

Speech to chant transformation with the phase vocoder

Axel Röbel¹ Joshua Fineberg²

¹IRCAM, analysis/synthesis group, Paris France

²Music Department, Harvard University, Boston USA

August 23, 2007

Lolita, by Joshua Fineberg

The compositional idea of the **imaginary opera *Lolita***:

- imaginary opera = rendering the **fantasies** of the main person, Humbert Humbert,
- Humbert is **oscillating between reality and fantasy**,
- the voices of the imagined persons are created by means of **transforming the actor's voice** into chant,
- premier in april 2008.

Technical background

Technical problems :

- speech to chant transformation generally requires **extreme time stretching and transposition**,
- output signal should sound as **natural as possible**,
- provide control for (partial) **speaker identity transformation**,
- provide control for **singing voice quality modifications**.

Techniques used

- transformations are done in the IRCAM **phase vocoder** implementation featuring **shape invariant** (or **waveform preserving**) signal transformations.
- **envelope modifications** based on cepstral domain *True Envelope* estimator.

System overview

speech transformation:

- **record speech signal** with desired text,
- **f0 analysis** of speech signal,
- **segment and label** speech signal (semi automatic),
- **transform duration and pitch** of speech signal according to desired melody,

Personality transformation

Lolita

Technical
background

System
overview

Personality
transformation

Demo

Remarks

- record chant signal with **constant pitch**,
- f0 analysis of speech signal,
- **segment and label** chant signal,
- **transform duration and pitch of chant signal** according to desired melody,
- gradually **mix the energy contours** and **spectral envelopes** of the speech and chant signals.

Lolita demo

Original voice example 1: [leave me alone](#)

Original voice example 2: [let them play](#)

Lolita extract: [play](#)

InterSpeech song

Lolita

Technical
background

System
overview

Demo

InterSpeech song

Remarks

- **no personality transformation,**
- **no use of singing voice spectral envelope,**
- **direct speech to chant transformation *without using singing voice time-domain envelopes,***
- **simplistic control *singing voice quality* by means of artificial *vibrato, time varying amplitude gain* and *modulation of spectral tilt.***

Recorded speech: [play speech](#)

InterSpeech song from recorded speech: [play song](#)

Synthetic speech (IRCAM Text-To-Speech):

[play speech](#)

InterSpeech song from synthetic speech: [play song](#)

Remarks

Preliminary work, can be largely improved:

- Use **recording specifications**, e.g. give pronunciation indications to the speaker
- Apply **spectral envelopes of singers**
- Apply **time-domain envelopes of singers**
- Improve **F0 and vibrato contour, time varying amplitude gain and modulation of spectral tilt**
- Use IRCAM TTS **concatenative synthesis** from a corpus of **recorded singing voice units** to obtain the basic sentence and apply the same IRCAM phase vocoder technique to obtain the wanted notes