An update on voice banking and voice reconstruction

Christophe Veaux, Junichi Yamagishi
The Centre for Speech Technology Research
University of Edinburgh
An unmet need for personalised voices

- **Neurological degenerative diseases**
  - Motor Neurone Disease
  - Parkinson’s
  - Multiple Sclerosis
- **Non-progressive neurological conditions**
  - Cerebral Palsy
  - Stroke
- **Speech impairment**
  - Cancer of the vocal cords (laryngectomy) or tongue (grossectomy)

**Degenerative diseases:**

- Some patients may already have speech disorders at the time of diagnosis
Building personalised voices

- **Voice Cloning**
  - HMM speech synthesis enable to adapt acoustic models to a new speaker
  - Significantly smaller amounts of speech data than previously required

- **Voice Reconstruction**
  - fixing statistical models so that they can generate natural sounding speech while keeping speaker identity

Better if the voice donors share the voice characteristics of the patients

**Diagram:**

- Short recording → Voice cloning → Voice catalogue → Voice Reconstruction → Repaired voice

- AVM → Voice catalogue
- Healthy reference → Voice catalogue

*patient* → *voice donors*
Voice for everyone by everyone

Healthy volunteers

Voice clone

Voice donation and banking

Voice catalogue

Voice repair

Short recording

Voice clone

Voice banking

Patients who have vocal problems

Patients who can speak well

Thursday, 23 May 13
Voice banking

Voice banking at School of Informatics and the Anne Rowling Regenerative Neurology Clinic

Semi-anechoic chamber of School of Informatics, University of Edinburgh

Anne Rowling Regenerative Neurology Clinic (Jan 2013): Voice banking studio will be constructed inside the clinic.
UK-wide voice database

- nearly 500 free volunteers (~ 400 hours in total )
- Start recordings of MSPs to get more scottish accents
UK-wide voice database

- nearly 500 free volunteers
  (~ 400 hours in total)

- Start recordings of MSPs to get more scottish accents

- text selection optimising HTS context coverage (across speaker)

- Reading style speech:
  newspaper, accent judgement sentences, rainbow passage

- “Conversational” style dialogues (switchboard, usenet)
  selected using a language model trained on AAC sentences
Catalog of Average Voice Models

- Approximately 4000 sentences (10 speakers) required to build an average voice.

- Voice donors are pooled into clusters to create **average voice models (AVM)** with **specific accent / age / gender**.

- **Automatic selection of the best AVM** for voice cloning.
Voice Reconstruction

**Post-process:** Fixing statistical models for each stream of the patient’s voice clone

Uses donors voice models:

- Substitution / Interpolation
- Projection on eigenspace
Voice Reconstruction

**Post-process:** Fixing statistical models for each stream of the patient’s voice clone

Uses donors voice models:

- Substitution / Interpolation
- Projection on eigenspace

- How to find the parameters which are to be substituted?
- How to choose the best voice donor?
First approaches

- **Baseline** model substitution
- Duration and aperiodicity models
- Global variances of log-F0, mel-cepstrum, aperiodicity
- Dynamic coefficients of mel-cepstrum and log-F0
- Voiced/unvoiced weights
- High-order coefficients of mel-cepstrum
- 1st coefficient of mel-cepstrum (energy)

- **Context-dependent** model substitution
- Substitute mel-cepstrum models that corresponds to **specific contexts** such as consonant, approximant, etc.
Distance between models

- KL distance between multi-dimensional distributions (Diagonal Covariance)

\[
D_{KL}(N_x, N_y) = \frac{1}{2} \sum_{i=1}^{N} \left[ \left( \frac{\sigma_{x,i}}{\sigma_{y,i}} \right)^2 + \left( \frac{\mu_{y,i} - \mu_{x,i}}{\sigma_{y,i}} \right)^2 - 2 \log \left( \frac{\sigma_{x,i}}{\sigma_{y,i}} \right) - 1 \right]
\]

Distance between models used to define a confidence measure
Confidence measure and donor selection
Confidence measure and donor selection
Confidence measure and donor selection

Model distance between patient and healthy donors for each parameter
Confidence measure and donor selection

Model distance between patient and healthy donors for each parameter
Confidence measure and donor selection

Model distance between patient and healthy donors for each parameter
Confidence measure and donor selection

Model distance between patient and healthy donors for each parameter

Average model distances between healthy speakers
Interpolation between models

• Interpolation of the model parameters based on the confidence measure

• Different configurations (component-wise, block-wise)
Subjective Evaluation (1/2)

- Two recordings of a same MND patient
  - one “healthy voice” recording (just after diagnosis)
  - one “disordered voice” recording (10 months later)

- Compared synthetic voices:
  - **HC**: Voice clone of “healthy speech”
  - **IC**: Voice clone of “impaired speech”
  - **AV**: Average voice model
  - **IR1**: Automatic reconstruction **Baseline**
  - **IR2**: Automatic reconstruction **Interpolation based on confidence measure**
  - **IR3**: Semi-supervised reconstruction
Subjective Evaluation (2/2)

Intelligibility Test

- Re-transcription of SUS sentences synthesised with the different voices
- Average word error rate (WER) over 40 native English listeners

![Graph showing WER (%) and Similarity (MOS-scale)]
Perspectives: subspace projection

For each context, the model space described by the healthy speaker adapted models is decomposed in eigenvectors. The reconstructed voice is estimated as the projection of the patient’s model on a **subspace** (principal components) that space.
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

- An easy case
  - original
  - voice clone
  - repaired
  - healthy voice

- An ‘extreme’ case
  - original
  - repaired
  - voice donor
  - healthy voice
voice reconstruction: example

- An easy case
  - original
  - voice clone
  - repaired
  - healthy voice

- An ‘extreme’ case
  - original
  - repaired
  - voice donor
  - healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

- An easy case
  - original
  - voice clone
  - repaired
  - healthy voice

- An ‘extreme’ case
  - original
    - repaired
  - voice donor
  - healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
voice reconstruction: example

• An easy case
  • original
  • voice clone
  • repaired
  • healthy voice

• An ‘extreme’ case
  • original
  • repaired
  • voice donor
  • healthy voice
Clinical trials

Feedbacks from patients and families

• Naturalness: too quick, voice slightly robotic
• Vocal identity: ranges between 3/5 and 4/5
• Intelligibility: ranges between 3/5 and 5/5
• All patients declared they were happy overall with their voice

General comments

"I'm pleased to have something that sounds like me" (patient)

"The sound of her voice coming from the living room has been a great boost to me" (partner)

"I am impressed with the quality of her synthesised voice" (relative)
Progress

• **Pre-NST**: funded by Euan MacDonald Centre for Motor Neurone Disease Research
  - initial voice bank of 150 donors
  - pilot trials with one patient

• **Work within NST**
  - greatly expanded voice bank of 500 donors
  - pilot trials with 25 patients
  - **Automatic voice repair**

• On-going developments:
  - clinic-based automatic voice reconstruction system, usable by clinicians
  - voices running on mobile devices

Thursday, 23 May 13
Thanks

http://www.smart-mnd.org/voicebank/about/home.html