all musical instruments the human voice is the most worthy because it produces both sound and words, while the others are of use only for sound" (Summa Musice, 13th century)
- Our representations (e.g. MIDI, Western notation) are inadequate for expressive singing
- Knowledge about singing from other disciplines (e.g. physiology, psychology, pedagogy) is rarely exploited in MIR
- Many MIR tasks involving singing have never been attempted

Our Plan

What we said we'd do

... introduce to the ISMIR community the exciting world of singing styles, the mechanisms of the singing voice and provide a guide to representations, engineering tools and methods for analysing and leveraging it.

Our aim

... for music information retrieval specialists to walk away with a newly sparked passion for singing and ideas of how to use our knowledge of singing, and singing information processing, to create new, exciting research.

Overview

10:00-10:05 Overview of this tutorial, brief introduction of three speakers

10:05-10:50 **Part 1: Singing Styles and Psychology of Singing** (45 min)
by **Simon Dixon**
questions (10 min)

11:00-12:15 **Part 2: Practical Guide to Singing Information Research** (45 min)
by **Matthias Mauch**
(11:30-12:00:break)
questions (10 min)

12:25-13:10 **Part 3: Singing Information Processing Systems** (45 min)
by **Masataka Goto**
questions (10 min)

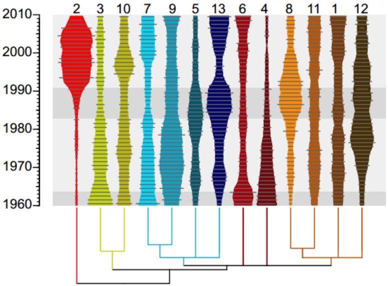
13:20-13:30 Conclusions

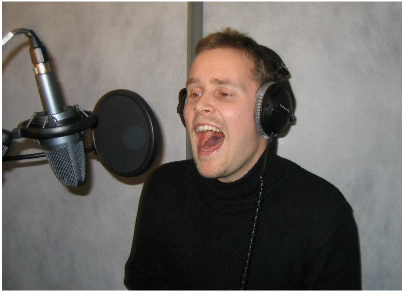
Part 1: Simon Dixon

- **Queen Mary University of London (2006-)**
  - Reader
  - Deputy Director of the Centre for Digital Music
- **Working on music informatics since 1996**
  - Mainly music signal analysis
  - E.g. automatic transcription, beat tracking, audio alignment
- **President of ISMIR (2014-2015)**

Part 2: Matthias Mauch

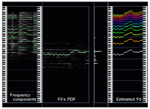
- Senior Applied Researcher in industry
- Visiting Lecturer at Queen Mary University of London
- Working on music informatics since 2006
- Passionate choir singer and pop singer



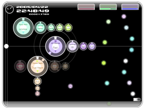


Part 3: Masataka Goto

- **Prime Senior Researcher / Leader of the Media Interaction Group, AIST**
  National Institute of Advanced Industrial Science and Technology
- **Working on music information research since 1992**
- **General chair of ISMIR 2009/2014**

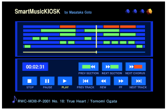


PreFEst



Musicream

http://songle.jp



SmartMusicKIOSK

Robot singer

http://songrium.jp

ISMIR 2015 Tutorial: Why singing is interesting

Part 1: Singing Styles and Psychology of Singing

Centre for Digital Music, Queen Mary University of London

Simon Dixon

2015/10/26

Part 1: Singing Styles and Psychology of Singing

- Singing Styles and Vocal Expression
- Physiology of the Singing Voice
- Intonation, Accuracy, Drift, Poor Singing
- MIR and Singing: Open Problems

Singing Styles and Vocal Expression

Singing Styles

- The voice is a versatile instrument
- It is universal: everyone has one, can use it, and it is suitable for music of all cultures
- It is portable, affordable and expressive
- Use cases: entertainment, art, worship, communication, social
- We observe a great diversity of styles of singing[1]
- Aesthetics (taste, appreciation of beauty) vary by style, and sometimes within styles

[1] J. Potter, ed. (2000). *The Cambridge Companion to Singing*. Cambridge, UK: Cambridge University Press.

Aesthetics: Natural or Artificial?

- Natural
  - Authenticity of expression (e.g. rock, pop, folk styles)
  - Speech-like quality (e.g. Broadway), directness
  - Clarity of lyrics: rap (lyrics foremost) vs opera (intelligibility sacrificed for volume)
  - Amplification destroyed the effort/reward equation
- Artificial
  - Purity of tone, effortless (e.g. Western classical: "objectifying control")
  - Training, discipline ("high" vs "low" culture)
  - Technical prowess (e.g. classical, jazz)
  - Performance, acting (e.g. rock, opera, musicals)
  - Microphone technique
  - Audio effects
- Exceptions to the general patterns disprove any simplistic view

Aesthetics: Other Factors

- Entertainment vs artistic or intellectual traditions
- Individuality
  - Choral: aim to act as one, breathing and articulating together; accurate but not expressive; no vanity
  - Pop: centrality of the star
- Historical shifts in Western art music
  - The "perfect voice is thus high, sweet and clear" (Isadore of Seville, d. 636)
  - "not effeminate, nasal, forced, strained, nor animal-like" (Scientiae artis musicae, Salomon, 1274)
  - Renaissance: small ranges; change music rather than register
  - Baroque: register switch (use of falsetto); throat articulation; don't move any part of body except glottis
  - 18th-19th century: smoothness, little/no vibrato, portamento, imperceptible register switch, no force, precise intonation
  - Garcia (1840): scientific approach: begin notes forcefully
  - Modern: power and unity of timbre across the range

Baroque Chorale: J.S. Bach

https://www.youtube.com/watch?v=MY-aowxVXfI

Broadway Belt: from “Oklahoma”

https://www.youtube.com/watch?v=rbm8u2PMzlc

Western Opera: Diane Damrau

https://www.youtube.com/watch?v=pZcaf9GfyWs

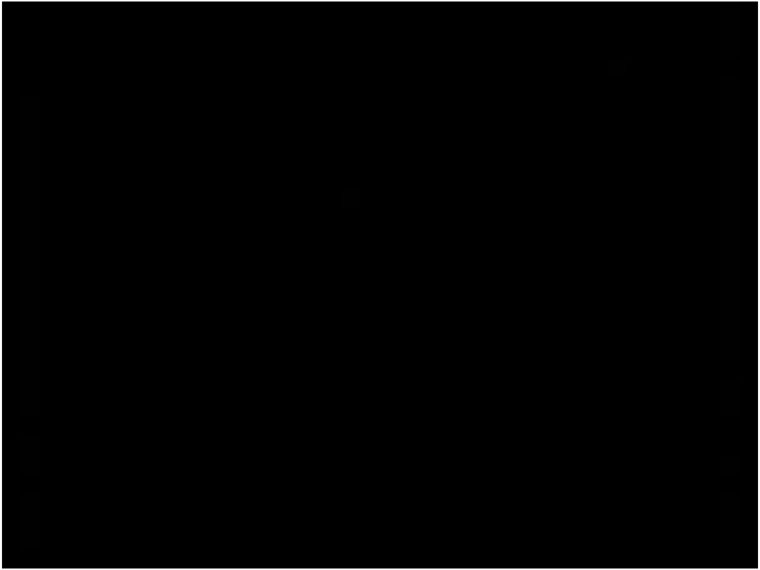
Beijing Opera

https://www.youtube.com/watch?v=mN9iXlfxpxI

European Art Song: Ian Bostridge

https://www.youtube.com/watch?v=DLsaSm5iG9o

Dutch Folk



https://www.youtube.com/watch?v=dVPguklp-Z4

Balkan: Neli Andreeva & Philip Kutev Choir

https://www.youtube.com/watch?v=-_gm0j1H1kc

Inuit Throat Singing

https://www.youtube.com/watch?v=XnPh3GGykaI

Tuvan Overtone Singing

https://www.youtube.com/watch?v=VTCJ5hedcVA

Pakastani Qawwali: Nusrat Fateh Ali Khan

https://www.youtube.com/watch?v=D9Ui2deAKr8

Indian Filmi: Lata Mangeshkar

https://www.youtube.com/watch?v=ubQ9hrKO6XI

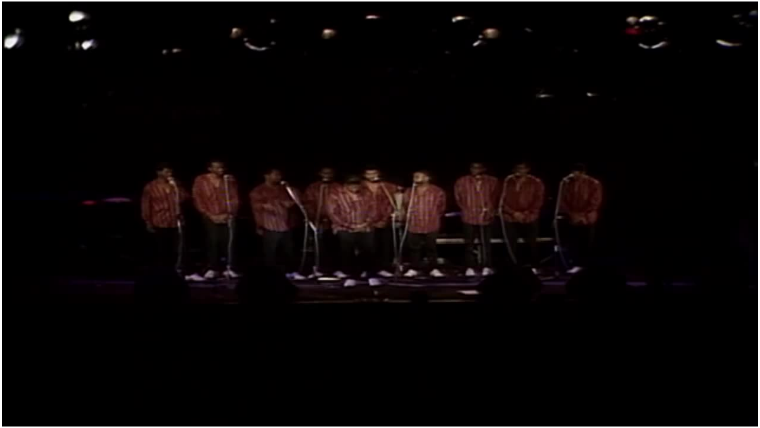
Indian “Beatboxing”: Sheila Chandra

https://www.youtube.com/watch?v=5_N1SWAT6L4

Japanese Enka: Otowa Shinobu

https://www.youtube.com/watch?v=hsWRRhxL838

South Africa: Ladysmith Black Mambazo



https://www.youtube.com/watch?v=288r0Mo1bFw

US Gospel: Fisk Jubilee Singers, 1909)

https://www.youtube.com/watch?v=GUvBGZnL9rE

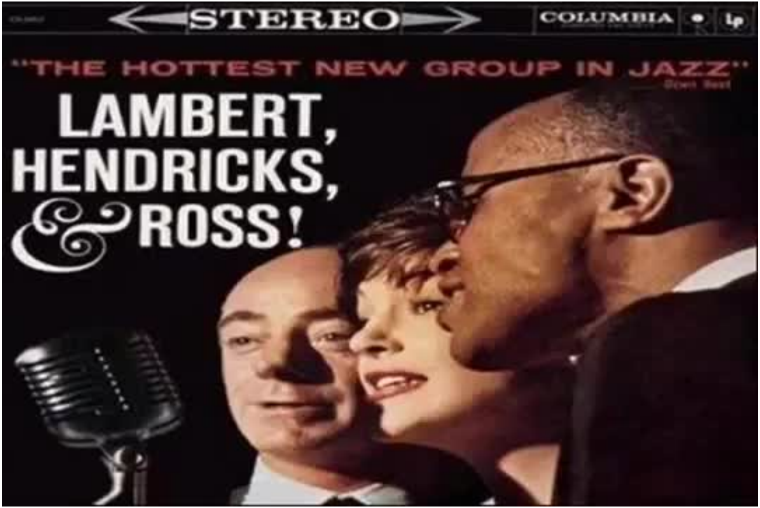
Sacred Harp (Shape Note Singing)

https://www.youtube.com/watch?v=eeQcrOpaCXs

Jazz Scat: Ella Fitzgerald

https://www.youtube.com/watch?v=T8Ji4uG4cac

Jazz Vocalese: Lambert, Hendricks and Ross



https://www.youtube.com/watch?v=LDbAsndZGW0

Jazz Acapella: Take 6

https://www.youtube.com/watch?v=tfHohRpcjo0

Pop Acapella: Vocal Sampling

https://www.youtube.com/watch?v=fW1dUnBhwL8

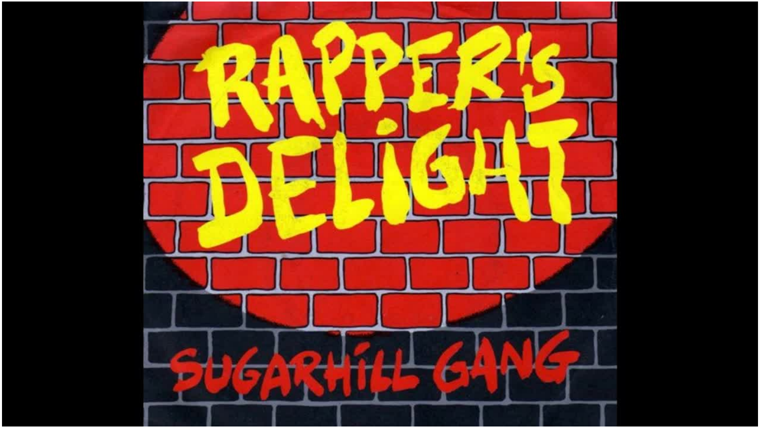
Vocal Acrobatics: Bobby McFerrin

https://www.youtube.com/watch?v=_4BhsYbXwf4

American Soul: James Brown

https://www.youtube.com/watch?v=ETNWrulIDic

Early Rap: Sugar Hill Gang



https://www.youtube.com/watch?v=rKTUAESacQM

Rap: Eminem

https://www.youtube.com/watch?v=XbGs_qK2PQA

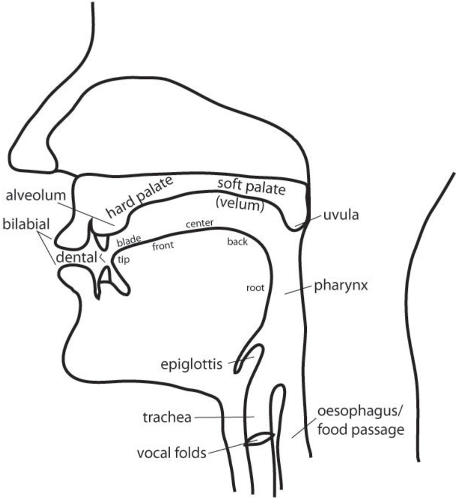
Heavy Rock meets MIR

https://www.youtube.com/watch?v=5MCq9nM-V4M

Physiology of the Singing Voice

How the Voice Works

- Respiratory system: compresses lungs to create airflow
- Vocal folds: chop airstream into a periodic pulsation
- Vocal tract: filters source waveform according to resonances (formants)[2]



[2] J. Sundberg (1987). *The Science of the Singing Voice*. DeKalb IL: Northern Illinois University Press.

Breathing

- Controlled by rib muscles diaphragm, abdominal wall
- Pressure at glottis determines loudness and affects pitch
- Lung pressure 0.4-1.5 kPa gives 65-87 dB SPL at 0.5m
- Air flow: alternating open phase (triangular pulse) and closed (or almost closed) phase, resulting in 12 dB/octave rolloff
- Slope of closing of glottis varies with loudness

Vocal Folds (Cords)

- Run from front (Adam's apple) to back (arytenoid cartilage), with an opening called the *glottis*
- Abducted (spread) and adducted (brought together) by laryngeal muscles operating on the arytenoid cartilages
- This determines the tension in the vocal cords
- Myoelastic theory: explains cyclic opening and closing of glottis
  - the vocal cords are initially closed
  - breath pressure is applied from beneath (*subglottic pressure*)
  - cords remain closed until sufficient pressure builds up to push them apart
  - air then escapes and the pressure drops
  - muscle tension brings the folds back together
- The rate of repetition of this cycle determines the pitch

Phonation Modes[3]

Continuum of tension in vocal cords:

- Completely relaxed (open): cords do not vibrate (*voiceless phonation*)
- Partially lax: high air flow, no closed phase (*breathy phonation*)
- Moderate tension: "sweet spot" of maximum vibration, normal state for spoken vowels (*flow phonation*)
- High tension: low air flow, long closed phase (*pressed phonation*)
- Pressed together (closed): vocal cords block airstream (*glottal stop*)

[3] P. Proutskova et al. (2012). "Breathy or Resonant – A Controlled and Curated Dataset for Phonation Mode Detection in Singing". In: *13th International Society for Music Information Retrieval Conference*, pp. 589–594.

Vocal Fold Oscillation Modes

- Vocal fry
  - Folds thick and relaxed
  - Multiple air bursts followed by a long closed phase
  - Two folds vibrate asynchronously
  - Occurs typically at the end of spoken phrases
- Modal (chest voice)
  - Symmetrical vibration
  - Open phase at least 50% of cycle
  - Sole register of classical tenor, baritone and bass
- Falsetto (head voice)
  - Folds thin and stretched
  - Symmetrical vibration
  - Almost no closed phase
  - Register of countertenor (with closed phase)

Pitch

- Singer's pitch range is determined by length and mass of vocal folds
- Classical voices
  - Soprano: 260-1050 Hz (C4-C6)
  - Alto: 175-700 Hz (F3-F5)
  - Tenor: 130-520 Hz (C3-C5)
  - Bass: 80-330 Hz (E2-E4)
- Vibrato (classical)
  - Pitch modulation via pulsations in cricothyroid muscle
  - Rate: 5-7 Hz
  - Depth (pitch variation): $\pm$0.5-1.5 semitones
- Vibrato (pop)
  - Amplitude modulation via variations in subglottal pressure

Vocal Tract

- Resonances occur in the vocal tract according to its configuration
- Up to 5 formants are relevant for singing
- Vowel quality: mainly determined by first 2 formants
- Voice quality: determined by individual factors (size, shape)
- Singer's formant
  - strong peak in spectral envelope of classical singers
  - clustering of the 3rd, 4th and 5th formants
  - bass (2.2 kHz), tenor (2.9 kHz), alto (3-3.5 kHz)
  - contributes to brilliance of sound and audibility over an orchestra without excessive effort

Examples of Singing Styles and Techniques

- Choral
  - no singer's formant, closer to speech than operatic singing
- Pop and country
  - more similar to speech (breathing patterns, lung pressure)
  - pressed phonation used for high pitches
  - (general) absence of low-larynx technique, diaphragm-oriented breathing, pure tone
- Theatrical (*belting*)
  - narrow pharynx, raised larynx, high lung pressure, long closed phase
  - loud, speech-like
  - boosts high overtones
  - extends range of chest register
- Overtone singing (some Asian cultures)
  - fixed F0
  - formant 2 or 3 is tuned to enhance a specific partial, sometimes stronger than F0
  - results in a new (additional) pitch

Intonation, Accuracy, Drift and Poor Singing

Poor Singers

- Reveal relationship between perception, memory and production; could identify interventions to help people sing
- Pfordresher compared imitation and discrimination task results to isolate causes of poor singing in non-musicians[4]
- Possible models:
  - perceptual deficit: would predict production covarying with perception, small intervals harder to reproduce than large, and little impact of auditory feedback (masking, augmenting)
  - motor deficit: predicting random direction of errors, large intervals harder than small, gravitation towards a "comfortable" pitch, no correlation with discrimination
- "Poor-pitch singing results from mismapping of pitch onto action, rather than problems specific to perceptual, motor, or memory systems."

[4] P. Pfordresher and S. Brown (2007). "Poor-Pitch Singing in the Absence of "Tone-Deafness"". In: *Music Perception* 25.2, pp. 95–115.

Poor Singers

- The majority of occasional singers can carry a tune[5]
- For a well-known tune at a slow tempo, nonmusicians are as proficient as professional singers
- Various categories of poor singers exist, mostly in the pitch domain, but sometimes in timing (selective impairment)
- Not normally the result of impoverished perception
- Absolute and/or relative accuracy in pitch and tempo suggest a multicomponent system underlying proficient singing
- Pitch accuracy (lack of bias) and precision (lack of spread) in singing familiar and unfamiliar melodies were investigated[6]
- Most participants had low systematic bias, but many had a large spread of results for each pitch class (i.e. were imprecise)

[5] Simone Dalla Bella and Magdalena Berkowska (2009). "Singing Proficiency in the Majority". In: *Ann. NY Acad. Sci.* 1169.1, pp. 99–107.

[6] P.Q. Pfordresher et al. (2010). "Imprecise singing is widespread". In: *J. Acoust. Soc. Am.* 128.4, pp. 2182–2190.

Pitch, Intervals and Temperament

- Octaves are divided into 12 (equal?) semitones
- For convenience: $p = 69 + 12\log_2(\frac{f}{440})$
- Musical intervals correspond to fundamental frequency (F0) ratios between constituent tones
- Consonant intervals correspond to simple whole-number ratios
  - 2:1 octave (12 semitones)
  - 3:2 perfect fifth (7 semitones)
- Problem: $2^7 \neq (\frac{3}{2})^{12}$
- For fixed-pitch instruments, some or all fifths are adjusted (tempered) when tuning in order to find a suitable compromise
- Variable pitch instruments (e.g. voice) adjust to the context
  - Temperament is not really needed
  - Different instances of the same note can have different pitches
  - Pitch drift: lack of a fixed reference pitch

Intonation and Drift

- Intonation is the pitch accuracy of a realisation of a note
- Assumes a reference (e.g. accompaniment or previous notes)
- Reported to be a main priority of choir rehearsals
- Drift: cumulative pitch error observed by unaccompanied singers over tens of seconds[7]
- Harmonic progressions can induce drift[8], but drift is also observed in solo singing

---

[7] Richard Seaton, Dennis Pim, and David Sharp (2013). "Pitch Drift in A Cappella Choral Singing". In: *Proc. Inst. Acoust. Ann. Spring Conf.* 35.1, pp. 358–364; Per-Gunnar Alldahl (2006). *Choral Intonation*. p. 4. Gehrman, Stockholm, Sweden.

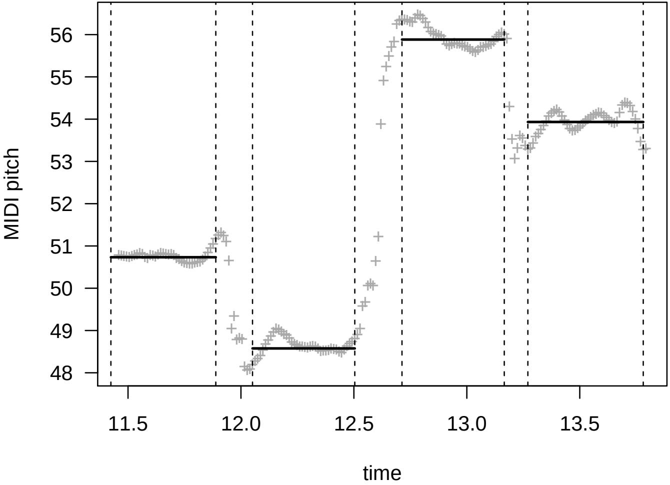
[8] Hiroko Terasawa (2004). *Pitch Drift in Choral Music*. Music 221A final paper. Center for Computer Research in Music and Acoustics; David M. Howard (2007). "Intonation Drift in A Capella Soprano, Alto, Tenor, Bass Quartet Singing With Key Modulation". In: *J. Voice* 21.3, pp. 300–315; J. Devaney, M. Mandel, and I. Fujinaga (2012). "A Study of Intonation in Three-Part Singing Using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)". . In: *ISMIR*, pp. 511–516.

Modelling Drift and Memory[9]

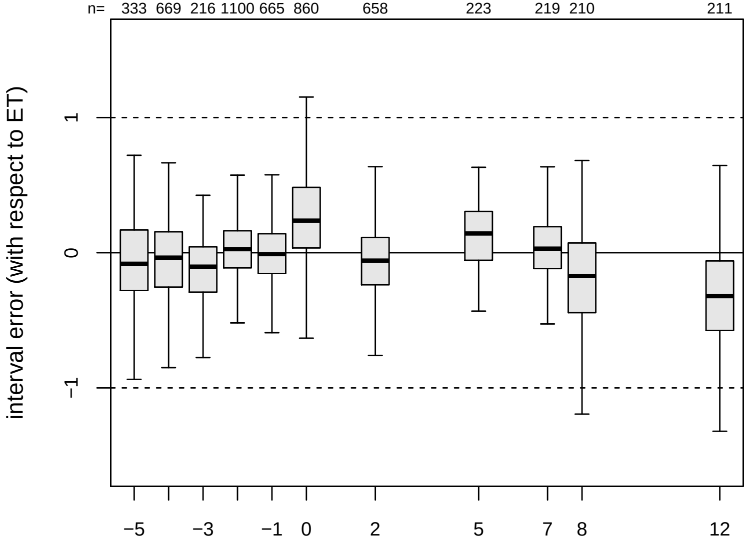
- Solo singing has no external reference pitch; the reference must be internal, in memory
- Drift corresponds to forgetting the reference pitch
- 24 singers of varying ability sang *Happy Birthday* three times (a run) for various conditions
- Semi-automatic analysis to track and segment pitch trajectories
- Median of pitch trajectory used as note-wise pitch
- Accuracy assessed in terms of:
  - Interval Error: relative to the score, assuming equal temperament
  - Pitch (Note) Error: relative to inferred tonic (linear fit)
  - Pitch Drift: between 1st and 3rd runs

[9] M. Mauch, K. Frieler, and S. Dixon (2014). "Intonation in Unaccompanied Singing: Accuracy, Drift and a Model of Reference Pitch Memory". In: *Journal of the Acoustical Society of America* 136.1, pp. 401–411.

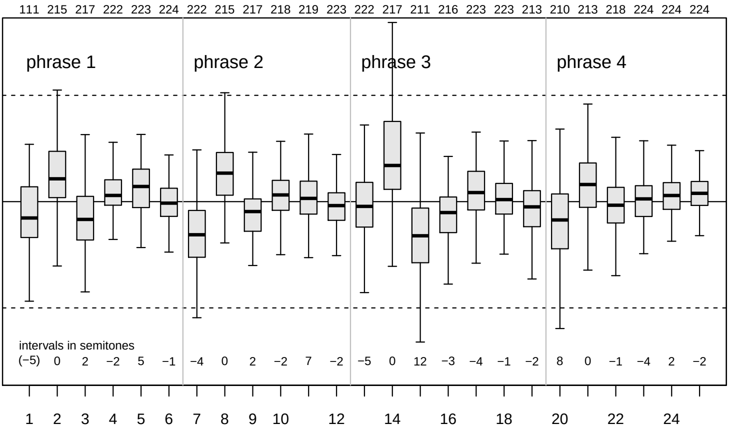
Example: Note Segmentation and Framewise/Notewise Pitch Estimates



Results: Interval Errors by Interval



Results: Interval Errors by Note Number



A Model of Reference Pitch Memory

- Assume that intonation is based on two components: a reference pitch r_i and the score information s_i relative to the reference:

$$p_i = r_i + s_i + \epsilon_i$$

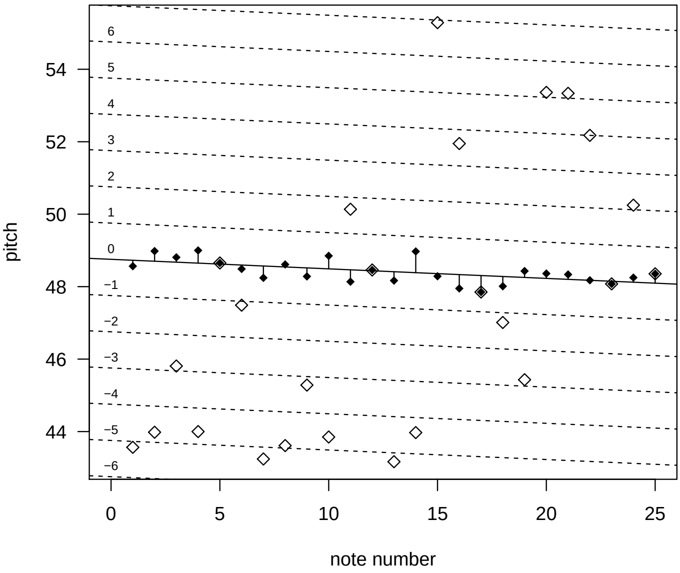
- Assume the memory of the reference r_i is given by the following causal process:

$$r_i = \mu r_{i-1} + (1 - \mu)(p_{i-1} - s_{i-1})$$

  where $p_{i-1} - s_{i-1}$ is a point estimate of the current reference pitch, and $\mu \in [0, 1]$ is a parameter relating to the memory of the previous reference pitch r_{i-1}

- Then $r_i = r_{i-1} + (1 - \mu)e_{i-1}$, i.e. the reference pitch is pulled in the direction of the observed error $e_{i-1} = (p_{i-1} - s_{i-1}) - r_{i-1}$

Example: Local Reference Pitch and Note Errors



Estimating the Memory Parameter μ

- Boundary case 1: $\mu = 0$
  - The previous note realisation is used for reference, with no further memory of the reference pitch
  - Errors are passed on fully, and variance increases with time; this is very different from our observed data
- Boundary case 2: $\mu = 1$
  - The reference pitch is maintained perfectly, unaffected by local errors
  - Variance is constant over time and there is no drift; this is again different from our observations
- Best fit: $\mu = 0.85$ (varying with singer)

Happy Birthday Study: Summary

- Median absolute pitch error = 19 cents; interval error = 27 cents
- Errors were correlated with choir experience and self-reported singing ability, but not with musical background
- Median absolute intonation drift = 11 cents
- Drift was significant in 22% of recordings
- Drift magnitude did not correlate with other measures of singing accuracy or singing experience
- Neither a static intonation memory model nor a memoryless interval-based intonation model account for the observations
- A simple causal tonal reference memory model provides a better fit

MIR and Singing: Open Problems

MIR Tasks Related to Singing

- Singing transcription and analysis
  - Predominant melody extraction
  - F0 estimation (monophonic, polyphonic)
  - Note segmentation 
  - Representation issues
- Vocal activity detection
- Singer identification
- Singing skill evaluation
- Vocal timbre analysis
- Lyric transcription and synchronisation
- Singing synthesis

Open Problems — Challenges for MIR

- Representation of singing
  - Event based representations (scores, MIDI) are insufficient
  - Continuous pitch tracks capture detail of intonation (ornaments, glides, vibrato, kobushi)[10], but *segmentation into notes* is difficult
  - *Integration of timbral information* (phonation, spectral characteristics, phonemes/lyrics) into singing representations
- Algorithms to compare and assess pitch tracks
- Holistic similarity (or skill) estimation that includes pitch, timing and timbre
- New MIREX tasks: assess a singer's naturalness, authenticity or purity of tone

[10] Y. Ikemiya, K. Itoyama, and H.G. Okuno (2014). "Transcribing Vocal Expression from Polyphonic Music". In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3127–3131.

Questions?

ISMIR 2015 Tutorial: Why singing is interesting

Part 2: Practical Guide to Singing Information Research

Centre for Digital Music, Queen Mary University of London

Matthias Mauch

2015/10/26

Section 1

Outline

Brief history of singing analysis tools

Pitch and note tracking state of the art

Annotation/transcription tools

Practical Intonation Analysis

Singing data resources

Section 2

Brief history of singing analysis tools

Tonoscope ($\sim$ 1914)

Figure: Carl Seashore and his tonoscope[1]

[1] Carl E Seashore (1914). "The Tonoscope". In: *The Psychological Monographs* 16.3, pp. 1–12.

Pitch tracking using phonophotography

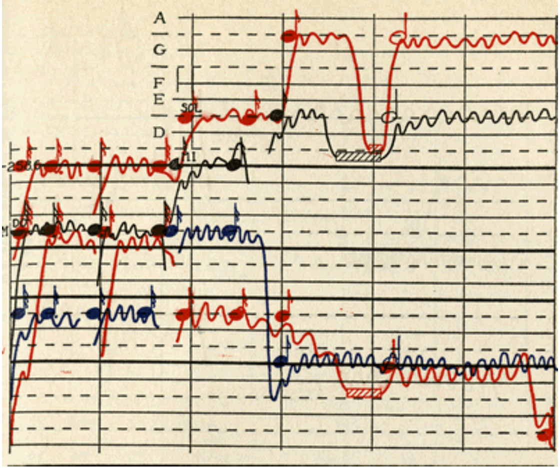


Figure: Example of phonophotography "score", superposition of four separately recorded and transcribed melody lines[3]

[2]Milton Metfessel (1928). *Phonophotography in folk music: American negro songs in new notation.* Univ. North Carolina Press.

Mid-20th century

1967 use of digital computers from the 1960s, e.g. cepstrum analysis[4]

1972 "An analyser has been developed which allows presentation of spectral analyses, amplitude and frequency vibrato **on paper, without need for photography**. Signal frequencies between 100 Hz and 10 kHz can be accepted [...]"[5] (emphasis mine)

1975 "[...] the problem of tracking the frequency of a single (monophonic) periodic signal is one that has been addressed extensively by the speech community. **Some groups consider this to be a solved problem**."[6]

1977 application to music archives[7]

[4] A. M. Noll (1967). "Cepstrum pitch determination". In: *The Journal of the Acoustical Society of America* 41.2, pp. 293–309.

[5] J. Seymour (1972). "Acoustic Analyses of Singing Voices. II. Frequency and Amplitude Vibrato Analyses". In: *Acta Acustica united with Acustica* 27.4, pp. 209–217.

[6] J. A. Moorer (1975). "On the segmentation and analysis of continuous musical sound by digital computer". PhD thesis. Standford University.

[7] B. Larsson (1977). *Pitch tracking in music signals*. Tech. rep., pp. 1–8.

The past 20 years

- pitch tracking becoming a commodity: RAPT,[8] PRAAT,[9] STRAIGHT,[10] YIN,[11] SRH,[12] Tartini ,[13] pYIN[14]

[8] D. Talkin (1995). "A Robust Algorithm for Pitch Tracking". In: *Speech Coding and Synthesis*, pp. 495–518.

[9] P. Boersma (2001). "Praat, a system for doing phonetics by computer". In: *Glot International* 5.9/10, pp. 341–345.

[10] H. Kawahara, J. Estill, and O. Fujimura (2001). "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT". . In: *Proceedings of MAVEBA*, pp. 59–64.

[11] A. de Cheveigné and H. Kawahara (2002). "YIN, a fundamental frequency estimator for speech and music". In: *The Journal of the Acoustical Society of America* 111.4, pp. 1917–1930. DOI: 10.1121/1.1458024.

[12] T. Drugman and A. Alwan (2011). "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics." In: *Proceedings of Interspeech 2011*, pp. 1973–1976.

[13] Philip McLeod (2008). "Fast, accurate pitch detection tools for music analysis". PhD thesis. University of Otago. Department of Computer Science.

[14] M. Mauch and S. Dixon (2014). "pYIN: a Fundamental Frequency Estimator Using Probabilistic Threshold Distributions". In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pp. 659–663.

Recent large-scale research

A recent large scale study of singing (11258 children!) assessed ability aurally, i.e. without measuring even pitch.[15]

[15] G. F. Welch et al. (2014). "Singing and social inclusion". In: *Frontiers in psychology* 5.

Section 3

Pitch and note tracking state of the art

Usage in community

our own survey[16]:

- sent to *ISMIR Community*, *Auditory* and *music-dsp*

Field of work		Position	
Music Inf./MIR	17 (55%)	Student	11 (35%)
Musicology	4 (13%)	Faculty Member	10 (32%)
Bioacoustics	3 (10%)	Post-doc	6 (19%)
Speech Processing	2 (5%)	Industry	4 (13%)
Experience			
Pitch track	18* (58%)		
Note track	16* (52%)		
Both	7 (23%)		
None	3 (10%)		

[16] M. Mauch, C. Cannam, et al. (2015). "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency". In: *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR 2015)*.

Usage in community

our own survey[16]:

- sent to *ISMIR Community*, *Auditory* and *music-dsp*

The DSP algorithms mentioned by survey participants were: YIN (5 participants), Custom-built (3), Aubio (2), and all following ones mentioned once: AMPACT, AMT, DESAM Toolbox, MELODIA, MIR Toolbox, Tartini, TuneR, SampleSumo, silbido, STRAIGHT and SWIPE.

[16] M. Mauch, C. Cannam, et al. (2015). "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency". In: *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR 2015)*.

General overview of pitch trackers

"F0 estimators often have three major components:

a) A pre-processing, or signal conditioning stage,
b) a generator of candidate estimates for the true period sought and
c) a 'post-processing' stage that selects the best candidate and refines the F0 estimate."

Talkin 1995

In addition: voiced/unvoiced detection (either as part of the third step, or as a separate one)

Monophonic pitch tracking, a nearly-solved problem

survey of pitch trackers for singing till 2013 Babacan *et al.*[17]

Table 1. Error Rates Across the Whole Dataset

	GPE (%)	FPE (C)	VDE (%)	FFE (%)
RAPT	1.01	21.96	1.05	1.99
RAPT*	**0.65**	21.98	1.05	**1.66**
STRAIGHTv	1.26	17.22	1.05	2.22
STRAIGHTv*	1.25	17.22	1.05	2.21
PRAATu	1.47	21.91	**0.81**	2.18
PRAAT	1.41	21.93	**0.81**	2.15
PRAAT*	1.41	21.94	**0.81**	2.13
SRHu	1.91	18.99	1.28	3.08
SRH	1.72	17.33	1.33	2.95
SRH*	1.61	17.36	1.33	2.84
SSHu	3.51	19.66	1.27	4.55
SSH	2.40	19.46	1.39	3.61
SSH*	1.91	19.43	1.39	3.16
YINvu	2.69	**8.38**	1.05	3.56
YINv	2.44	12.79	1.05	3.32
YINv*	0.91	12.95	1.05	1.9

[17] O. Babacan et al. (2013). "A Comparative Study of Pitch Extraction Algorithms on a Large Variety of Singing Sounds". In: *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pp. 7815–7819.

Partially solved problem: predominant pitch

- predominant pitch trackers (e.g. PreFEst,[18] MELODIA[19])
- vocal activity detection[20,21]

---

[18] M. Goto (2004). "A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals". In: *Speech Communication* 43.4, pp. 311–329.

[19] J. Salamon and E. Gómez (2012). "Melody extraction from polyphonic music signals using pitch contour characteristics". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 20.6, pp. 1759–1770.

[20] B. Lehner, G. Widmer, and R. Sonnleitner (2014). "On the reduction of false positives in singing voice detection". In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 7480–7484.

[21] M. Mauch, H. Fujihara, et al. (2011). "Timbre and Melody Features for the Recognition of Vocal Activity and Instrumental Solos in Polyphonic Music". In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pp. 233–238.

Pitch tracking implementations I

- YIN Java implementation in Tarsos: `https://github.com/JorenSix/TarsosDSP`, Matlab implementation see `http://www.auditory.org/postings/2002/26.html`, Vamp implementation in pYIN: `https://code.soundsoftware.ac.uk/projects/pyin`
- MELODIA (Vamp plugin) `http://mtg.upf.edu/technologies/melodia`
- pYIN Vamp plugin and source code: `https://code.soundsoftware.ac.uk/projects/pyin` or Python implementation: `https://github.com/ronggong/pypYIN`
- STRAIGHT (Matlab, available upon request) `http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_e.html`

Pitch tracking implementations II

- SWIPE Matlab: http://www.cise.ufl.edu/~acamacho/publications/swipep.m SPTK/Python: http://pysptk.readthedocs.org/en/latest/sptk.html#f0-analysis
- Tartini for Supercollider http://doc.sccode.org/Classes/Tartini.html or standalone http://miracle.otago.ac.nz/tartini/
- Aubio http://aubio.org/ or in Vamp: http://aubio.org/vamp-aubio-plugins/
- RAPT (Matlab) http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/fxrapt.html
- mirtoolbox https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox
- Cepstral Pitch Tracker https://code.soundsoftware.ac.uk/projects/cepstral-pitchtracker

Note tracking

- much less explored (only few papers[22,23,24,25,26])
- ill-defined for music that is not sung from a fixed note representation

---

[22] T. De Mulder et al. (2004). "Recent improvements of an auditory model based front-end for the transcription of vocal queries". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*. Vol. 4, pp. iv-257–iv-260. DOI: 10.1109/ICASSP.2004.1326812.

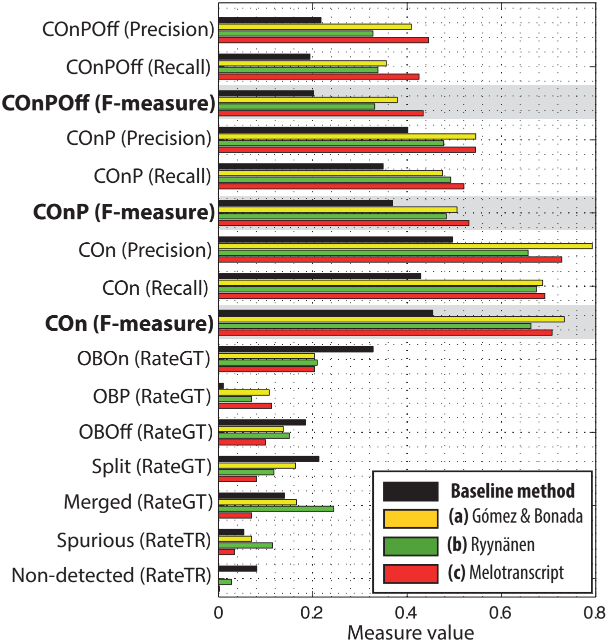
[23] M. P. Ryynänen and A. P. Klapuri (2008). "Automatic transcription of melody, bass line, and chords in polyphonic music". In: *Computer Music Journal* 32.3, pp. 72–86.

[24] E. Gómez and J. Bonada (2013). "Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing". In: *Computer Music Journal* 37.2, pp. 73–90.

[25] M. Mauch, C. Cannam, et al. (2015). "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency". In: *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR 2015)*.

[26] N. Kroher and E. Gómez (in review). "Automatic Transcription of Flamenco Singing from Polyphonic Music Recordings". In: *arXiv preprint arXiv:1510.04039*

Note tracking — low performance



Melody *note* tracking implementations I

- Aubio `http://aubio.org/`
- pYIN: `https://code.soundsoftware.ac.uk/projects/pyin`
- CANTE `http://cofla-project.com/cante.html`
- Cepstral Pitch Tracker `https://code.soundsoftware.ac.uk/projects/cepstral-pitchtracker`

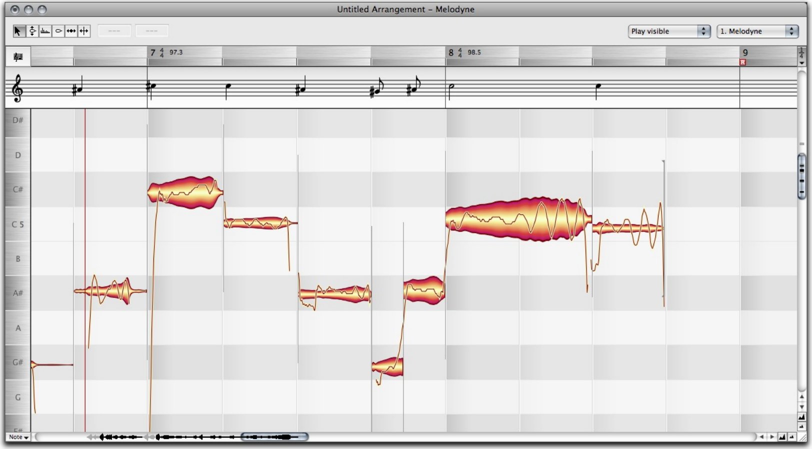
Section 4

Annotation/transcription tools

Survey

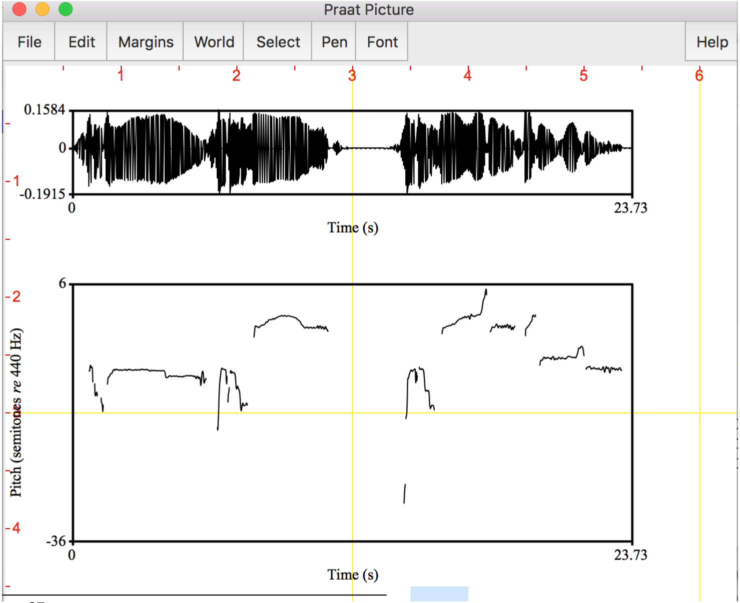
The tools with graphical user interfaces mentioned by survey participants were: Sonic Visualiser (12 participants), Praat (11), Custom-built (3), Melodyne (3), Raven (and Canary) (3), Tony (3), WaveSurfer (3), Cubase (2), and the following mentioned once: AudioSculpt, Adobe Audition, Audacity, Logic, Sound Analysis Pro, Tartini and Transcribe!.

Melodyne



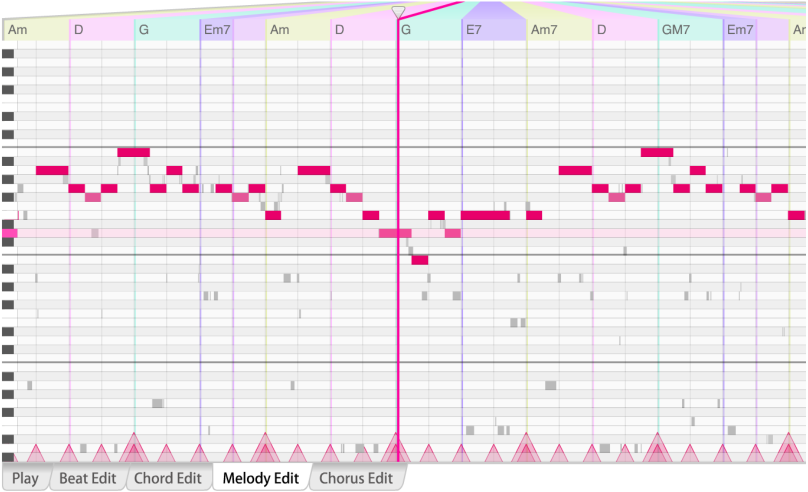
Untitled Arrangement – Melodyne
Play visible
1. Melodyne
7 4/4 97.3
8 4/4 98.5
9
D#
D
C#
C 5
B
A#
A
G#
G
Note

Praat[27]



[27] P. Boersma (2001). “Praat, a system for doing phonetics by computer”. In: *Glot International* 5.9/10, pp. 341–345.

Songle.jp[28]

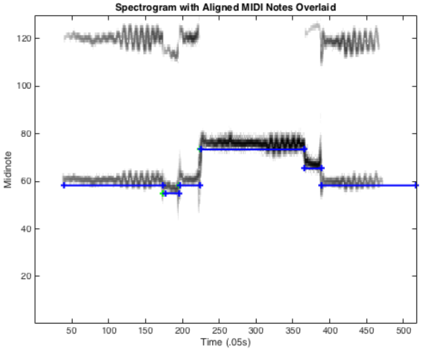


[28] M. Goto et al. (2011). "Songle: A Web Service for Active Music Listening Improved by User Contributions". In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pp. 311–316.

AMPACT[29]

"Automatic Music Performance Analysis and Comparison Toolkit"

- monophonic sung/MIDI alignment
- perceived pitch
- vibrato rate/depth
- note slope calculation



[29] J. Devaney, M. Mandel, and I. Fujinaga (2012). "A Study of Intonation in Three-Part Singing Using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)". . In: *13th International Society of Music Information Retrieval Conference*, pp. 511–516.

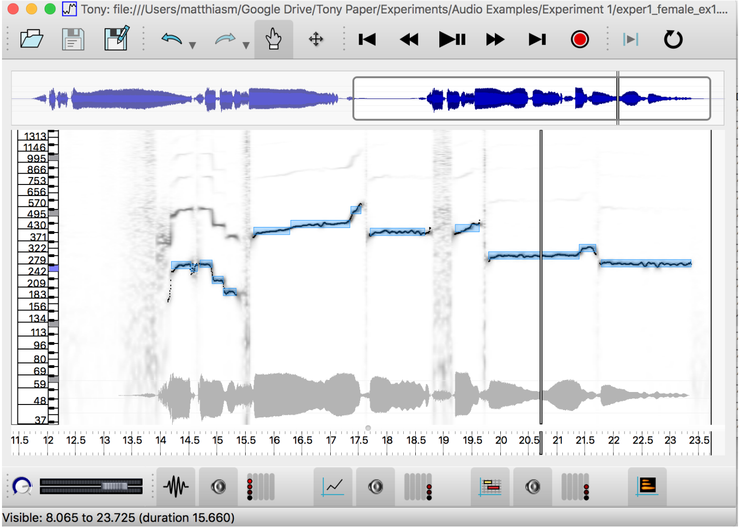
Requirements

	Melodyne	Praat	Sonic Visualiser
estimate pitch	✔	✔	✔
estimate notes	✔	~✔	✔
note/pitch correction	✔	✘	✘
note/pitch sonification	✘	✘	✘
save note/ pitch track	~✔	✘	✔
load note/ pitch track	✘	✘	✔

Requirements

	Melodyne	Praat	Sonic Visualiser	?
estimate pitch	✔	✔	✔	✔
estimate notes	✔	~✔	✔	✔
note/pitch correction	✔	✘	✘	✔
note/pitch sonification	✘	✘	✘	✔
save note/ pitch track	~✔	✘	✔	✔
load note/ pitch track	✘	✘	✔	✔

Tony[30]



[30] M. Mauch, C. Cannam, et al. (2015). "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency". In: *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR 2015)*.

Section 5

Practical Intonation Analysis

A few possible research questions

- vibrato: vibrato depth and rate in different instruments[31]
- scale: equal tempered vs. just intonation[32]
- poor vs. good singing: accuracy and precision[33]
- intonation drift: does intonation reference change over time?[34]

---

[31] L. Yang, M. Tian, and E. Chew (2015). "Vibrato Characteristics and Frequency Histogram Envelopes in Beijing Opera Singing". In: *Fifth Biennial Mathematics and Computation in Music International Conference (MCM2015)*.

[32] J. Devaney, M. Mandel, and I. Fujinaga (2012). "A Study of Intonation in Three-Part Singing Using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)". . In: *13th International Society of Music Information Retrieval Conference*, pp. 511–516.

[33] P. Q. Pfordresher and S. Brown (2007). "Poor-Pitch Singing in the Absence of "Tone Deafness"". In: *Music Perception* 25.2, pp. 95–115. DOI: 10.1525/mp.2007.25.2.95.

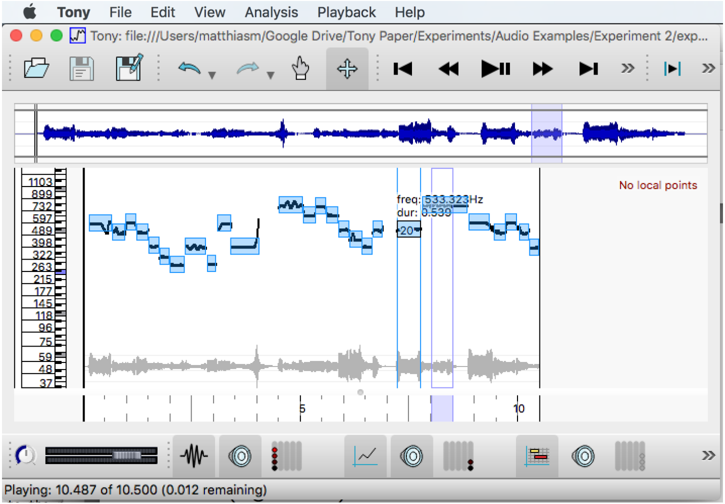
[34] M. Mauch, K. Frieler, and S. Dixon (2014). "Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory". In: *Journal of the Acoustical Society of America* 136.1, pp. 401–411.

Note pitches

Most psychological studies use note-based frequency estimates[35]

[35] P. Q. Pfordresher and S. Brown (2007). "Poor-Pitch Singing in the Absence of "Tone Deafness"". In: *Music Perception* 25.2, pp. 95–115. DOI: 10.1525/mp.2007.25.2.95; S. Dalla Bella, J. Giguère, and I. Peretz (2007). "Singing proficiency in the general population". In: *Journal of the Acoustical Society of America* 121.2, p. 1182. DOI: 10.1121/1.2427111; J. Devaney and D. P. W. Ellis (2008). "An Empirical Approach to Studying Intonation Tendencies in Polyphonic Vocal Performances". In: *Journal of Interdisciplinary Music Studies* 2.1&2, pp. 141–156.

Tony demo 1 (Erhu)



Intonation

Intonation is defined as "accuracy of pitch in playing or singing" ,[36] or "the act of singing or playing in tune" .[37]

$$p = 69 + 12\log_2 \frac{f_0}{440}. \tag{1}$$

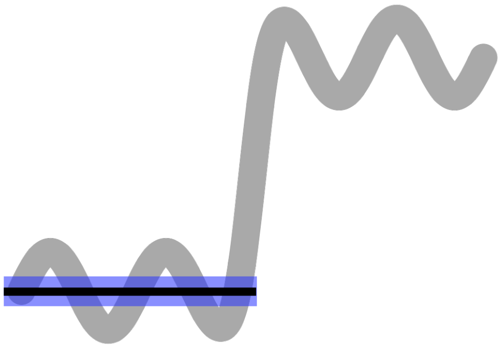
- pitch strongly associated with fundamental frequency
- we usually measure pitch as in (1), i.e. in semitones with middle C corresponding to $p = 60$

[36] Julia Swannell (1992). *The Oxford Modern English Dictionary*. p. 560. Oxford University Press, USA.

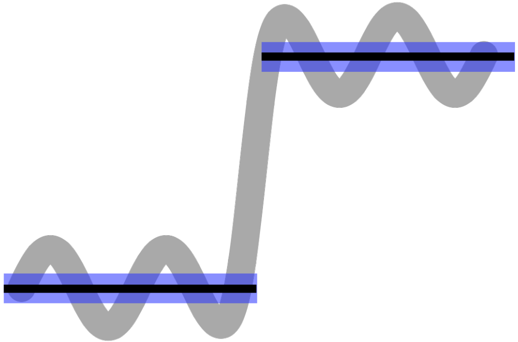
[37] Michael Kennedy (1980). *The Concise Oxford Dictionary of Music*. p. 319. Oxford University Press, Oxford, United Kingdom.

Interval error

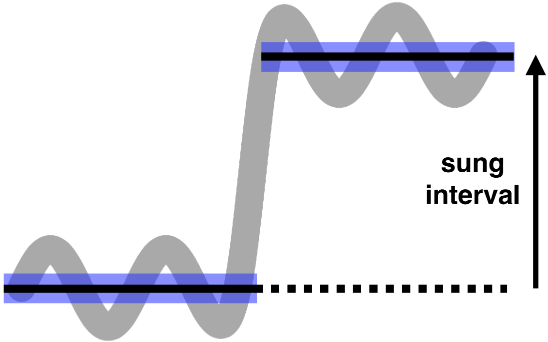
Interval error



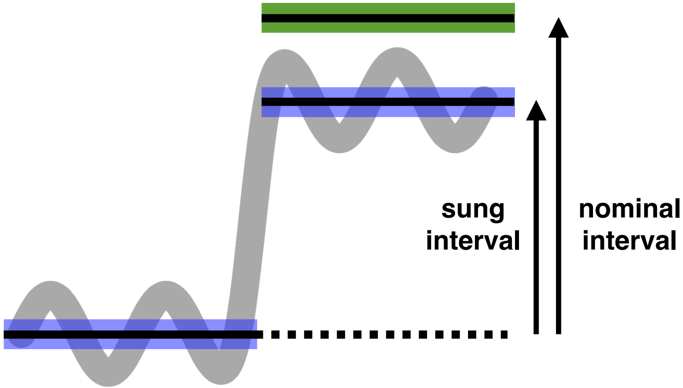
Interval error



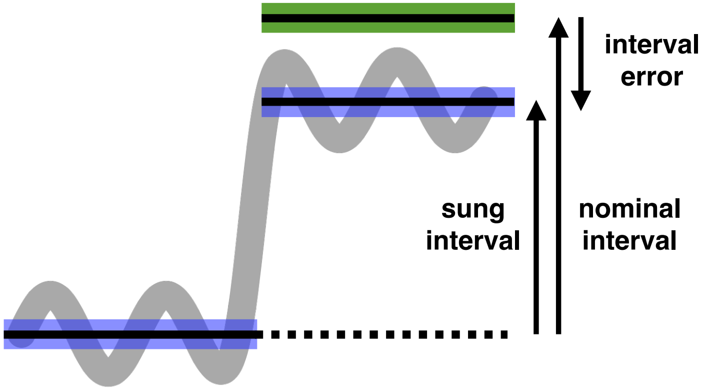
Interval error



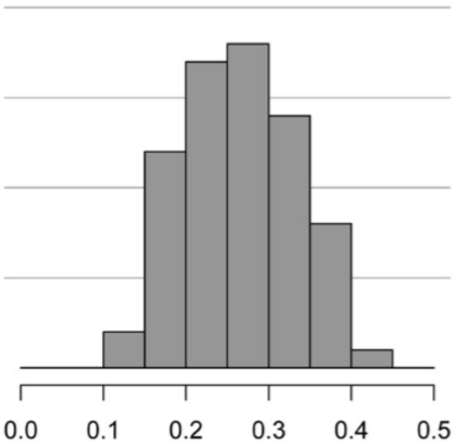
Interval error



Interval error



Interval error stats



from our paper[38]

[38] M. Mauch, K. Frieler, and S. Dixon (2014). "Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory". In: *Journal of the Acoustical Society of America* 136.1, pp. 401–411.

Tony demo 2 (Happy Birthday) + IPython Notebook

jupyter Tutorial Singing Analysis Last Checkpoint: 13 hours ago (autosaved)

File Edit View Insert Cell Kernel Help Python 2

Code Cell Toolbar: None

Intonation analysis example

Imports

In [118]:

```
# numpy and matplotlib-related
import matplotlib
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline
```

In [119]:

```
# other funky imports
import bz2
from xml.etree import ElementTree
```

Constants and helper functions

In [120]:

```
fs = 44100
```

Tony demo 3 (Happy Birthday singing!) + updated IPython Notebook

link to notebook online, don't use in presentation :)

What can go wrong? Things to consider.

Timing. You're interested in intonation, but people sing at different speeds, so you don't know whether intonation differences are caused by the different speed.

Solution:

- provide click track to singers

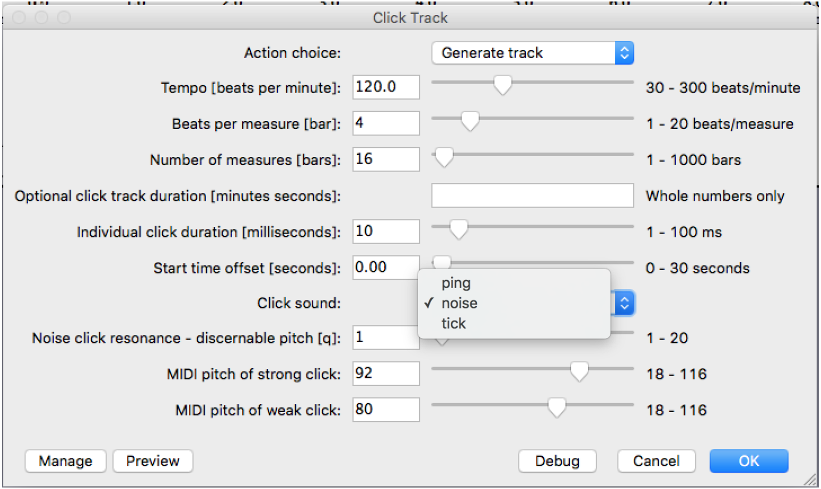
What can go wrong? Things to consider.

Unintended pitch bias. You're using Audacity's standard clicks, but after a while you notice that they're actually pitched, so you're unsure if singers pick up their pitch from the clicks.

Solution:

- use un-pitched noise clicks
- eliminate other pitched sounds (fans etc.) from the environment

What can go wrong? Things to consider.



What can go wrong? Things to consider.

Audio analysis fail. You recorded in a reverberant or noisy room, and in the analysis stage you realise that it's not clear what is the clean singing signal, and what is echo or noise.

Solution:

- use rooms with little reverb
- seek quiet rooms and CHECK FOR HUMS!
- use close mic'ing (we used a headset)

What can go wrong? Things to consider.

Singers out of range. You recorded singers and asked them to sing a melody, giving them a particular pitch. Some singers had to try very hard to reach the high notes (or couldn't reach them), while for others the melody was in their comfort range.

Solution:

- nothing to worry about if testing for reaction to vocal strain was your aim ;)
- find pitch that singers are comfortable with

What can go wrong? Things to consider.

Data unsharable. You recorded singers but didn't ask them whether you could (anonymously) share their singing recordings with the community.

Solution:

- do ask them in writing (and record the answers)
- by the way: also make sure you've got ethics approval for your experiments (this is usually very easy, at least at UK universities)

What can go wrong? Things to consider.

Poor singers. Some of the singers either don't sing the right song, or sing it so badly that it's unrecognisable.

Solution:

- devise a rule in advance to remove poor singers (e.g. all that have an interval error > 1 semitone in more than 20% of the notes.

Section 6

Singing data resources

- Meertens Tune Collections
  http://www.liederenbank.nl/mtc/
- Dawn Black's Singing Voice Audio Dataset
  http://isophonics.net/SingingVoiceDataset
- QBSH Corpus http://neural.cs.nthu.edu.tw/jang2/dataSet/childSong4public/QBSH-corpus/
- IRMAS (instrument recognition, but has 778 voice samples)
  http://www.mtg.upf.edu/download/datasets/irmas
- iKala Dataset http://mac.citi.sinica.edu.tw/ikala/
- MTG-QBH
  http://mtg.upf.edu/download/datasets/mtg-qbh
- Molina's evaluation framework incl. some data
  http://www.atic.uma.es/ismir2014singing/
- Jiajie Dai's Singing Experiment data http://figshare.com/articles/Media_Content_for_Analysis_of_Intonation_Trajectories_in_Solo_Singing_/1482221

- Polina Proutskova's phonation modes dataset http://www.doc.gold.ac.uk/~mas02pp/phonation_modes/
- MedleyDB http://medleydb.weebly.com/downloads.html
- RWC (Musical Instrument Sound) https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html
- TONAS Dataset http://mtg.upf.edu/download/datasets/tonas
- ccmixter corpus (separate vocal tracks) http://www.loria.fr/~aliutkus/kam/
- Jamendo corpus http://www.mathieuramona.com/wp/data/jamendo/
- Ultrastar Database (annotations for vocals, gender, ..., but no audio for popular songs) http://openaudio.eu/
- RWC Melody line annotations and more detailed vocal/instrumental activity https://staff.aist.go.jp/m.goto/RWC-MDB/AIST-Annotation/

- Tunebot Database
  `http://music.cs.northwestern.edu/data/tunebot/`
- MARG database `http://marg.snu.ac.kr/?page_id=767`
- NTENT Singing Voice Database (by Liliya Tsirulnik and Shlomo Dubnov.)
  `http://liliyatsirulnik.wix.com/svdb`
- Mixing Secrets Dataset `https://sisec.inria.fr/professionally-produced-music-recordings/`
- Cofla Flamenco Annotations
  http://cofla-project.com/corpus.html

Bibliography I

Babacan, O. et al. (2013). "A Comparative Study of Pitch Extraction Algorithms on a Large Variety of Singing Sounds". In: *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pp. 7815–7819.

Boersma, P. (2001). "Praat, a system for doing phonetics by computer". In: *Glot International* 5.9/10, pp. 341–345.

Cheveigné, A. de and H. Kawahara (2002). "YIN, a fundamental frequency estimator for speech and music". In: *The Journal of the Acoustical Society of America* 111.4, pp. 1917–1930. DOI: 10.1121/1.1458024.

Dalla Bella, S., J. Giguère, and I. Peretz (2007). "Singing proficiency in the general population". In: *Journal of the Acoustical Society of America* 121.2, p. 1182. DOI: 10.1121/1.2427111.

Bibliography II

De Mulder, T. et al. (2004). "Recent improvements of an auditory model based front-end for the transcription of vocal queries". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*. Vol. 4, pp. iv-257–iv-260. DOI: 10.1109/ICASSP.2004.1326812.

Devaney, J. and D. P. W. Ellis (2008). "An Empirical Approach to Studying Intonation Tendencies in Polyphonic Vocal Performances". In: *Journal of Interdisciplinary Music Studies* 2.1&2, pp. 141–156.

Devaney, J., M. Mandel, and I. Fujinaga (2012). "A Study of Intonation in Three-Part Singing Using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT)". In: *13th International Society of Music Information Retrieval Conference*, pp. 511–516.

Drugman, T. and A. Alwan (2011). "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics." In: *Proceedings of Interspeech 2011*, pp. 1973–1976.

Bibliography III

Gómez, E. and J. Bonada (2013). "Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing". In: *Computer Music Journal* 37.2, pp. 73–90.

Goto, M. (2004). "A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals". In: *Speech Communication* 43.4, pp. 311–329.

Goto, M. et al. (2011). "Songle: A Web Service for Active Music Listening Improved by User Contributions". In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pp. 311–316.

Kawahara, H., J. Estill, and O. Fujimura (2001). "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT". In: *Proceedings of MAVEBA*, pp. 59–64.

Bibliography IV

Kennedy, Michael (1980). *The Concise Oxford Dictionary of Music*. p. 319. Oxford University Press, Oxford, United Kingdom.

Kroher, N. and E. Gómez (in review). "Automatic Transcription of Flamenco Singing from Polyphonic Music Recordings". In: *arXiv preprint arXiv:1510.04039*.

Larsson, B. (1977). *Pitch tracking in music signals*. Tech. rep., pp. 1–8.

Lehner, B., G. Widmer, and R. Sonnleitner (2014). "On the reduction of false positives in singing voice detection". In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 7480–7484.

Mauch, M., C. Cannam, et al. (2015). "Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency". In: *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR 2015)*.

Bibliography V

Mauch, M. and S. Dixon (2014). “pYIN: a Fundamental Frequency Estimator Using Probabilistic Threshold Distributions”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pp. 659–663.

Mauch, M., K. Frieler, and S. Dixon (2014). “Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory”. In: *Journal of the Acoustical Society of America* 136.1, pp. 401–411.

Mauch, M., H. Fujihara, et al. (2011). “Timbre and Melody Features for the Recognition of Vocal Activity and Instrumental Solos in Polyphonic Music”. In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pp. 233–238.

McLeod, Philip (2008). “Fast, accurate pitch detection tools for music analysis”. PhD thesis. University of Otago. Department of Computer Science.

Bibliography VI

Metfessel, Milton (1928). *Phonophotography in folk music: American negro songs in new notation.* Univ. North Carolina Press.

Moorer, J. A. (1975). "On the segmentation and analysis of continuous musical sound by digital computer". PhD thesis. Standford University.

Noll, A. M. (1967). "Cepstrum pitch determination". In: *The Journal of the Acoustical Society of America* 41.2, pp. 293–309.

Pfordresher, P. Q. and S. Brown (2007). "Poor-Pitch Singing in the Absence of "Tone Deafness"". In: *Music Perception* 25.2, pp. 95–115. DOI: 10.1525/mp.2007.25.2.95.

Ryynänen, M. P. and A. P. Klapuri (2008). "Automatic transcription of melody, bass line, and chords in polyphonic music". In: *Computer Music Journal* 32.3, pp. 72–86.

Bibliography VII

Salamon, J. and E. Gómez (2012). "Melody extraction from polyphonic music signals using pitch contour characteristics". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 20.6, pp. 1759–1770.

Seashore, Carl E (1914). "The Tonoscope". In: *The Psychological Monographs* 16.3, pp. 1–12.

Seymour, J. (1972). "Acoustic Analyses of Singing Voices. II. Frequency and Amplitude Vibrato Analyses". In: *Acta Acustica united with Acustica* 27.4, pp. 209–217.

Swannell, Julia (1992). *The Oxford Modern English Dictionary*. p. 560. Oxford University Press, USA.

Talkin, D. (1995). "A Robust Algorithm for Pitch Tracking". In: *Speech Coding and Synthesis*, pp. 495–518.

Welch, G. F. et al. (2014). "Singing and social inclusion". In: *Frontiers in psychology* 5.

Bibliography VIII

Yang, L., M. Tian, and E. Chew (2015). “Vibrato Characteristics and Frequency Histogram Envelopes in Beijing Opera Singing”. In: *Fifth Biennial Mathematics and Computation in Music International Conference (MCM2015)*.

ISMIR 2015 Tutorial: Why singing is interesting

Part 3: Singing Information Processing Systems

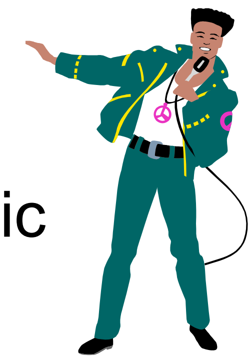
AIST (National Institute of Advanced Industrial Science and Technology)

Masataka Goto

2015/10/26

Singing Information Processing

❑ “Singing information processing” [Goto, *et al.*, 2009-]



- Music information research for singing voices

  Signal processing + machine learning + interfaces + …

- Singing is one of the most important elements of music

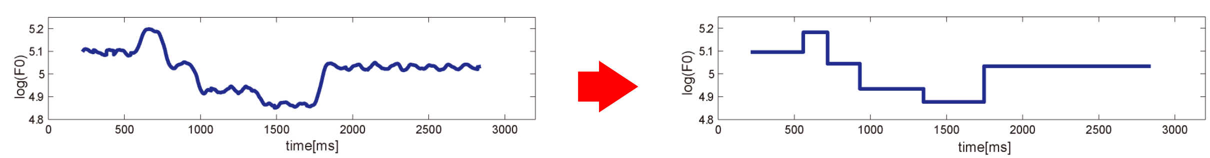
  Many people listen to music with a focus on singing

- Attract attention not only from a scientific point of view but also from the standpoint of industrial applications

  - Automatic singing pitch correction (e.g., *“Auto-Tune”*)

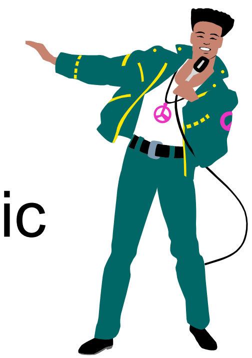
    Intentionally used to achieve a desired effect such as *T-Pain (USA)* and *Perfume (Japan)*

F0 of natural singing　　　　After Auto-Tune



Singing Information Processing

- **"Singing information processing"** [Goto, *et al.*, 2009-]
  - Music information research for singing voices
    Signal processing + machine learning + interfaces + …
  - Singing is one of the most important elements of music
    Many people listen to music with a focus on singing
  - Attract attention not only from a scientific point of view but also from the standpoint of industrial applications
    - Automatic singing pitch correction (e.g., *"Auto-Tune"*)
    - Singing synthesis (e.g., *"VOCALOID"*)
    - Query-by-humming (e.g., *"midomi"*)
    - Singing skill evaluation for *Karaoke*
  - The concept is broad and still emerging



Singing Information Processing Systems

- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

Singing Information Processing Systems

- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

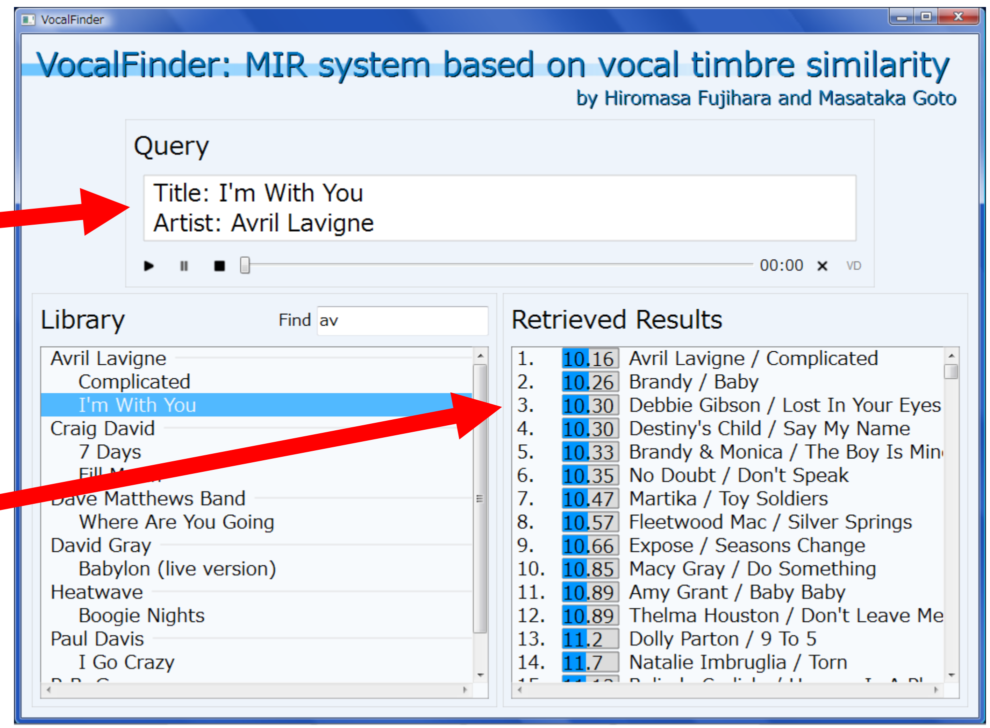
VocalFinder

[Fujihara, Goto, 2007-]

- **Music information retrieval based on singing voice timbre**
  - Retrieve and discover songs that have vocal timbre similar to a query song

Query song

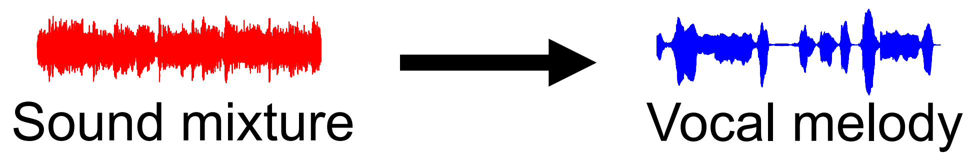
Retrieved songs



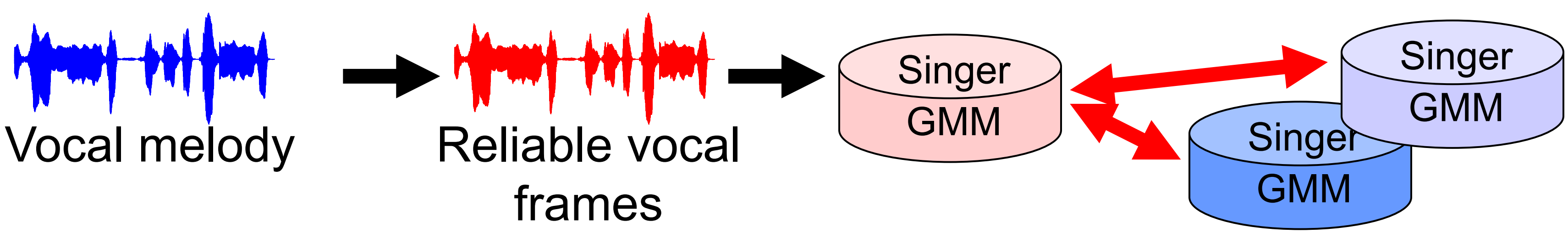
Avril Lavigne: I'm With You
Britney Spears: Oops!
Celine dion: Because You Loved Me

VocalFinder: Technology

- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*

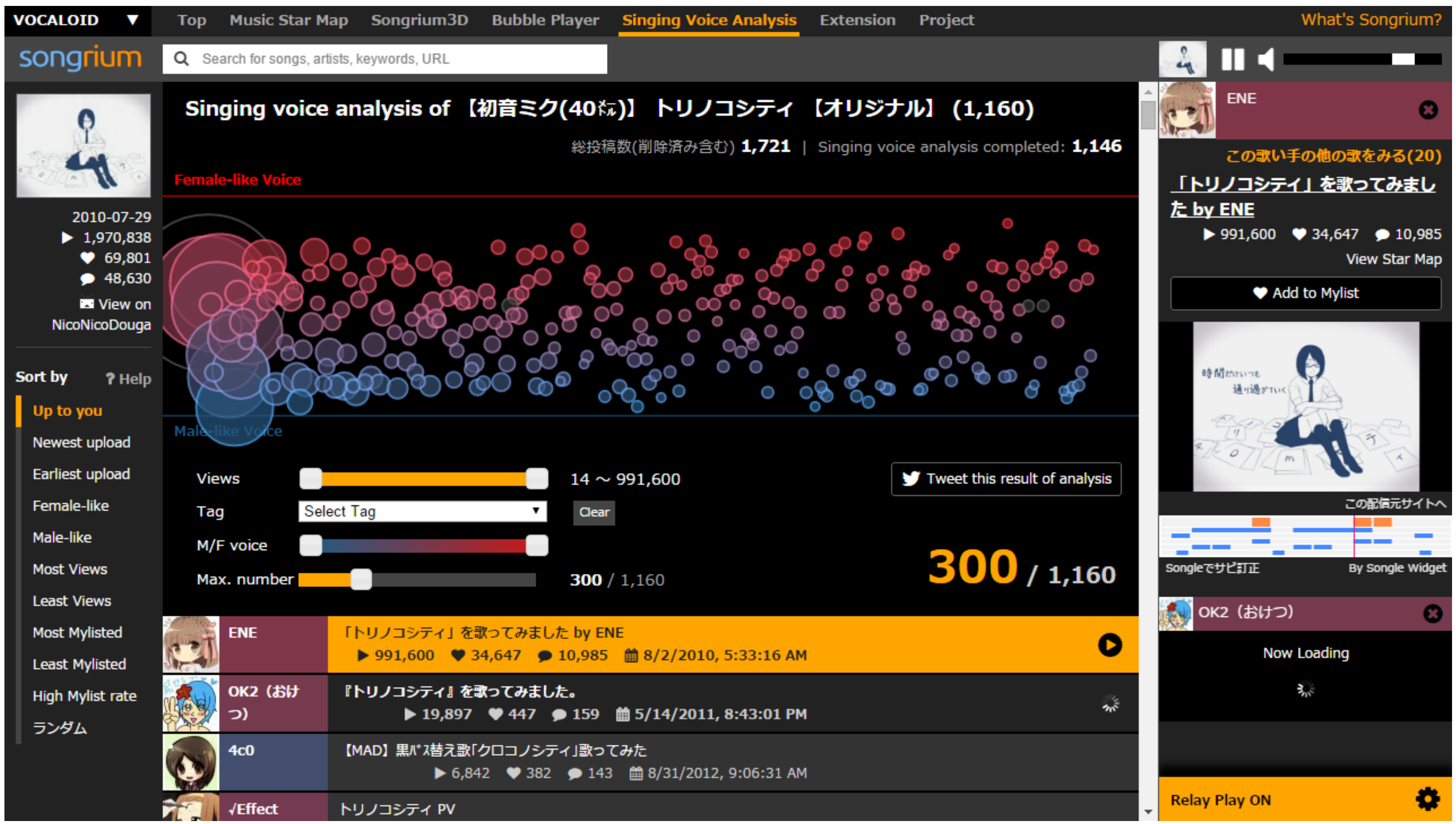


- **Vocal timbre modeling method** [Fujihara, et al., 2005-]
  - Train singer GMM for each song by using feature vectors on reliable vocal frames



Male/Female Estimation

- ❑ **Music browsing based on male/female estimation**
  - Visualize a song collection by using male/female estimation



You can try this at **http://songrium.jp/sings**

Male/Female Estimation: Technology

- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*

- **Male/Female modeling method**
  - Train male/female SVM classifier by using feature vectors on reliable vocal frames

Singer ID

[Fujihara, *et al.*, 2005-]

❑ Singer identification (ID) for polyphonic music recordings

- Identify the name of the singer who sang the input song

  Similar to speaker recognition

- You can retrieve a song without metadata

Song: *"Open Arms"*
Artist: *Journey*

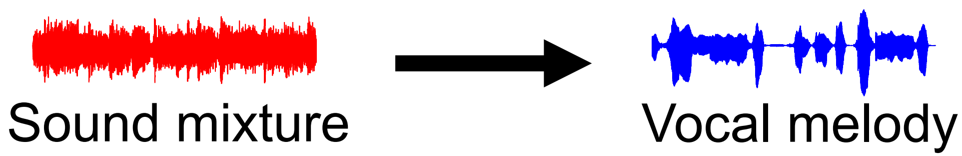
Singer name:
Steve Perry

Song: *"With or Without You"*
Artist: *U2*

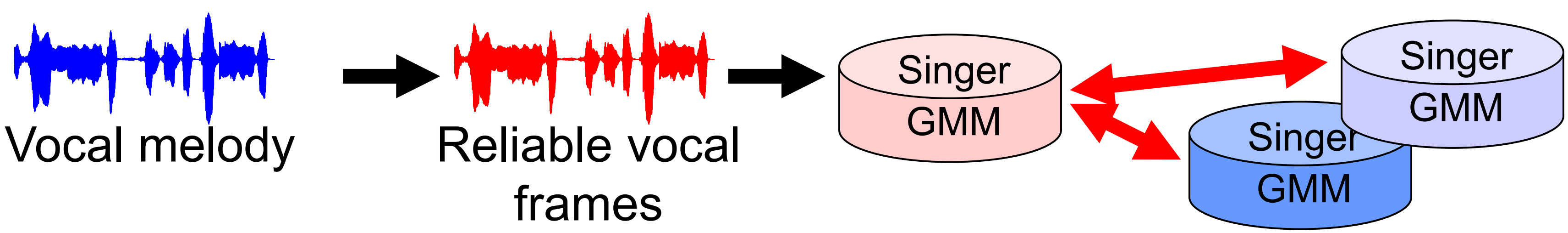
Singer name:
Bono

Singer ID: Technology

- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*



- **Vocal timbre modeling method** [Fujihara, et al., 2005-]
  - Train singer GMM for each singer by using feature vectors on reliable vocal frames



Singing Information Processing Systems

- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

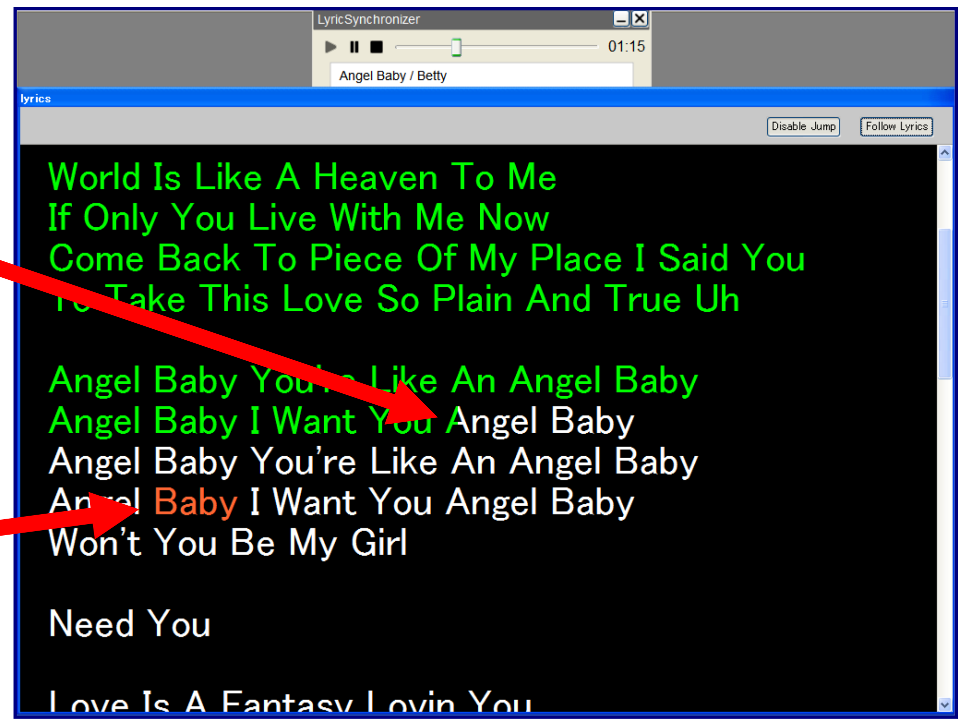
LyricSynchronizer

[Fujihara, Goto, Okuno, 2006-]

- **Automatic synchronization of lyrics with polyphonic music recordings**
  - Display scrolling lyrics with the phrase currently being sung highlighted during playback of a song

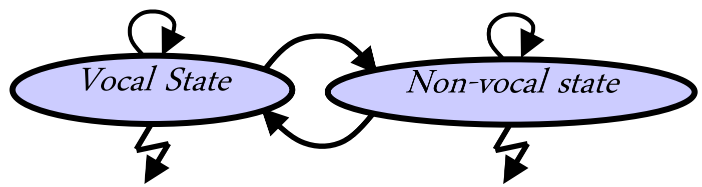
The current playback position

You can listen from a clicked word



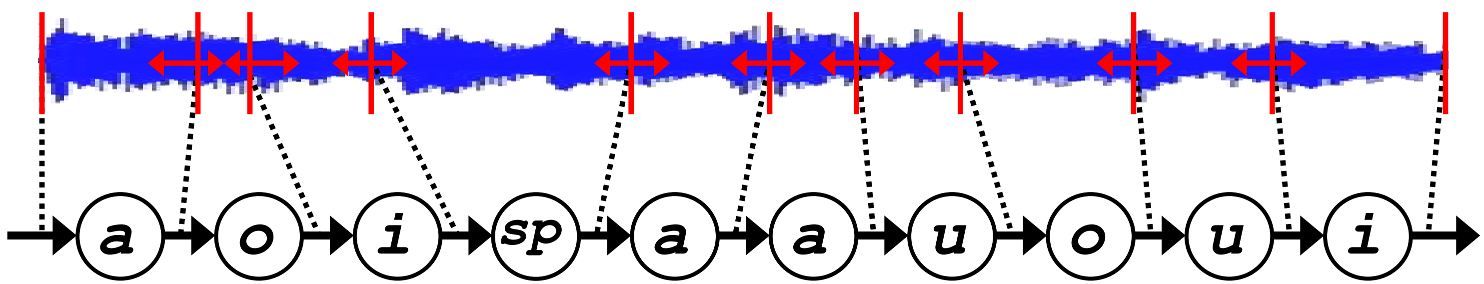
LyricSynchronizer: Technology

- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*
- **Automatic lyrics synchronization method** [Fujihara, Goto, Okuno, 2006-]
  - Detect vocal sections by using HMM



Vocal Activity Detection (VAD)

  - Locate each phoneme in resynthesized vocal melody by using the Viterbi (forced) alignment technique



Lyric Synchronization: References

- A. Loscos, P. Cano and J. Bonada, "Low-delay singing voice alignment to text," in Proc. of ICMC 99, 1999.
- Y. Wang, M. Kan, T. Nwe, A. Shenoy, and J. Yin, “Lyrically: automatic synchronization of acoustic musical signals and textual lyrics,” in Proceedings of ACM Multimedia 2014, pp.212-219, 2004.
- C. H. Wong, W. M. Szeto and K. H. Wong, "Automatic lyrics alignment for cantonese popular music," Multimedia Systems, vol.4-5, no. 12, pp.307-323, 2007.
- M. Muller, F. Kurth, D. Damm, C. Fremerey and M. Clausen, "Lyrics-based audio retrieval and multimodal navigation in music collections," in Proc. of ECD L 2007, pp.112-123, 2007.
- K. Lee and M. Cremer, “Segmentation-based lyrics-audio alignment using dynamic programming.,” in Proc. ISMIR, 2008, pp.395-400.
- A. Mesaros and T. Virtanen, “Automatic alignment of music audio and lyrics,” in Proc. DAFx, 2008, pp.321-324.
- M.-Y. Kan, Y. Wang, D. Iskandar, T. L. Nwe and A. Shenoy, "Lyrically: Automatic synchronization of textual lyrics to acoustic music signals," IEEE Transactions on Audio, Speech, and Language Processing, vol.16, no. 2, pp.338-349, 2008.

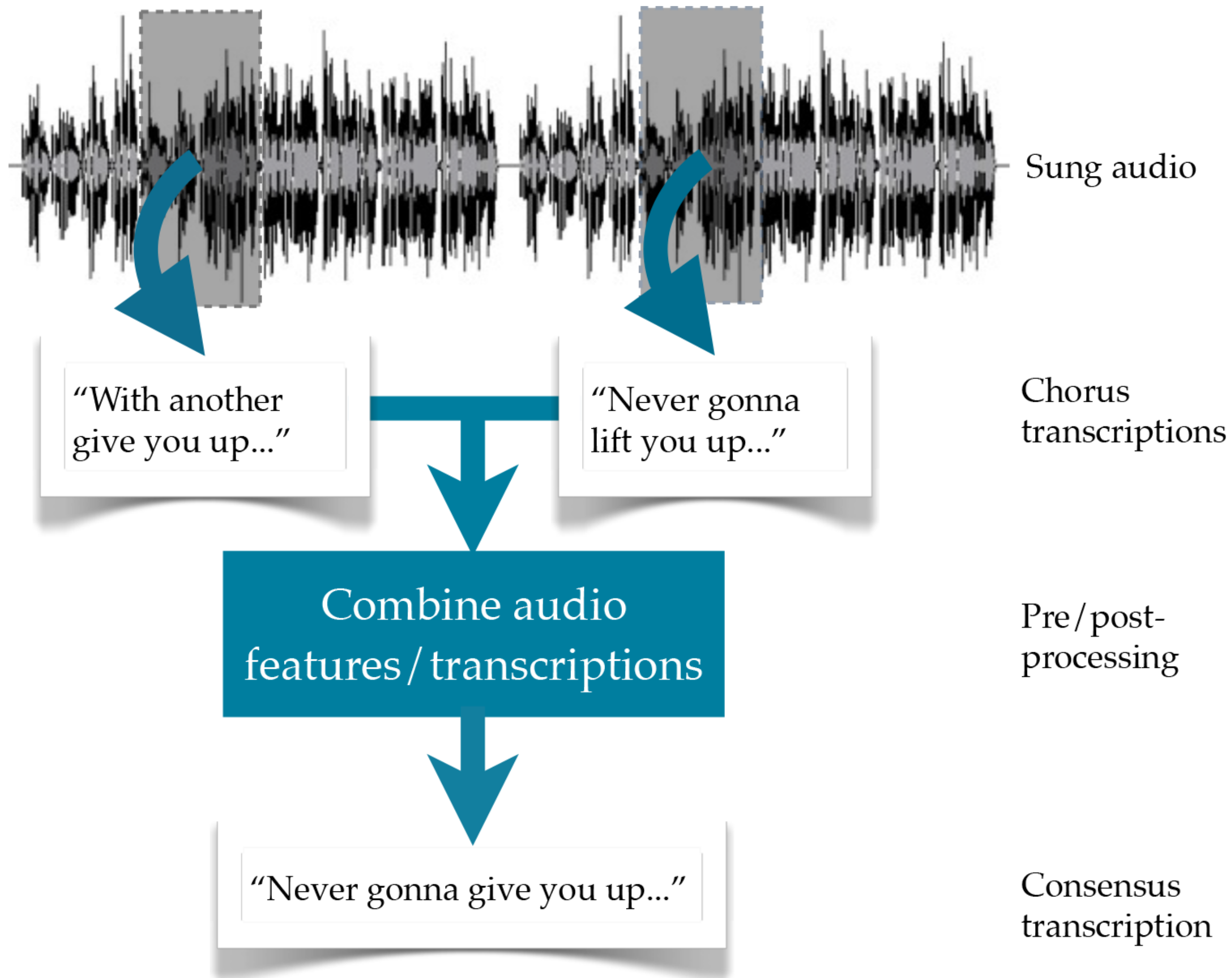
Lyric Synchronization: References

- K. Chen, S. Gao, Y. Zhu and Q. Sun, "Popular song and lyrics synchronization and its application to music information retrieval," in Proc. ofMMCN'06, 2006.
- H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata and H. G. Okuno, "Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals," in Proc. of ISM 2006, pp.257-264, 2006.
- D. Iskandar, Y. Wang, M.-Y. Kan and H. Li, "Syllabic level automatic synchronization of music signals and text lyrics," in Proc. ACM Multimedia 2006, pp.659-662, 2006.
- H. Fujihara and M. Goto, "Three techniques for improving automatic synchronization between music and lyrics: Fricative detection, filler model, and novel feature vectors for vocal activity detection," in Proc. of ICASSP 2008, 2008.
- H. Fujihara, M. Goto, J. Ogata and H. G. Okuno, "LyricSynchronizer: Automatic synchronization system between musical audio signals and lyrics," IEEE Journal of Selected Topics in Signal Processing, vol.5, no. 6, pp.1252-1261, 2011.
- M. Mauch, H. Fujihara and M. Goto, "Integrating additional chord information into HMM-based lyrics-to-audio alignment," IEEE Transactions on Audio, Speech, and Language Processing, vol.20, no. 1, pp.200-210, 2012.

Lyric Transcription

[McVicar, Ellis, Goto, 2014]

- **Automatic transcription of lyrics**
  - Use repeated choruses to improve automatic lyric recognition (solo sung voice)



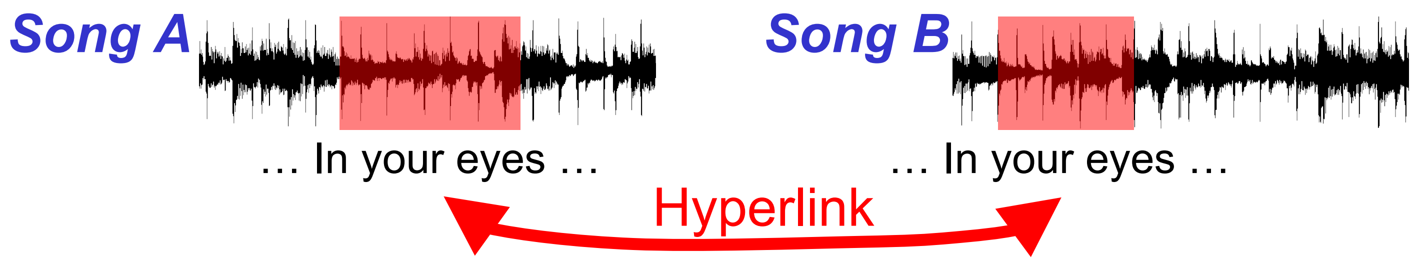
Lyric Transcription: References

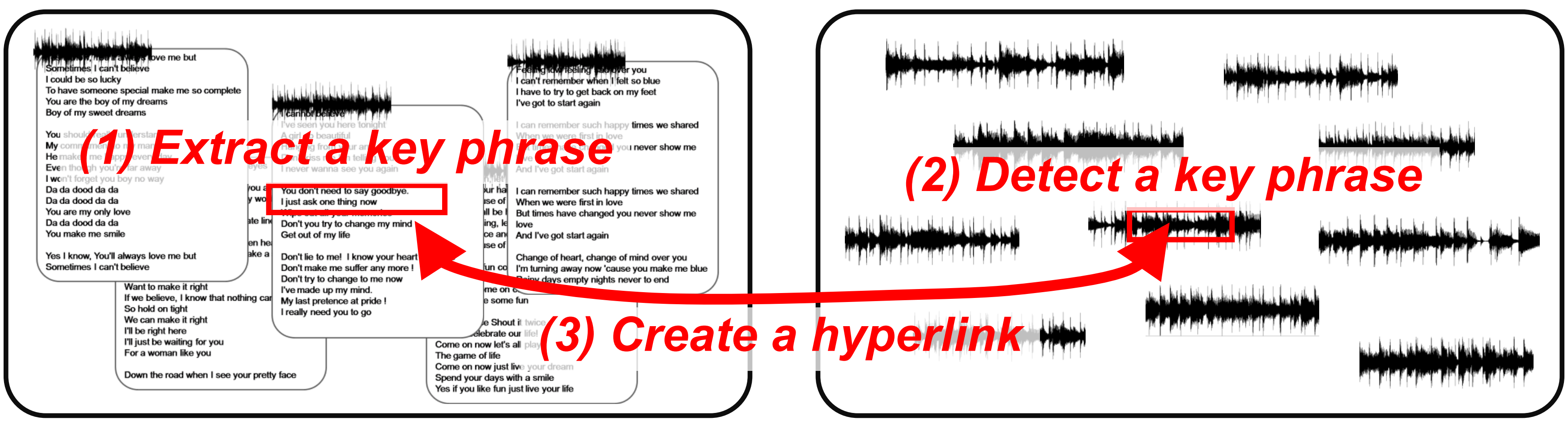
- C.-K. Wang, R. -Y. Lyu and Y.-C. Chiang, "An automatic singing transcription system with multilingual singing lyric recognizer and robust melody tracker," in Proc. of Eurospeech 2003, pp.1197-1200, 2003.
- A. Sasou, M. Goto, S. Hayamizu and K. Tanaka, "An autoregressive, non-stationary excited signal parameter estimation method and an evaluation of a singing-voice recognition," in Proc. of ICASSP 2005, pp.1-237-240, 2005.
- M. Suzuki, T. Hosoya, A. Ito and S. Makino, "Music information retrieval from a singing voice using lyrics and melody information," EURASIP Journal on Advances in Signal Processing, vol.2007, 2007.
- A. Mesaros and T. Virtanen, “Automatic recognition of lyrics in singing,” EURASIP Journal on Audio, Speech, and Music Processing, vol.2010, 2010.
- A. Mesaros and T. Virtanen, “Recognition of phonemes and words in singing,” in Proceedings of ICASSP 2010, pp.2146-2149, 2010.
- M. McVicar, D. P. Ellis and M. Goto, "Leveraging repetition for improved automatic lyric transcription in popular music," in Proc. of ICASSP 2014, pp.3141-3145, 2014.
- M. Gruhne, K. Schmidt and C. Dittmar, "Phoneme recognition in popular music," in Proc. of ISMIR 2007, pp.369-370, 2007.
- W.-H. Tsai and H.-M. Wang, "Automatic identification of the sung language in popular music recordings," J New Music Res., vol.36, no. 2, pp.105-114, 2007.

Hyperlinking Lyrics

[Fujihara, Goto, Ogata, 2008-]

- **Creating hyperlinks between phrases in lyrics**
  - Create a hyperlink from a phrase in the lyrics of a song to the same phrase in the lyrics of another song





Songs with text lyrics **Songs without text lyrics**

Hyperlinking Lyrics: Technology

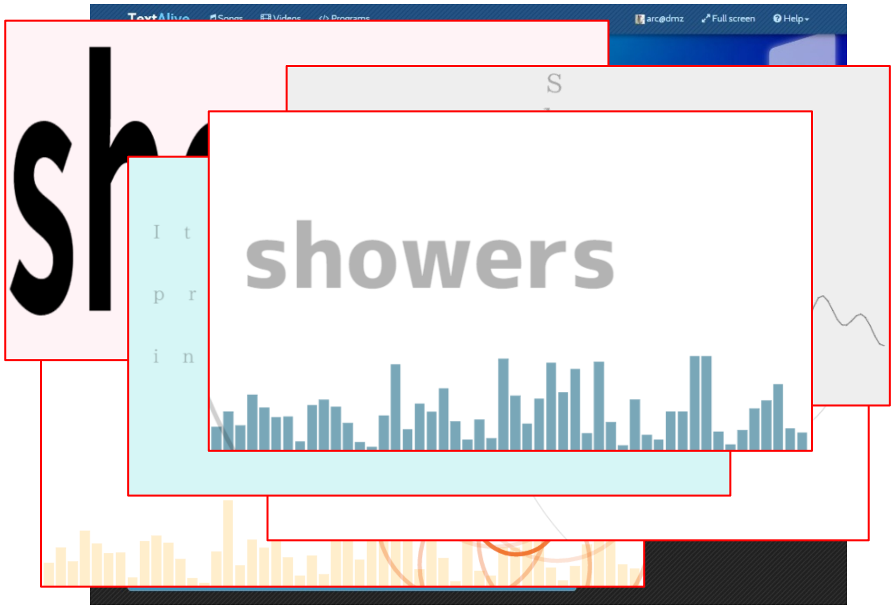
- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*

Sound mixture → Vocal melody

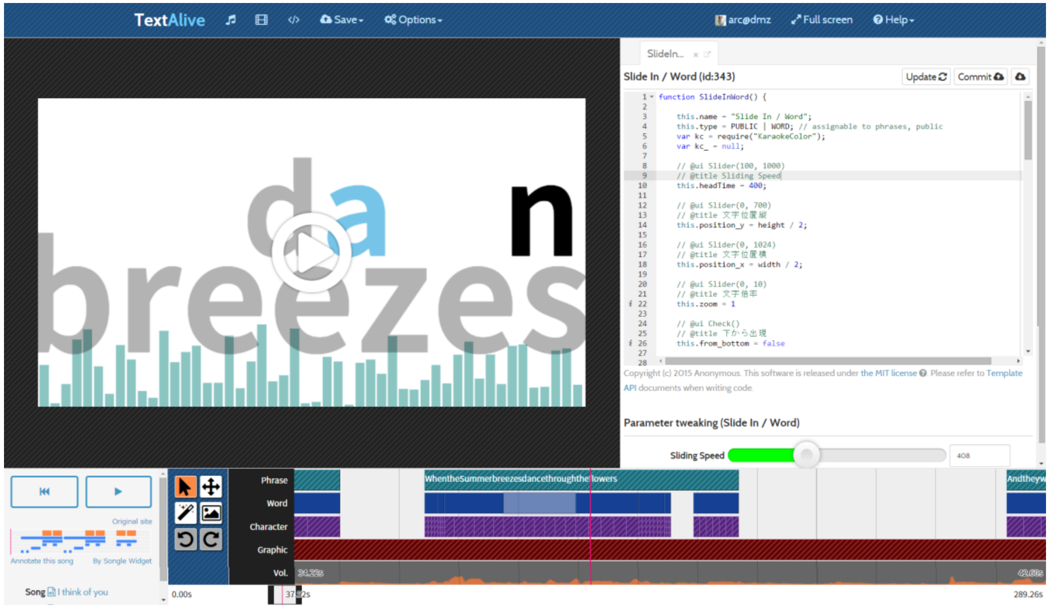
- **Keyword spotting method** [Fujihara, Goto, Ogata, 2008-]
  - Locate each key phrase in extracted vocal melody by using the Viterbi alignment with phoneme HMMs

Lyric Animation: TextAlive [Kato, Nakano, Goto, 2014-]

- **Interactive editing of lyrics animation based on automatic video composition on the web browser**
  - Reduce manual labors: **music understanding** techniques
  - Compose videos on-the-fly: **live programming** techniques



Change animation styles in just one click



Edit animation details interactively with intuitive user interfaces

Lyric Animation: TextAlive [Kato, Nakano, Goto, 2014-]

- **Interactive editing** of lyrics animation based on **automatic video composition** on the web browser
  - Reduce manual labors: **music understanding** techniques
  - Compose videos on-the-fly: **live programming** techniques

Available on http://textalive.jp

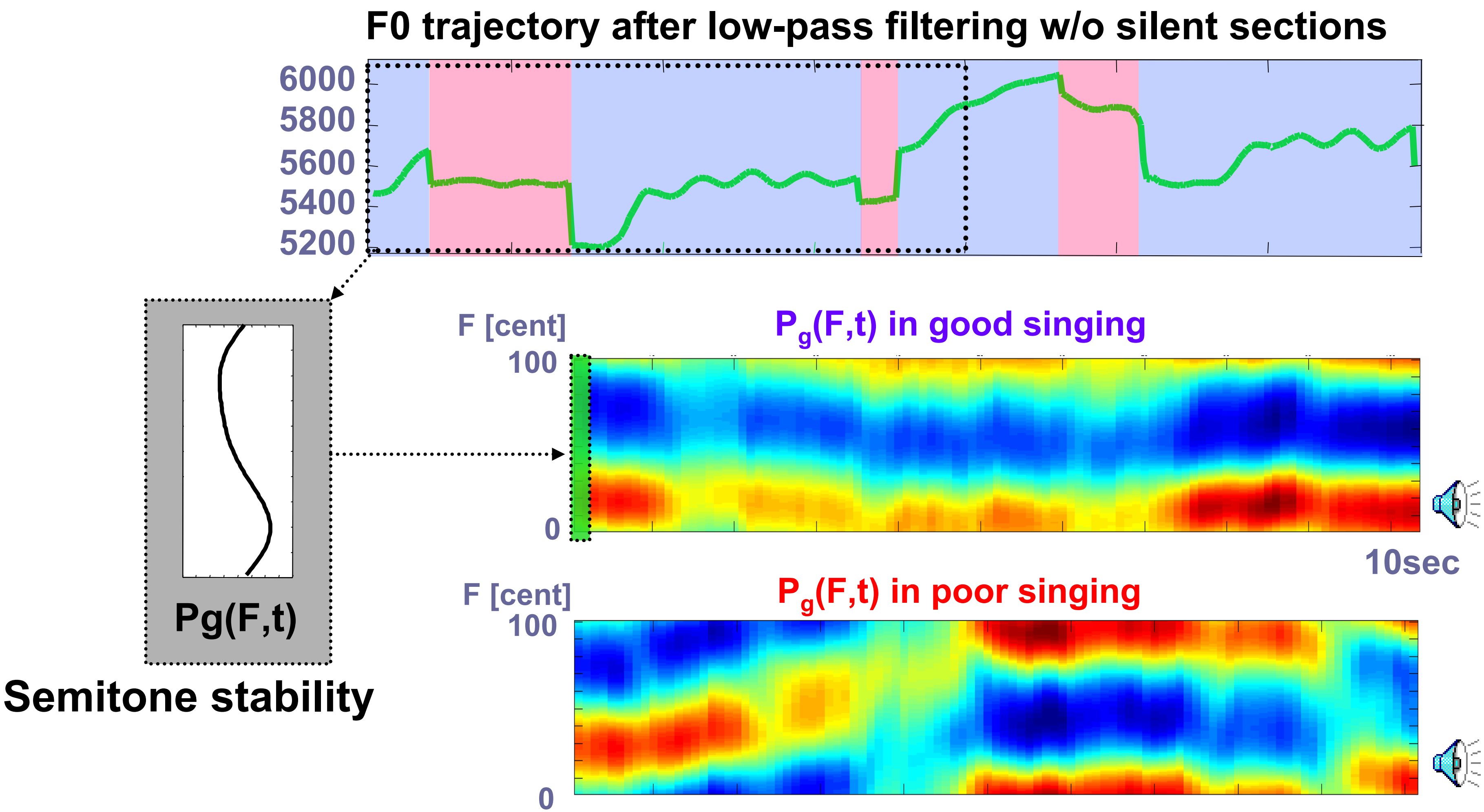
Singing Information Processing Systems

- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

Singing Skill Evaluation

[Nakano, Goto, Hiraga, 2006]

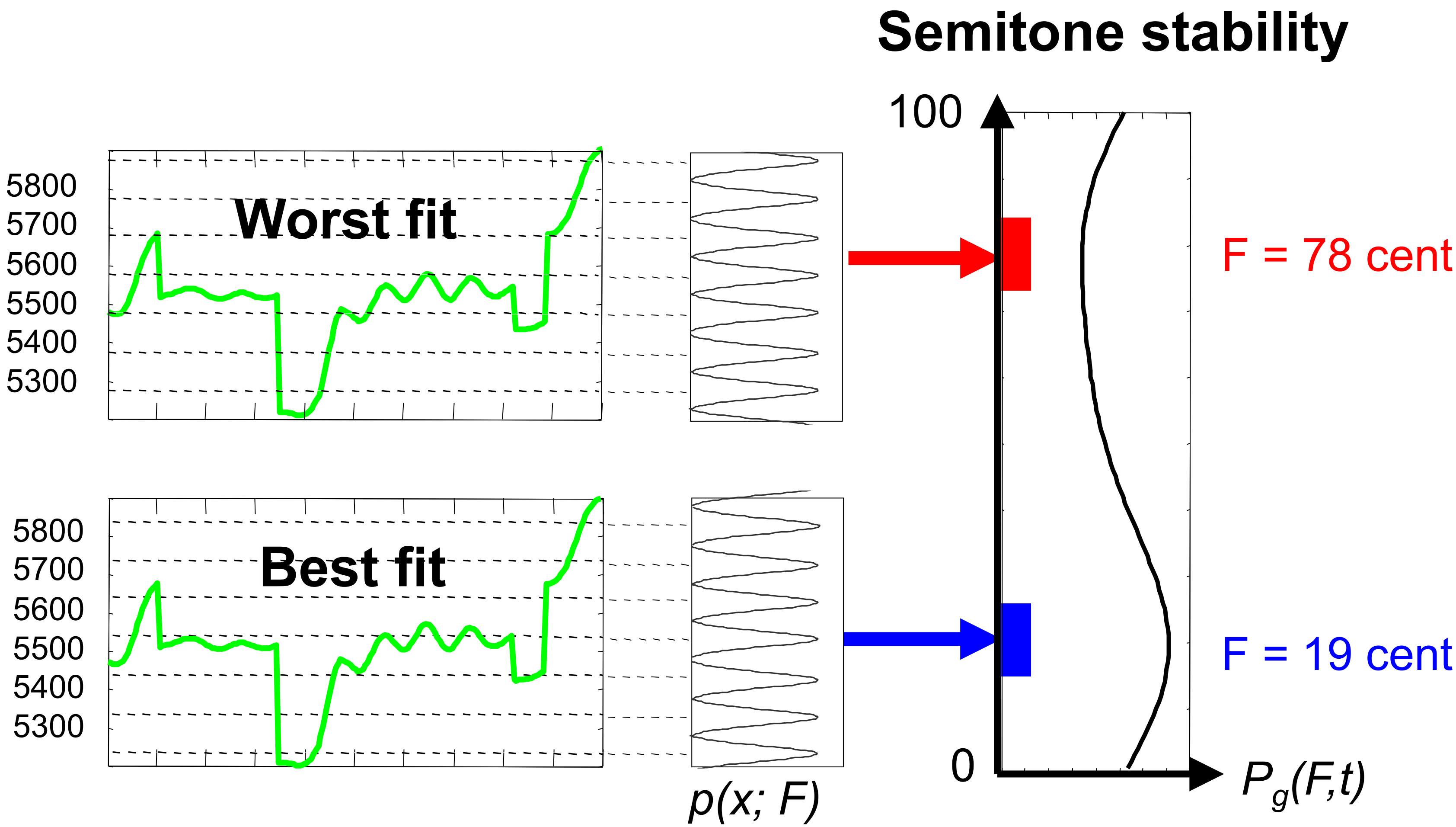
- **Evaluate pitch interval accuracy w/o score**



Singing Skill Evaluation

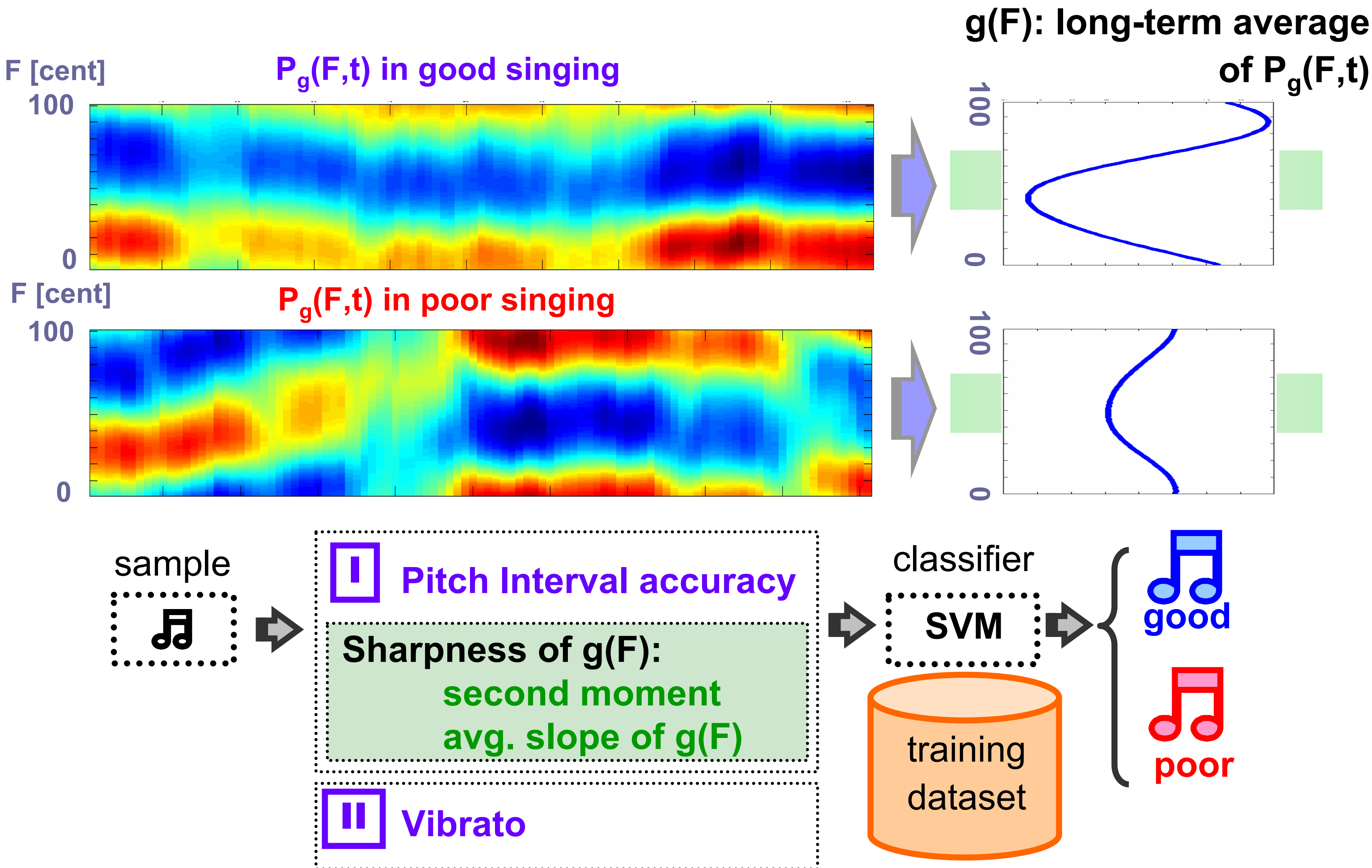
[Nakano, Goto, Hiraga, 2006]

- Evaluate pitch interval accuracy w/o score



Singing Skill Evaluation

[Nakano, Goto, Hiraga, 2006]

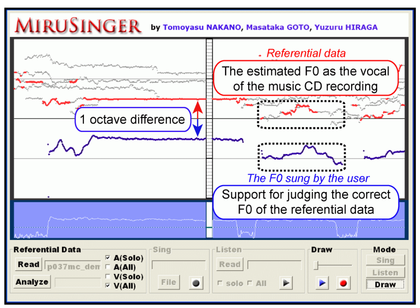


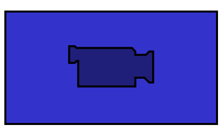
MiruSinger

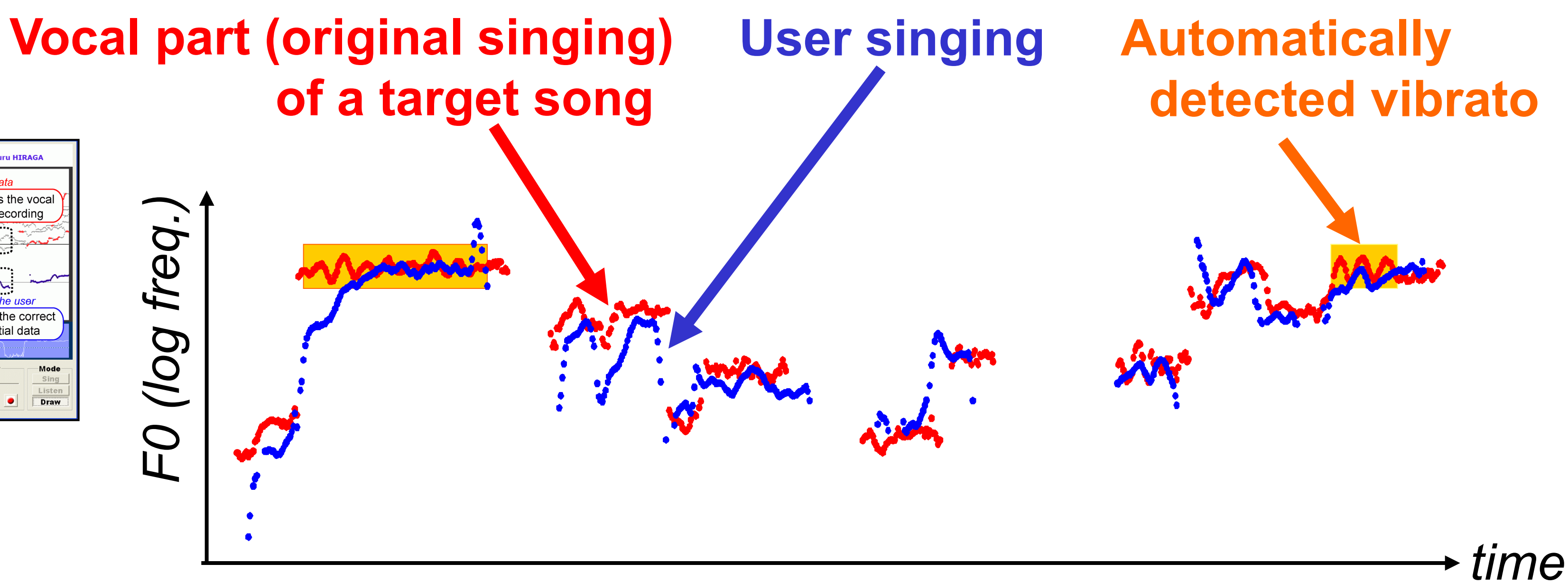
[Nakano, Goto, Hiraga, 2007-]

❑ Singing skill visualization and training

- Help you imitate the vocal part of a target song
- Analyze and visualize vocal singing with reference to the vocal part of a target song
- Real-time feedback of F0 and vibrato sections

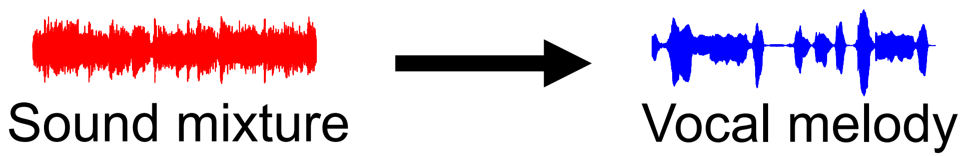






MiruSinger: Technology

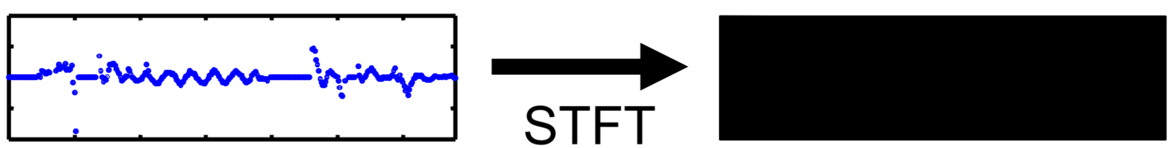
- **Automatic vocal extraction method** [Goto, 1999-]
  - Segregate vocal melody from polyphonic sound mixtures by using predominant-F0 estimation method *PreFEst*



- **Automatic Vibrato Detection Method**

[Nakano, Goto, Hiraga, 2006-]

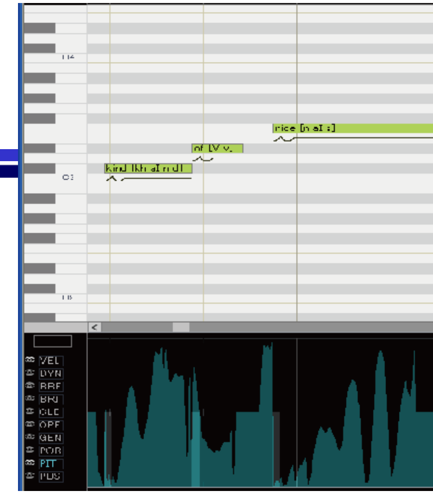
  - Calculate vibrato likeliness by using STFT of delta F0



Singing Information Processing Systems

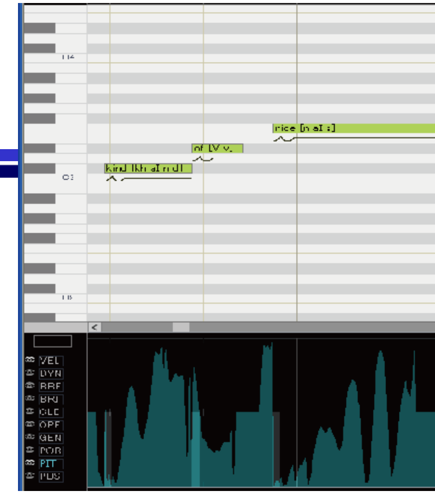
- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

Singing Synthesis



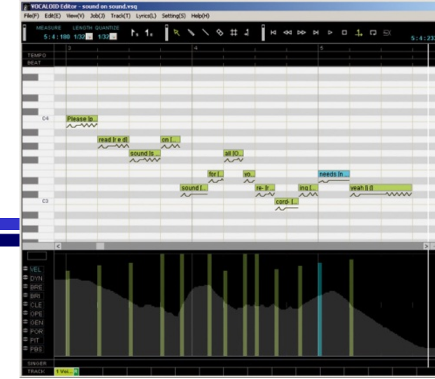
- **Text-to-singing synthesis**
  - Input: Note-level score information with its lyrics
    + Singing synthesis parameters
    such as pitch (F0) and dynamics (power)
- **Speech-to-singing synthesis**
  - Input: Speaking voice reading the lyrics of a song
- **Singing-to-singing synthesis**
  - Input: Singing voice singing the lyrics of a song

Singing Synthesis



- **Text-to-singing synthesis**
  - Input: Note-level score information with its lyrics
    + Singing synthesis parameters
    such as pitch (F0) and dynamics (power)
- **Speech-to-singing synthesis**
  - Input: Speaking voice reading the lyrics of a song
- **Singing-to-singing synthesis**
  - Input: Singing voice singing the lyrics of a song

Text-to-Singing Synthesis

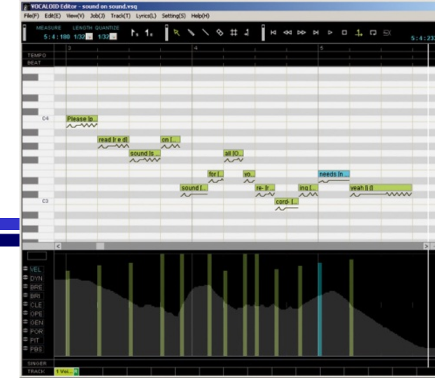


❑ **Singing synthesis engine “VOCALOID2”**

- Singing synthesis software ***“Hatsune Miku”*** was released on August 31st, 2007
- Virtual singer was embodied (illustrated) by a cartoon girl

  This has inspired many people to create, share, and remix

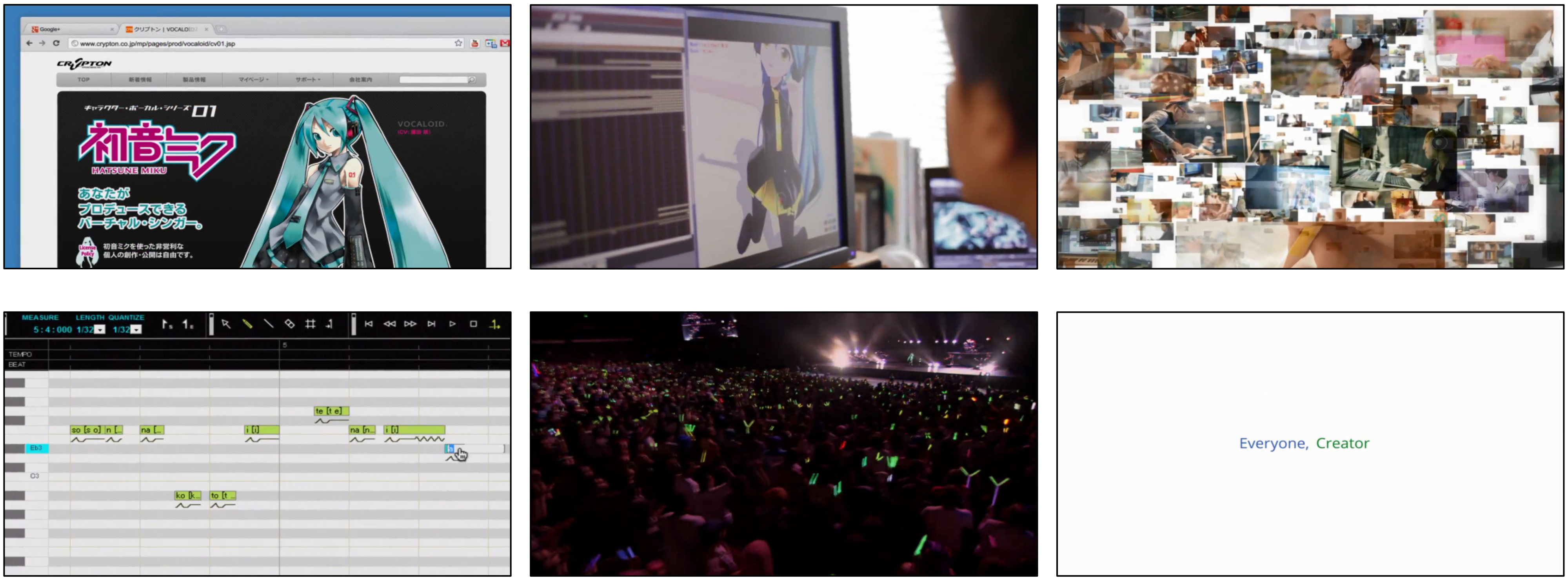
Text-to-Singing Synthesis



- **Singing synthesis engine "VOCALOID2"**
  - Singing synthesis software ***"Hatsune Miku"*** was released on August 31st, 2007
  - Virtual singer was embodied (illustrated) by a cartoon girl
    This has inspired many people to create, share, and remix
  - Both amateur and professional musicians started using singing synthesizers as their main vocals
  - A lot of different voices have already been on the market

Hatsune Miku Phenomenon

- **Commercial film of "Google Chrome" browser**
  - 1 minute introduction of *Hatsune Miku* Phenomenon



Bronze Lion Award, Cannes Lions International Festival of Creativity, June 2012

Used with permission from Crypton Future Media, INC.

Hatsune Miku Phenomenon

❑ **Live concerts featuring *Hatsune Miku***

- Tokyo, Sapporo, Wakayama, Yokohama, Osaka, etc.
- Los Angeles, New York, Singapore, Hong Kong, Taipei, Jakarta, Shanghai, etc.
- Opening act for Lady Gaga's USA concert tour in 2014

MIKUNOPOLIS 2011

SNOW MIKU 2015

Hatsune Miku Phenomenon

- **Live concerts featuring *Hatsune Miku***
  - Tokyo, Sapporo, Wakayama, Yokohama, Osaka, etc.
  - Los Angeles, New York, Singapore, Hong Kong, Taipei, Jakarta, Shanghai, etc.
  - Opening act for Lady Gaga’s USA concert tour in 2014
  - US television debut “Late Show with David Letterman” in 2014
  - Hatsune Miku’s Opera at Théâtre du Châtelet in Paris in 2013

Hatsune Miku Phenomenon

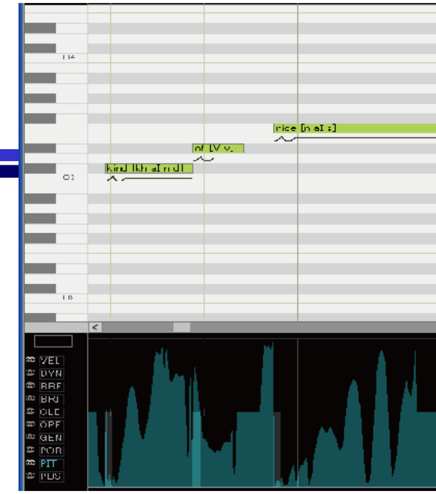
- The most surprising change

Singing synthesis breaks down the long-cherished view that "listening to a non-human singing voice is worthless", emerging the "culture in which people actively enjoy songs with synthesized singing voices as the main vocals"

Future of Singing Synthesis

- **Music technologies have already changed music cultures in the history of music**
  - The piano and guitars were brand-new technologies when people started using them
- **Sound synthesis**
  - Sound synthesizers are widely used and have become indispensable to popular music production
- **Singing synthesis**
  - There is no reason that the same will not happen for singing
  - It is historically inevitable that singing synthesizers will become widely used worldwide and likewise indispensable

Singing Synthesis



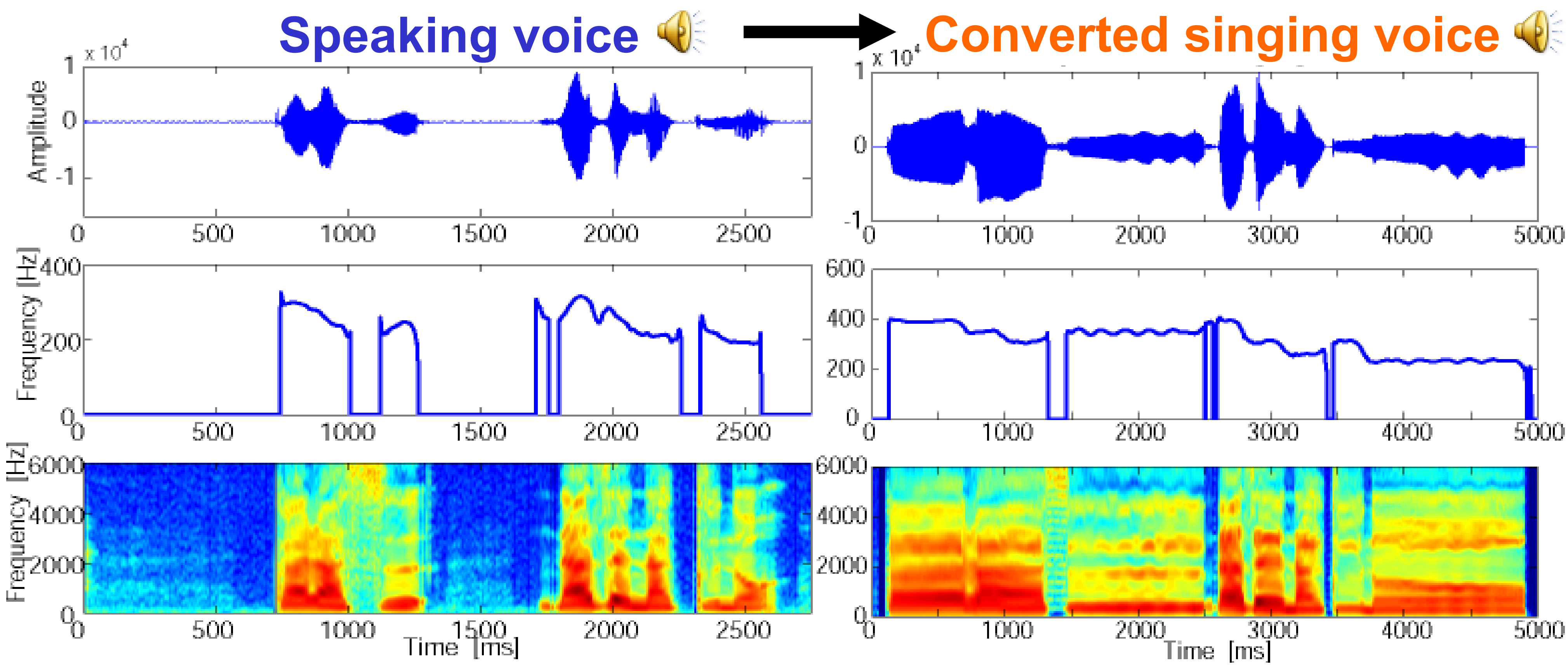
- Text-to-singing synthesis
  - Input: Note-level score information with its lyrics
    + Singing synthesis parameters
    such as pitch (F0) and dynamics (power)
- Speech-to-singing synthesis
  - Input: Speaking voice reading the lyrics of a song
- Singing-to-singing synthesis
  - Input: Singing voice singing the lyrics of a song

[Saitou, Goto, 2007-]

Speech-to-Singing Synthesis

❑ SingBySpeaking

- Convert a **speaking voice** to a **singing voice** by changing F0, phoneme duration, and singing formant



SingBySpeaking [Saitou, Goto, 2007-]

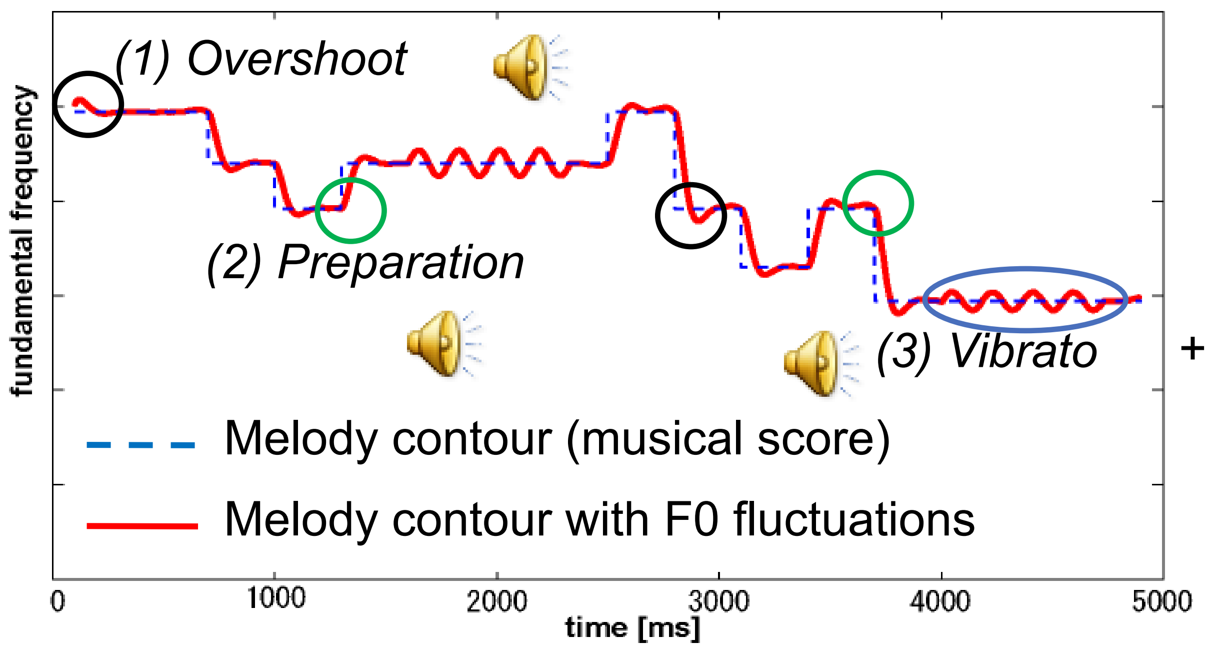
❑ Automatic Lyrics Synchronization Method

- Locate each phoneme in the speaking voice
  by using the Viterbi alignment with phoneme HMMs

❑ F0 Contour Generation Method

- Add four types of F0 fluctuations on musical notes

Musical score

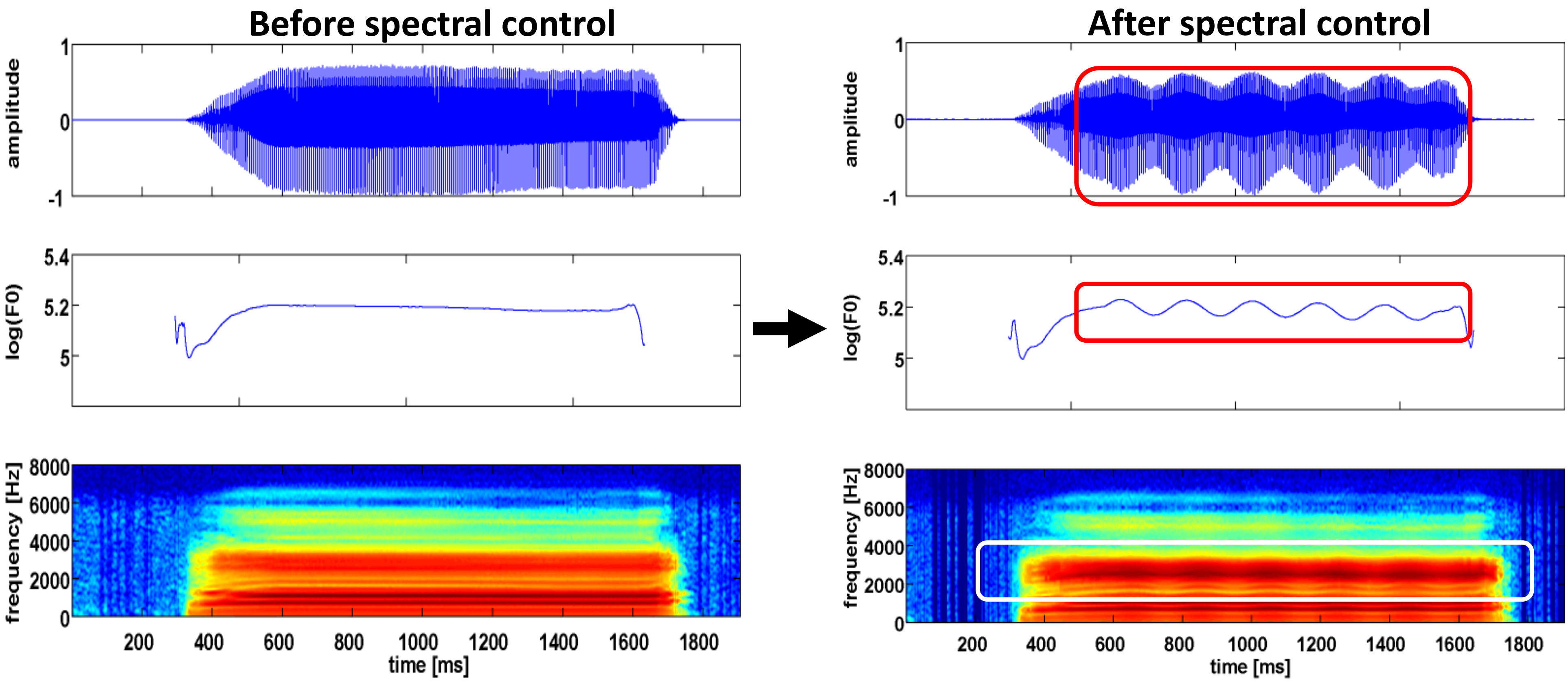


All

+ (4) Fine fluctuations

SingBySpeaking [Saitou, Goto, 2007-]

❑ Spectral Control Method



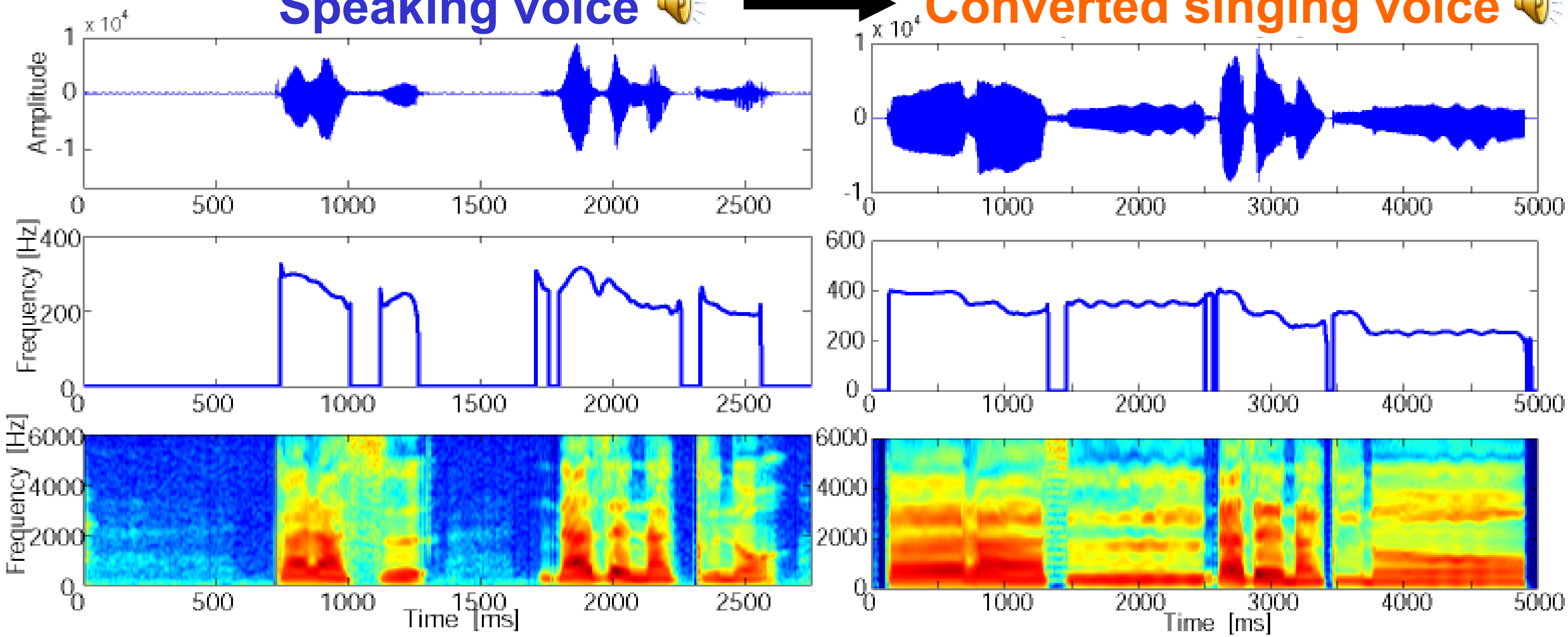
Original input   + *Singer's formant*   + *Vibrato & AM*   + *All*

SingBySpeaking [Saitou, Goto, 2007-]

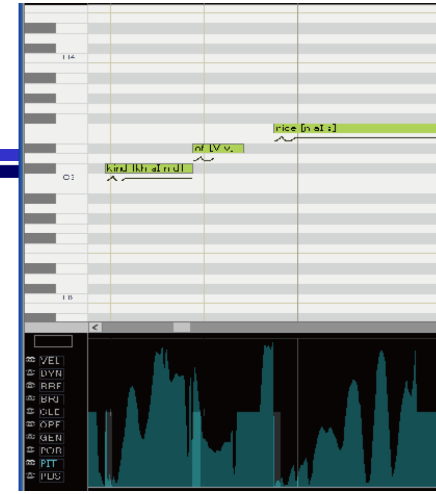
❑ Speech-to-singing synthesis

- Convert a **speaking voice** to a **singing voice** by changing F0, phoneme duration, and singing formant

Speaking voice → **Converted singing voice**



Singing Synthesis

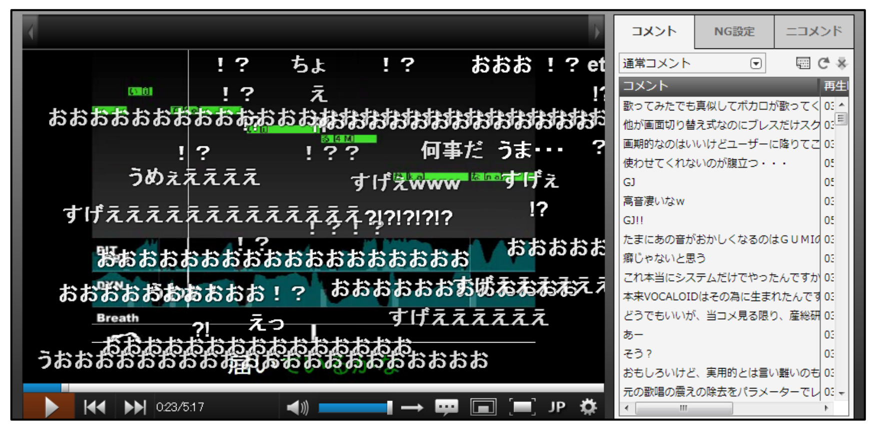


- **Text-to-singing synthesis**
  - Input: Note-level score information with its lyrics
    + Singing synthesis parameters
    such as pitch (F0) and dynamics (power)
- **Speech-to-singing synthesis**
  - Input: Speaking voice reading the lyrics of a song
- **Singing-to-singing synthesis**
  - Input: Singing voice singing the lyrics of a song

Singing-to-Singing Synthesis: VocaListener

"Packaged" by VocaListener

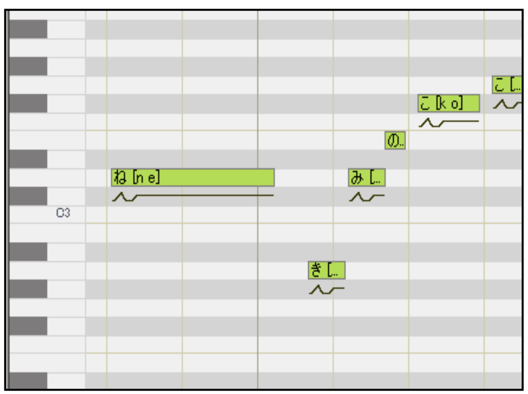
2010/10/04 [sm12320140]



[Nakano, Goto, 2008-]

What is VocaListener?

- **VocaListener** synthesizes natural singing voices by analyzing and imitating human singing
  - Imitate the pitch, dynamics, phoneme timing, and breath of the singer's voice
  - Estimate parameters of singing synthesizer *"VOCALOID"*



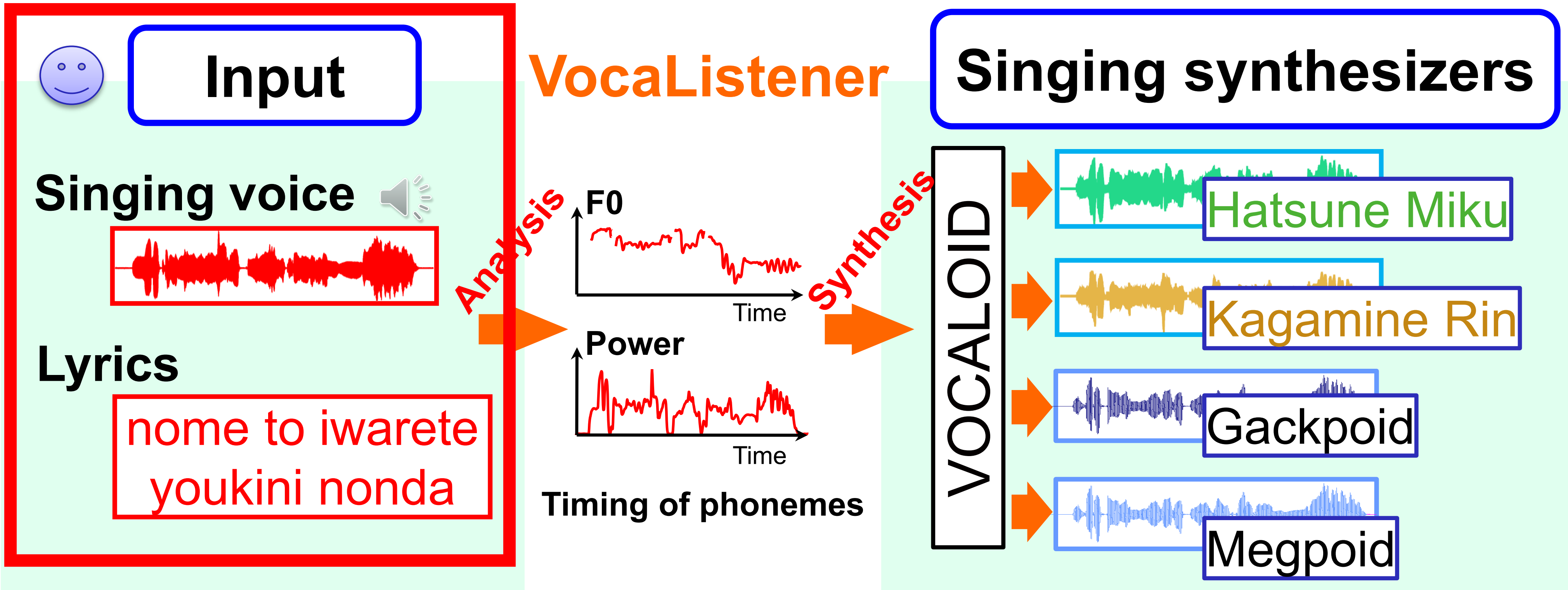
Original VOCALOID

VOCALOID + VocaListener

[Nakano, Goto, 2008-]

VocaListener

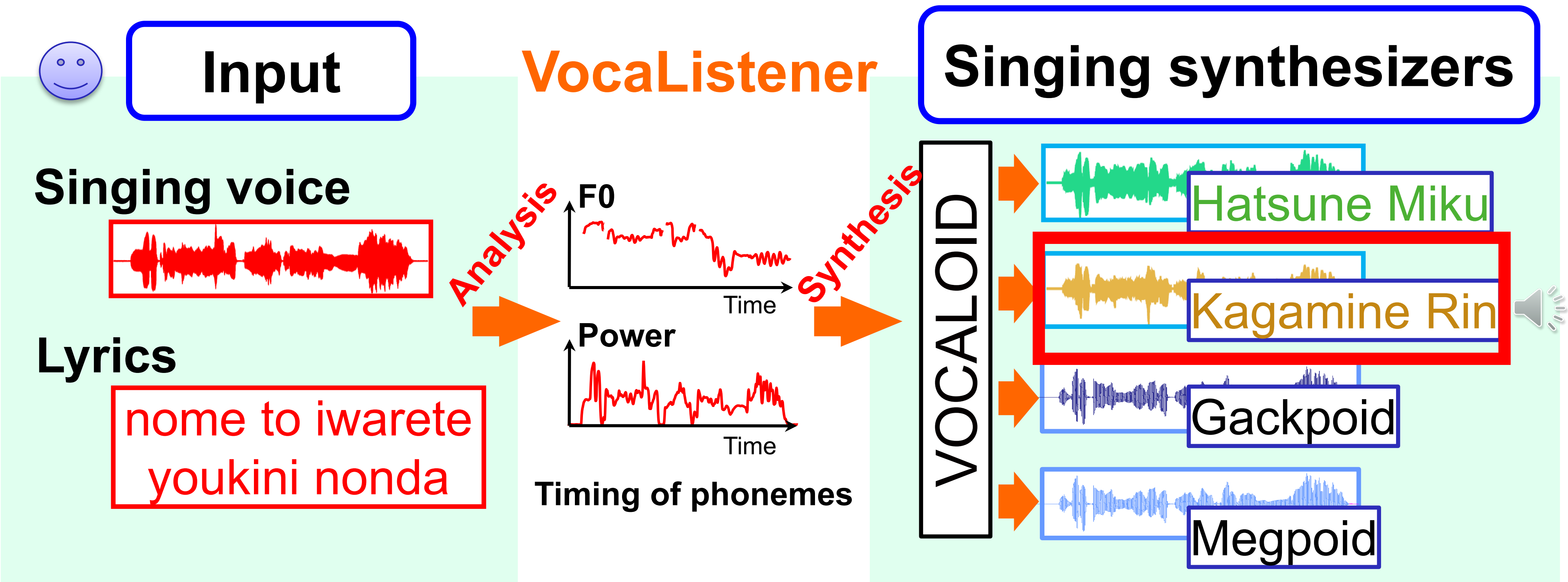
- Generate a musical score by analyzing the input singing voice
- Estimate synthesis parameters for each virtual singer



[Nakano, Goto, 2008-]

VocaListener

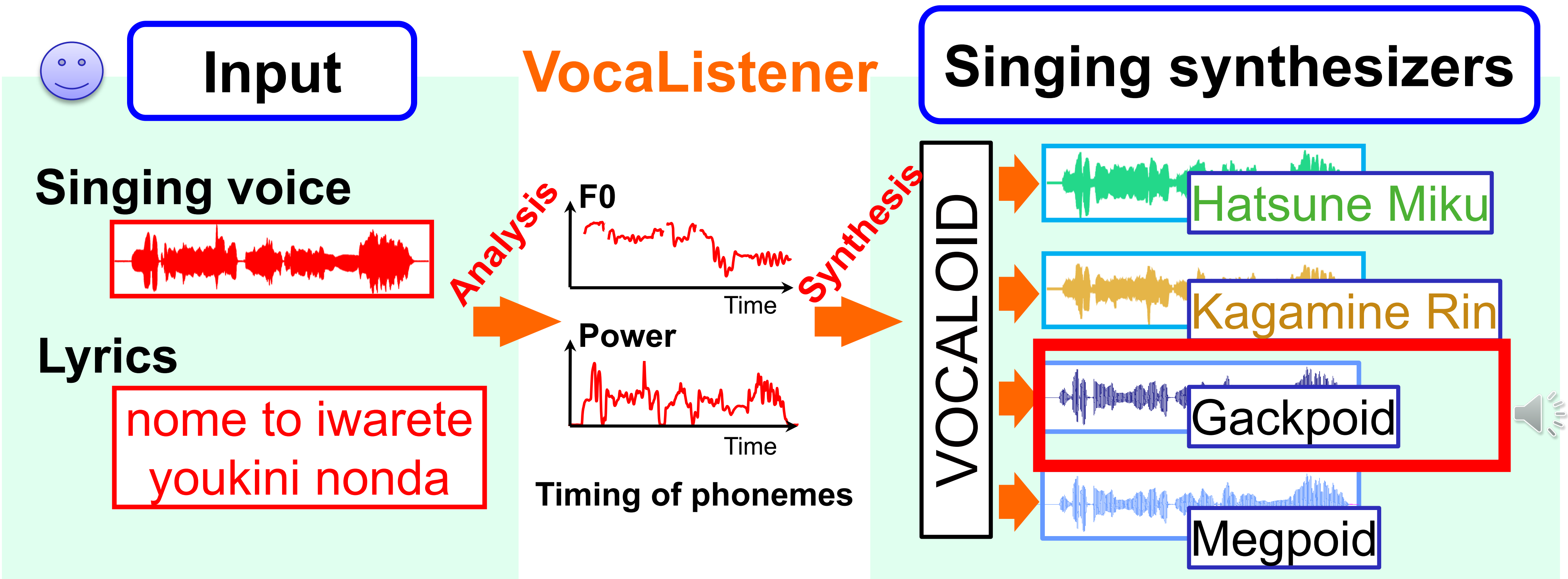
- Generate a musical score by analyzing the input singing voice
- Estimate synthesis parameters for each virtual singer



[Nakano, Goto, 2008-]

VocaListener

- Generate a musical score by analyzing the input singing voice
- Estimate synthesis parameters for each virtual singer

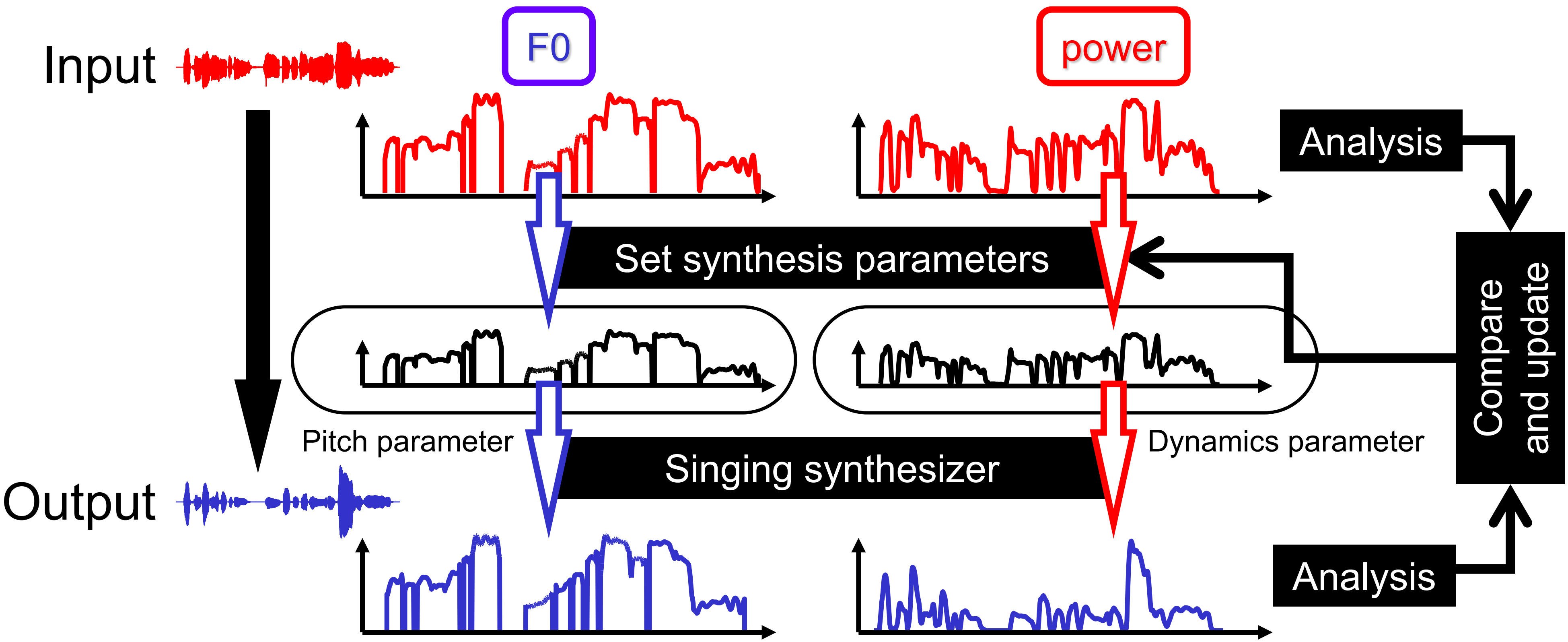


VocaListener

[Nakano, Goto, 2008-]

❑ Why is this difficult?

- Synthesized results are different because of singer DBs
- We needed iterative parameter estimation

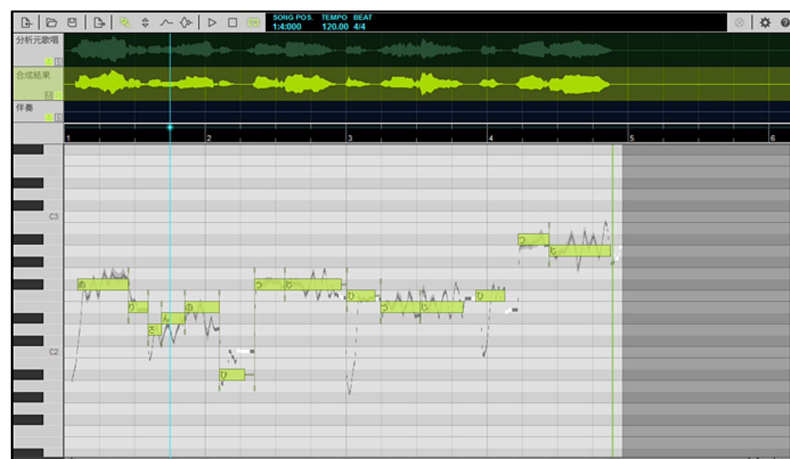


VocaListener as a Commercial Product

- 2011/06: Press release of a product version
  - As a Job Plugin of **VOCALOID3** by **YAMAHA Corp.**
- 2012/10: The product appears on the market
  - **“VOCALOID3 Job Plugin VocaListener”**
- 2015/08: The upgraded version is released
  - **“VOCALOID4 Job Plugin VocaListener”**

VocaListener

VOCALOID™4 Job Plugin




Sagashituzuketa FutarinohoNtono Prologue
探し続けた 二人のほんとの プロローグ

Robot Singer HRP-4C Miim

VocaListener + VocaWatcher

AIST, Japan

Shuuji Kajita, Tomoyasu Nakano,

Masataka Goto, Yosuke Matsusaka,

Shin'ichiro Nakaoka, and Kazuhito Yokoi

[Kajita, Nakano, Goto, Matsusaka, Nakaoka, Yokoi, 2010-]

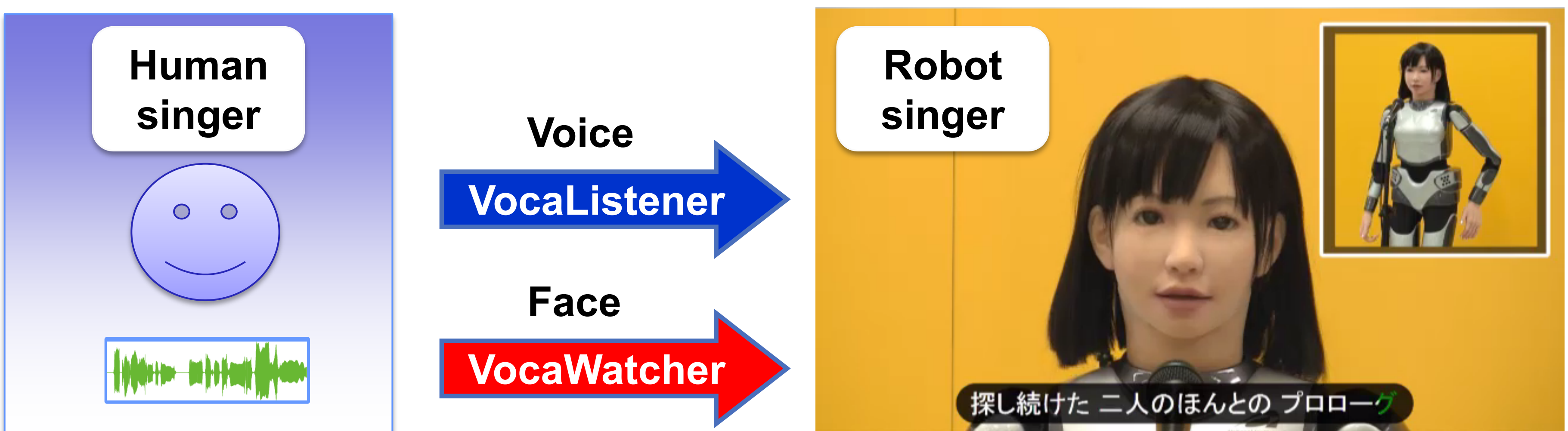
Humanoid Robot Singer: HRP-4C Miim

- Imitating a human singer
  - **VocaListener**

    Imitate vocal expressions to synthesize singing voices
  - **VocaWatcher**

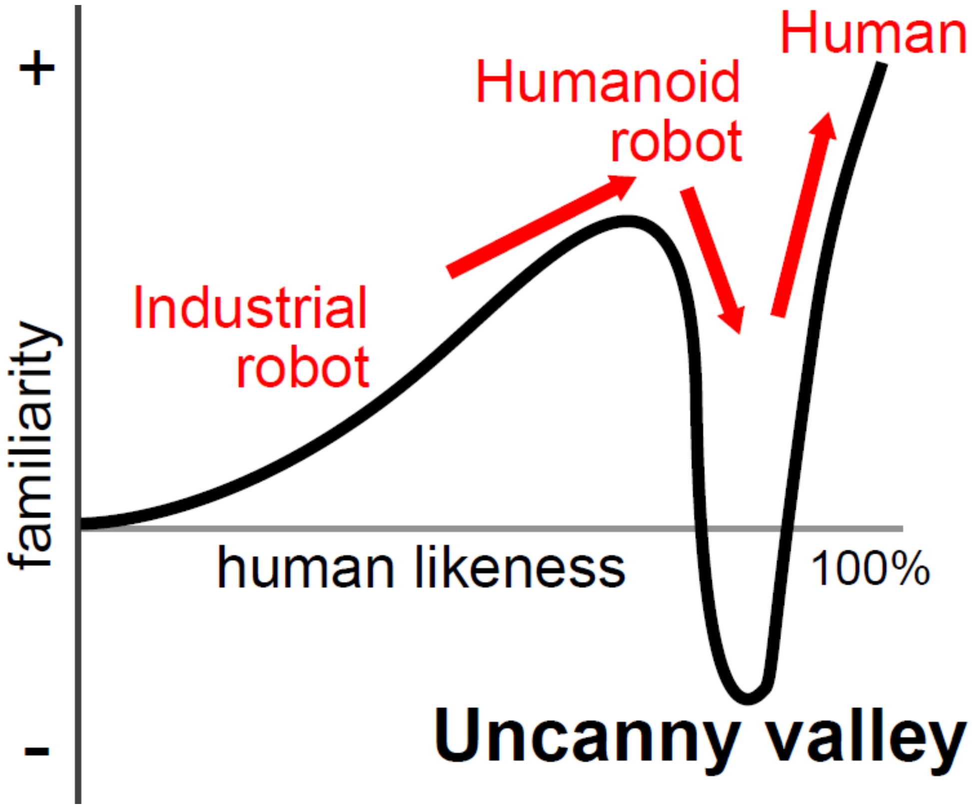
    Imitate facial expressions to generate robot motions

PROLOGUE 2010



- “**Uncanny Valley**”

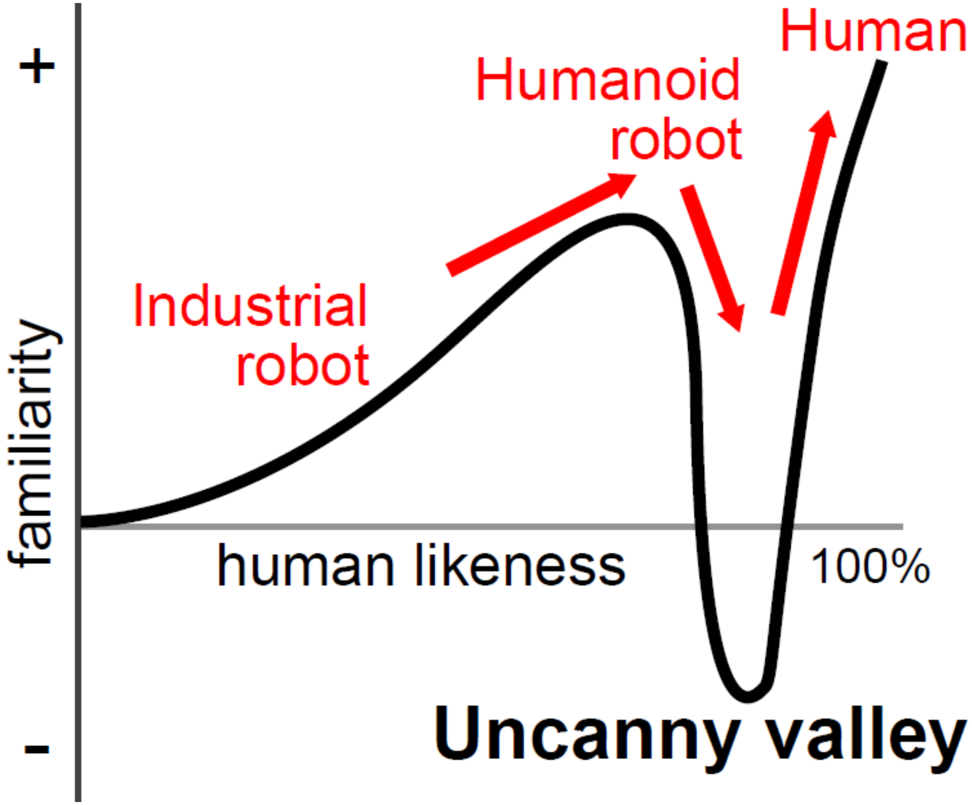
Uncanny? Creepy? Cute?



Toward Future Technologies

- **Let's overcome the "Uncanny Valley"!**
  - We should not be afraid of jumping into the valley
  - Otherwise, we cannot go beyond
  - Let's challenge!

Uncanny? Creepy? Cute?



Robot Singer HRP-4C Miim

VocaListener + VocaWatcher

Video clips are available on the web!

http://staff.aist.go.jp/t.nakano/VocaListener/

http://staff.aist.go.jp/t.nakano/VocaWatcher/

Singing Information Processing Systems

- **Vocal Timbre Analysis**
  - MIR based on vocal timbre similarity
  - Male/female estimation
  - Singer identification
- **Lyric Transcription and Synchronization**
  - Lyric synchronization/transcription
  - Lyric animation (kinetic typography)
- **Singing Skill Evaluation**
  - Singing skill evaluation/visualization/training
- **Singing Synthesis**
  - Text-to-singing synthesis
  - Speech-to-singing synthesis
  - Singing-to-singing synthesis
  - Robot singer

Summary

- **Summary**
  - Introduce singing information processing systems
- **Let's work on singing information processing!**
  - Singing possesses aspects of both speech and music
  - Many unsolved research problems

    Automatic recognition of singing (lyrics) is
    the most difficult class of speech recognition (ASR)
    because of loud accompaniment and large fluctuations

    Singing synthesis requires dynamic, complex, and expressive
    changes in the voice pitch, power, and timbre
  - Research activities on speech and music will be integrated

References

MIR based on singing voices / Query-by-Humming (QBH)

- H. Fujihara and M. Goto, "A music information retrieval system based on singing voice timbre," in Proc. of ISMIR 2007, pp.467-470, 2007.
- T. Kageyama, K. Mochizuki and Y. Takashima, "Melody retrieval with humming," in Proc. of ICMC 93, pp.349-351, 1993.
- A. Ghias, J. Logan, D. Chamberlin and B. Smith, "Query by humming: Musical information retrieval in an audio database," in Proc. of ACM Multimedia 1995, vol.95, pp.231-236, 1995.
- T. Sonoda, M. Goto and Y. Muraoka, "A WWW-based melody retrieval system," in Proceedings of ICMC 98, pp.349-352, 1998.
- R. Dannenberg, W. Birmingham, B. Pardo, C. Meek, N. Hu and G. Tzanetakis, "A comparative evaluation of search techniques for query-by-humming using the musart testbed," Journal of the American Society for Information Science and Technology, vol.58, pp.687-701, 2007.
- E. Unal, E. Chew, P. Georgiou and S. Narayanan, "Challenging uncertainty in query by humming systems: A fingerprinting approach," IEEE Transactions on Audio, Speech, and Language Processing, vol.16, pp.359-371, 2008.
- J.-S. R. Jang and H.-R. Lee, "A general framework of progressive filtering and its application to query by singing/humming," IEEE Transactions on Audio, Speech, and Language Processing, vol.16, no.2, pp.350--358, 2008.
- W.-H. Tsai, Y.-M. Tu and C.-H. Ma, "An FFT-based fast melody comparison method for query-by-singing/humming systems," Pattern Recognition Letters, vol.33, no.16, pp.2285-2291, 2012.
- Y. Ohishi, M. Goto, K. Ito and K. Takeda, "A stochastic representation of the dynamics of sung melody," in Proc. of ISMIR 2007, pp.371-372, 2007.
- M. Suzuki, T. Ichikawa, A. Ito and S. Makino, "Novel tonal feature and statistical user modeling for query-by-humming," Journal of Information Processing, vol.17, pp.95-105, 2009.
- D. Little, D. Raffensperger and B. Pardo, "A query by humming system that learns from experience," in Proc. of ISMIR 2007, pp.335-338, 2007.
- A. Duda, A. Nurnberger and S. Stober, "Towards query by singing / humming on audio databases," in Proc. of ISMIR 2007, pp.331-334, 2007.

References

Vocal Timbre Analysis / Singer Identification

- A. Kanato, T. Nakano, M. Goto and H. Kikuchi, "An automatic singing impression estimation method using factor analysis and multiple regression," in Proc. of ICMC SMC 2014, pp.1244-1251, 2014.
- B. Whitman, G. Flake and S. Lawrence, "Artist detection in music with minnowmatch," in Proc. of NNSP 2001, pp.559-568, 2001.
- A. L. Berenzweig, D. P. W. Ellis and S. Lawrence, "Using voice segments to improve artist classification of music," in Proc. of AES-22 Intl. Conf. on Virt., Synth., and Ent. Audio, 2002.
- Y. E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features," in Proc. of ISMIR 2002, pp.164-169, 2002.
- T. Zhang, "Automatic singer identification," in Proc. of ICME 2003, pp.33-36, 2003.
- M. A. Bartsch, Automatic Singer Identification in Polyphonic Music. PhD thesis, The University of Michigan, 2004.
- N. C. Maddage, C. Xu and Y. Wang, "Singer identification based on vocal and instrumental models," in Proc. of ICPR'04, vol.2, pp.375-378, 2004.
- W.-H. Tsai and H.-M. Wang, "Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.1, pp.330-341, 2006.
- T. L. Nwe and H. Li, "Exploring vibrato-motivated acoustic features for singer identification," IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.2, pp.519-530, 2007.
- J. Shen, B. Cui, J. Shepherd and K.-L. Tan, "Towards efficient automated singer identification in large music databases," in Proc. of SIGIR '06, pp.59-66, 2006.
- A. Mesaros, T. Virtanen and A. Klapuri, "Singer identification in polyphonic music using vocal separation and pattern recognition methods," in Proc. of ISMIR 2007, 2007.
- H. Fujihara, T. Kitahara, M. Goto, K. Komatani, T. Ogata and H. G. Okuno, "Singer identification based on accompaniment sound reduction and reliable frame selection," in Proc. of ISMIR 2005, pp.329-336, 2005.

References

- H. Fujihara, M. Goto, T. Kitahara and H. G. Okuno, "A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre similarity-based music information retrieval," IEEE Transactions on Audio, Speech, and Language Processing, vol.18, no.3, pp.638-648, 2010.
- W.-H. Tsai, H.-M. Wang, D. Rodgers, S.-S. Cheng and H. M. Yu, "Blind clustering of popular music recordings based on singer voice characteristics," in Proc. of ISMIR 2003, pp.167-173, 2003.
- T. Nakano, K. Yoshii and M. Goto, "Vocal timbre analysis using latent Dirichlet allocation and cross-gender vocal timbre similarity," in Proc. of ICASSP 2014, pp.5239-5343, 2014.

References

Lyric Synchronization

- A. Loscos, P. Cano and J. Bonada, "Low-delay singing voice alignment to text," in Proc. of ICMC 99, 1999.
- Y. Wang, M. Kan, T. Nwe, A. Shenoy, and J. Yin, “Lyrically: automatic synchronization of acoustic musical signals and textual lyrics,” in Proceedings of ACM Multimedia 2014, pp.212-219, 2004.
- C. H. Wong, W. M. Szeto and K. H. Wong, "Automatic lyrics alignment for cantonese popular music," Multimedia Systems, vol.4-5, no.12, pp.307-323, 2007.
- M. Muller, F. Kurth, D. Damm, C. Fremerey and M. Clausen, "Lyrics-based audio retrieval and multimodal navigation in music collections," in Proc. of ECD L 2007, pp.112-123, 2007.
- M.-Y. Kan, Y. Wang, D. Iskandar, T. L. Nwe and A. Shenoy, "Lyrically: Automatic synchronization of textual lyrics to acoustic music signals," IEEE Trans. on Audio, Speech, and Language Processing, vol.16, no.2, pp.338-349, 2008.
- K. Chen, S. Gao, Y. Zhu and Q. Sun, "Popular song and lyrics synchronization and its application to music information retrieval," in Proc. of MMCN'06, 2006.
- H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata and H. G. Okuno, "Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals," in Proc. of ISM 2006, pp.257-264, 2006.
- D. Iskandar, Y. Wang, M.-Y. Kan and H. Li, "Syllabic level automatic synchronization of music signals and text lyrics," in Proc. ACM Multimedia 2006, pp.659-662, 2006.
- H. Fujihara and M. Goto, "Three techniques for improving automatic synchronization between music and lyrics: Fricative detection, filler model, and novel feature vectors for vocal activity detection," in Proc. of ICASSP 2008, 2008.
- K. Lee and M. Cremer, “Segmentation-based lyrics-audio alignment using dynamic programming.,” in Proc. ISMIR, 2008, pp.395-400.
- A. Mesaros and T. Virtanen, “Automatic alignment of music audio and lyrics,” in Proc. DAFx, 2008, pp.321-324.
- H. Fujihara, M. Goto, J. Ogata and H. G. Okuno, "LyricSynchronizer: Automatic synchronization system between musical audio signals and lyrics," IEEE J. of Selected Topics in Signal Processing, vol.5, no.6, pp.1252-1261, 2011.
- M. Mauch, H. Fujihara and M. Goto, "Integrating additional chord information into HMM-based lyrics-to-audio alignment," IEEE Transactions on Audio, Speech, and Language Processing, vol.20, no.1, pp.200-210, 2012.

References

Lyric Transcription

- C.-K. Wang, R. -Y. Lyu and Y.-C. Chiang, "An automatic singing transcription system with multilingual singing lyric recognizer and robust melody tracker," in Proc. of Eurospeech 2003, pp.1197-1200, 2003.
- A. Sasou, M. Goto, S. Hayamizu and K. Tanaka, "An autoregressive, non-stationary excited signal parameter estimation method and an evaluation of a singing-voice recognition," in Proc. of ICASSP 2005, pp.1-237-240, 2005.
- M. Suzuki, T. Hosoya, A. Ito and S. Makino, "Music information retrieval from a singing voice using lyrics and melody information," EURASIP Journal on Advances in Signal Processing, vol.2007, 2007.
- A. Mesaros and T. Virtanen, “Automatic recognition of lyrics in singing,” EURASIP Journal on Audio, Speech, and Music Processing, vol.2010, 2010.
- A. Mesaros and T. Virtanen, “Recognition of phonemes and words in singing,” in Proceedings of ICASSP 2010, pp.2146-2149, 2010.
- M. McVicar, D. P. Ellis and M. Goto, "Leveraging repetition for improved automatic lyric transcription in popular music," in Proc. of ICASSP 2014, pp.3141-3145, 2014.
- M. Gruhne, K. Schmidt and C. Dittmar, "Phoneme recognition in popular music," in Proc. of ISMIR 2007, pp.369-370, 2007.
- W.-H. Tsai and H.-M. Wang, "Automatic identification of the sung language in popular music recordings," J New Music Res., vol.36, no.2, pp.105-114, 2007.

Hyperlinking Lyrics

- H. Fujihara, M. Goto and J. Ogata, "Hyperlinking Lyrics: A method for creating hyper links between phrases in song lyrics," in Proc. of ISMIR 2008, pp.281-286, 2008.

References

Singing Skill Evaluation / Singing Training / Other topics

- W, T, Bartholomew, "A physical definition of "good voice quality " in the male voice," J. Acoust. Soc. Am., vol.55, pp.838-844, 1934.
- T. Saitou and M. Goto, "Acoustic and perceptual effects of vocal training in amateur male singing," in Proc. of Interspeech 2009, pp.832-835, 2009.
- D. Ruinskiy and Y. Lavner, "An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals," IEEE Transactions on Audio, Speech, and L anguage Processing, vol.15, pp.838-850, 2007.
- T. Nakano, J. Ogata, M, Goto and Y. Hiraga, "Analysis and automatic detection of breath sounds in unaccompanied singing voice," in Proc. of ICMPC 2008, pp.387-390, 2008.
- p, Lal, "A comparison of singing evaluation algorithms," in Proc. of Interspeech 2006, pp, 2298-2301, 2006,
- T. Nakano, M. Goto and Y. Hi raga, "An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features," in Proc. of Interspeech 200..pp.1706-1709, 2006.
- P. Prasert, K. Iwano and S. Furui, "An automatic singing voice evaluation method for voice training systems," in Proceeding of the 2008 Spring Meeting of the Acoustical Society of Japan, pp.911-912, 2008.
- R. Daido, S. Hahm, Ito, S. Makino and A. Ito, "A system for evaluating singing enthusiasm for karaoke," in Proc. of ISMIR 2011, pp.31-36, 2011.
- D. M, Howard and G, F. Welch, "Microcomputer-based singing ability assessment and development," Applied Acoustics, vol.27, pp.89-102, 1989.
- D. Hoppe, M. Sadakata and P. Desain, "Development of real-time visual feedback assistance in singing training: a review," Journal of Computer Assisted L earning, vol.22, pp.308-316, 2006.
- T. Nakano, M. Goto and y, Hiraga, "MiruSinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data," in Proc. of ISM 2007 Workshops (Demonstrations), pp.75-76, 2007.

References

Singing Synthesis

- P. R. Cook, "Singing voice synthesis: History, current work, and future directions," Computer Music Journal, vol.20, no.3, pp.38-46, 1996.
- P. Depalle, G. Garcia and X. Rodet, "A virtual castrato," in Proc. of ICMC '94, pp.357-360, 1994.
- J. Sundberg, "The KTH synthesis of singing," Advances in Cognitive Psychology. Special issue on Music Performance, vol.2, no.2-3, pp.131-143, 2006.
- P. R. Cook, Identification of Control Parameters in an Articulatory Vocal Tract Model, With Applications to the Synthesis of Singing. PhD thesis, Stanford University, 1991.
- J. Bonada and X. Serra, "Synthesis of the singing voice by performance sampling and spectral models," IEEE Signal Processing Magazine, vol.24, no.2, pp.67-79, 2007.
- D. Schwarz, "Corpus-based concatenative synthesis," IEEE Signal Processing Magazine, vol.24, no.2, pp.92-104, 2007.
- H. Kenmochi and H. Ohshita, "Vocaloid - commercial singing synthesizer based on sample concatenation," in Proc. of Interspeech 2007, pp.4009-4010, 2007.
- K. Saino, H. Zen, Y. Nankaku, A. Lee and K. Tokuda, "An HMM-based singing voice synthesis system," in Proc. of Interspeech 2006, pp.1141-1144, 2006.
- K. Saino, M. Tachibana and H. Kenmochi, "A singing style modeling system for singing voice synthesizers," in Proc. of Interspeech 2010, pp.2894-2897, 2010.
- T. Nose, M. Kanemoto, T. Koriyama and T. Kobayashi, "A style control technique for singing voice synthesis based on multiple-regression HSMM," in Proc. of Interspeech 2013, pp.378-382, 2013.
- M. Umbert, J. Bonada and M. Blaauw, "Generating singing voice expression contours based on unit selection," in Proc. of SMAC 2013, 2013.
- H. Kawahara and H. Katayose, "Scat singing generation using a versatile speech manipulation system STRAIGHT," The Journal of the Acoustical Society of America (The 141 st Meeting of the Acoustical Society of America), vol.109, no.5, pp.2425-2426, 2001.

References

- H. Kawahara, "Application and extensions of STRAIGHT-based morphing for singing voice manipulations based on vowel centred approach," in Proc. of the 19th International Congress on Acoustics 2007 (ICA 2007), pp.2018-2021, 2007.
- M. Morise, M. Onishi, H. Kawahara and H. Katayose, "v.morish'09: A morphing-based singing design interface for vocal melodies," in Proc. of ICEC 2009, pp.185-190, 2009.
- H. Kawahara, M. Morise, H. Banno and V. G. Skuk, "Temporally variable multi-aspect N-way morphing based on interference-free speech representations," in Proc. of APSIPA ASC 2013, 2013.
- J. Janer, J. Bonada and M. Blaauw, "Performance-driven control for sample-based singing voice synthesis," in Proc. of DAFx-06, pp.41-44, 2006.
- T. Nakano and M. Goto, "VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation," in Proc. of SMC 2009, pp.343-348, 2009.
- H. Kenmochi, "VOCALOID and Hatsune Miku phenomenon in japan," in Proc. of the First Interdisciplinary Workshop on Singing Voice (InterSinging 2010), pp.1-4, 2010.
- Crypton Future Media, "What is the HATSUNE MIKU movement?", http://www.crypton.co.jp/download/pdf/info_miku_e.pdf, 2008.
- T. Saitou, M. Goto, M. Unoki and M. Akagi, "Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices," in Proc. of WASPAA 2007, pp.215-218, 2007.
- M. Goto, T. Nakano, S. Kajita, Y. Matsusaka, S. Nakaoka and K. Yokoi, "VocaListener and VocaWatcher: Imitating a human singer by using signal processing," in Proc. of ICASSP 2012, pp.5393-5396, 2012.
- T. Nakano and M. Goto, "VocaListener2: A singing synthesis system able to mimic a user's singing in terms of voice timbre changes as well as pitch and dynamics," in Proc. of ICASSP 20ll, pp.453-456, 2011.
- S. Kajita, T. Nakano, M. Goto, Y. Matsusaka, S. Nakaoka and K. Yokoi, "VocaWatcher: Natural singing motion generator for a humanoid robot," in Proc. of IROS 2011, 2011.

References

VocaRefiner / Singing Voice Conversion

- T. Nakano and M. Goto, "VocaRefiner: An interactive singing recording system with integration of multiple singing recordings," in Proc. of SMC 2013, pp.115-122, 2013.
- F. Villavicencio and J. Bonada, "Applying voice conversion to concatenative singing-voice synthesis," in Proc. of Interspeech 2010, pp.2162-2165, 2010.
- H. Doi, T. Toda, T. Nakano, M. Goto and S. Nakamura, "Singing voice conversion method based on many-to-many eigenvoice conversion and training data generation using a singing-to-singing synthesis system," in Proc. of APSIPA ASC 2012, 2012.
- H. Doi, T. Toda, T. Nakano, M. Goto and S. Nakamura, "Evaluation of a singing voice conversion method based on many-to-many eigenvoice conversion," in Proc. of Interspeech 2013, pp.1067-1071, 2013.
- K. Kobayashi, H. Doi, T. Toda, T. Nakano, M. Goto, G. Neubig, S. Sakti and S. Nakamura, "An investigation of acoustic features for singing voice conversion based on perceptual age," in Proc. of Inter speech 2013, pp.1057-1061, 2013.
- K. Kobayashi, T. Toda, T. Nakano, M. Goto, G. Neubig, S. Sakti and S. Nakamura, "Regression approaches to perceptual age control in singing voice conversion," in Proc. of 1CASSP 2014, pp.7954-7958, 2014.

Psychology / Physiology / Vocal Pedagogy

- D. Deutsch, ed., The Psychology of Music. Academic Press, 1982.
- I. R. Titze, Principles of Voice Production. The National Center for Voice and Speech, 2000.
- F. Husler and Y. Rodd-Marling, Singing: The Physical Nature of the Vocal Organ. A Guide to the Unlocking of the Singing Voice. Hutchinson & Co, 1965.

References

Voice Percussion

- O. Gillet and G. Richard, "Drum loops retrieval from spoken queries," Journal of Intelligent Information Systems, vol.24, no.2-3, pp.159-177, 2005.
- O. Gillet and G. Richard, "Indexing and querying drum loops databases," in Proc. of CBMI 2005, 2005.
- A. Kapur, M. Benning and G. Tzanetakis, "Query-by-beatboxing: Music retrieval for the DJ," in Proc. of ISMIR 2004, pp.170--177, 2004.
- A. Hazan, "Towards automatic transcription of expressive oral percussive performances," in Proc. of IUI 2005, pp.296-298, 2005.
- E. Sinyor, C. McKay, R. Fiebrink, D. McEnnis and I. Fujinaga, "Beatbox classification using ace," in Proc. of ISMIR 2005, pp.672-675, 2005.
- T. Nakano, M. Goto, J. Ogata and Y. Hiraga, "Voice Drummer: A music notation interface of drum sounds using voice percussion input," in Proc. of UIST 2005 (Demos), pp.49-50, 2005.

The references above are shown in the following papers:

- **M. Goto: Singing Information Processing, Proceedings of the 12th IEEE International Conference on Signal Processing (IEEE ICSP 2014), pp.2431-2438, October 2014.**
- **M. Goto, T. Saitou, T. Nakano, and H. Fujihara: Singing Information Processing Based on Singing Voice Modeling, Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010), pp.5506-5509, March 2010.**

ISMIR 2015 Tutorial

Conclusions

Simon Dixon Queen Mary University of London, UK

Masataka Goto AIST, Japan

Matthias Mauch Queen Mary University of London, UK

2015/10/26

Why singing is interesting

- inherent reasons
  - people love to sing
  - people love to listen to other people singing
- scientific reasons
  - scientific discovery in music psychology: how people sing, and how people perceive singing
  - scope for historical and cultural analysis: how people's singing differs and changes
- MIR reasons
  - many MIR tasks relating to singing can be improved, and new ones explored!
  - there's a lot of data out there (even annotated), which we can exploit

Why singing is interesting

- inherent reasons
  - people love to sing
  - people love to listen to other people singing
  - people love to listen to computers singing
- scientific reasons
  - scientific discovery in music psychology: how people sing, and how people perceive singing
  - scope for historical and cultural analysis: how people's singing differs and changes
- MIR reasons
  - many MIR tasks relating to singing can be improved, and new ones explored!
  - there's a lot of data out there (even annotated), which we can exploit