Advanced Signal Processing Seminar SS 2010

Voice Transformation Algorithms

Florian Krebs

Signal Processing and Speech Communication Laboratory

29.06.2010

Overview

- Definition and applications
- Components of typical Voice Conversion systems
- Example

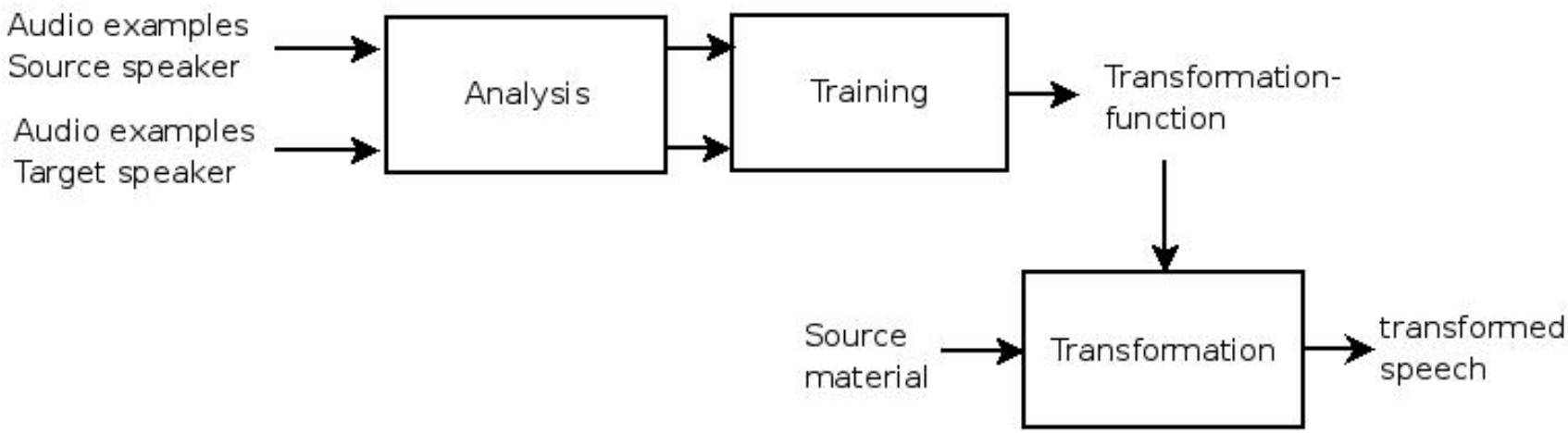
Voice Transformation / Conversion

"Voice conversion aims at transforming the characteristics of the speech signal uttered by a speaker (Source Speaker), in such a way that a human listener could believe that the transformed speech is produced by another specific speaker (Target Speaker)."

Applications

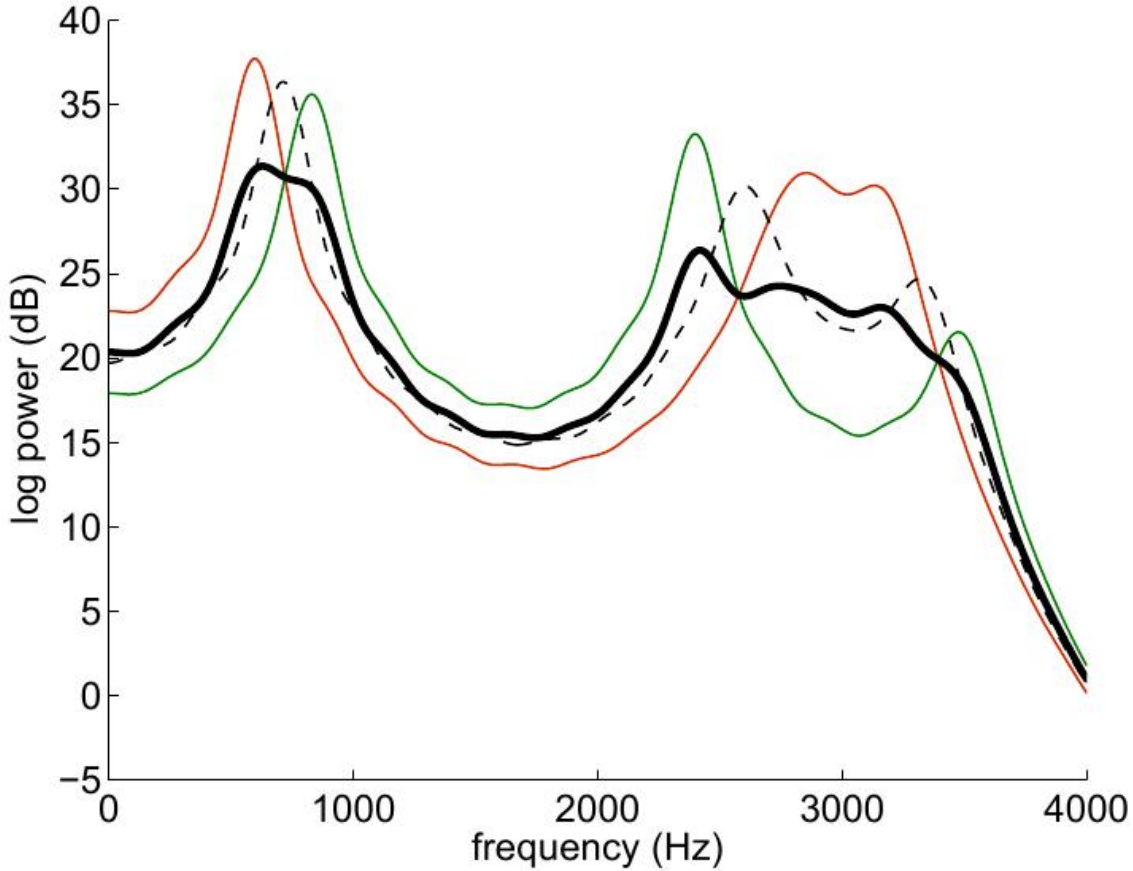
- Different voices for TTS systems
- Voice dubbing for movies or music productions
- Create new voices

Components


Audio examples
Source speaker
Audio examples
Target speaker
Analysis
Training
Transformation-
function
Source
material
Transformation
transformed
speech

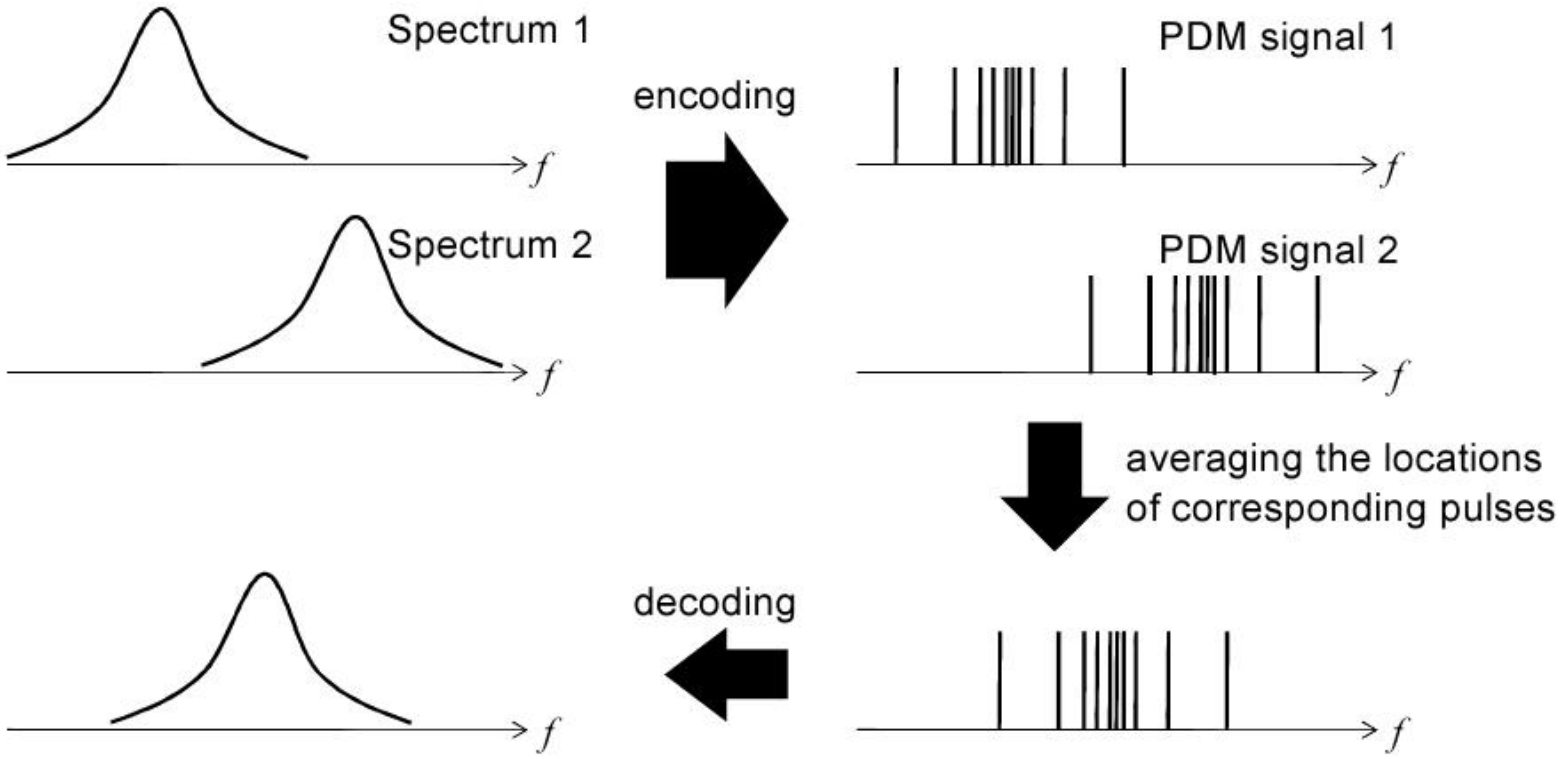
Analysis (MFCC)

Spectral averaging in the cepstral domain



Analysis (LSF)

Spectral interpolation based on line spectrum representation



Training

- Mapping of source and target elements
  - dynamic time warping
  - forced alignment speech recognition
  - Hidden Markov modelling
- Estimate the transfer function
  - mapping codebooks
  - Gaussian mixture models (GMM)
  - neural networks

Transformation

- maximum a posteriori (MAP) estimation
- maximum likelihood linear regression (MLLR)

MAP 1

Bayes theorem:

$$P(\lambda|O) = \frac{P(O|\lambda) \times P(\lambda)}{P(O)} \quad (1)$$

$P(\lambda|O)$ ... posterior probability
$P(O|\lambda)$ ... likelihood
$P(\lambda)$ ... prior
$P(O)$ ... evidence

MAP 2

ML estimation:

$$\hat{\lambda} = arg\max_{\lambda}\{P(O|\lambda)\} \tag{2}$$

MAP estimation:

$$\hat{\lambda} = arg\max_{\lambda}\{P(O|\lambda) \times P(\lambda)\} \tag{3}$$

$P(\lambda|O)$ ... posterior probability
$P(O|\lambda)$ ... likelihood

MLLR 1

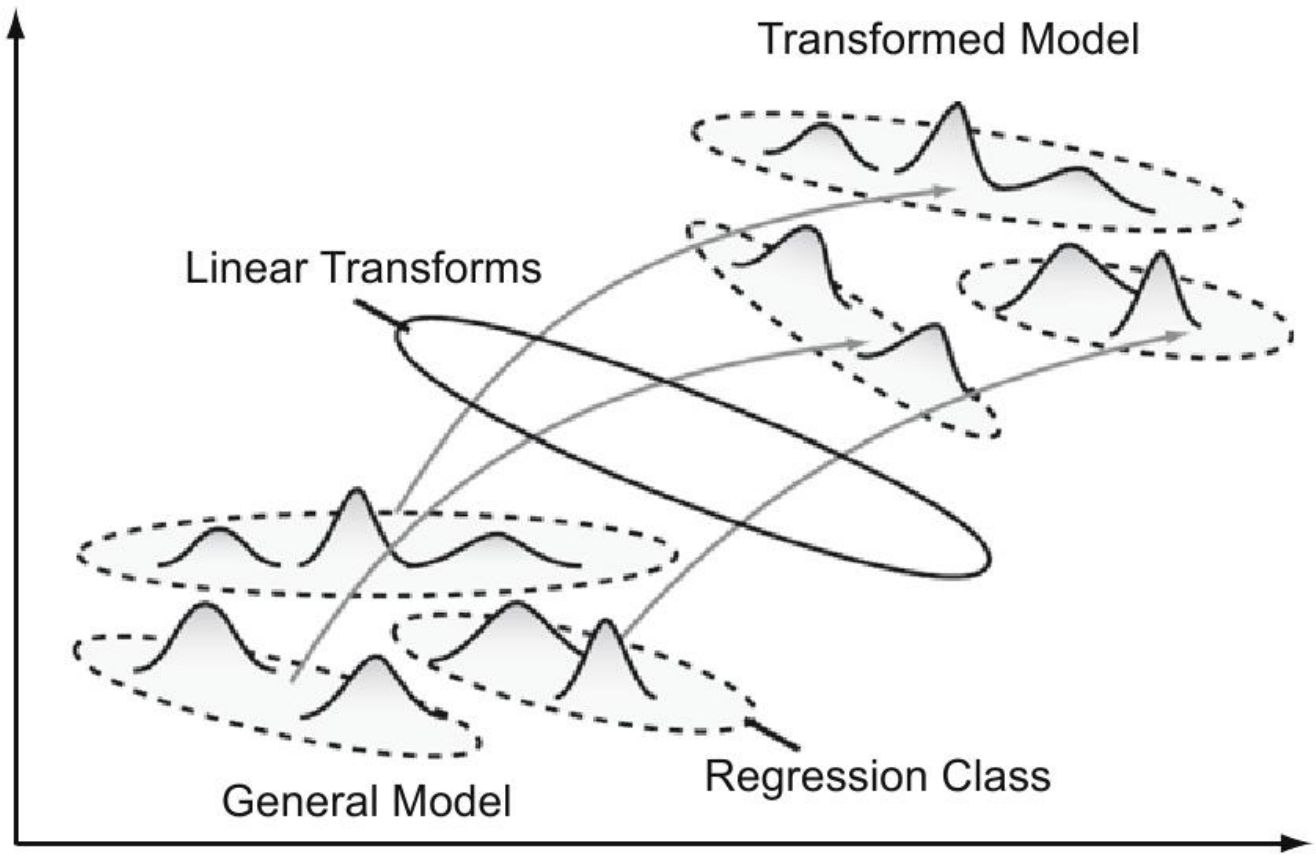


Figure: Overview of linear-transformation-based adaptation technique

MLLR 2

$$\hat{\mu}_j = A_{r(j)}\mu_j + b_{r(j)} \tag{4}$$

$$\hat{\Sigma}_j = H_{r(j)}\Sigma_j H_{r(j)}^{\mathrm{T}} \tag{5}$$

μ_j ... mean vector of the jth state-output source distribution
$\hat{\mu}_j$ ... mean vector of the transformed source distribution
Σ_j ... covariance matrices
$A_{r(j)}$ ... mean linear-transformation matrix
$H_{r(j)}$ ... covariance linear-transformation matrix
$b_{r(j)}$ ... mean bias vector for the $r(j)$th regression class.

Example: Kain and Macon 1

Analysis

- Speech corpus
  - 10 speakers speak the same 50 sentences
- Features
  - 16th order line spectral frequencies (LSFs)
  - LPC residual
  - pitch-synchronous sinusoidal analysis over 2 pitch periods
  - warped using Bark scale

Example: Kain and Macon 2

Training

- forced alignment using the CSLU speech toolkit
- locally linear transformation function with two unknown parameters v_i and Γ_i:

$$F(x) = \sum_{i=1}^{Q} h_i(x)[v_i + \Gamma_i \Sigma_i^{-1}(x - \mu_i)] \tag{6}$$

Q ... number of multivariate Gaussian functions
$h_i(x)$ ... posterior probability that the i^{th} Gaussian component generated the spectral target vector x

- GMMs are trained to produce the unknown parameters v_i and Γ_i to minimize the mean squared error between target und source spectrum.

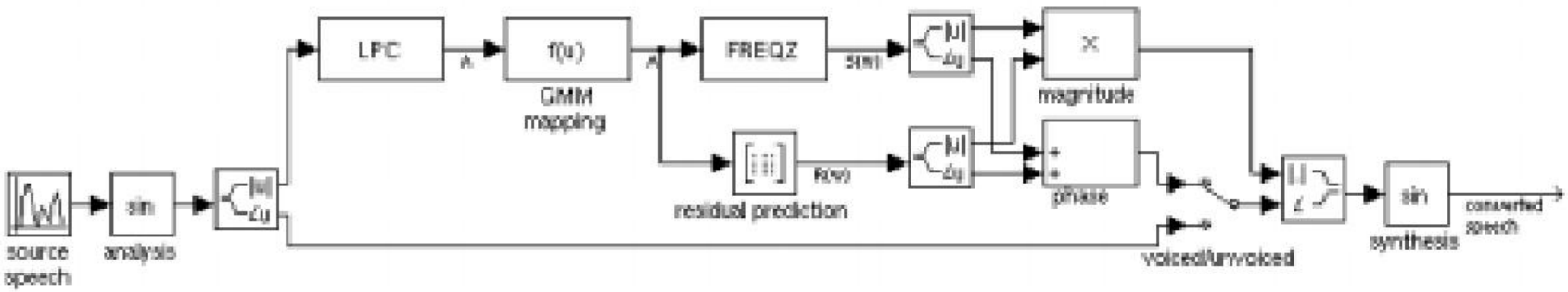
Example: Kain and Macon 3

Transformation

- spectral vectors from the source speaker are converted using the transformation function
- 1998:
  - the pitch of the source speaker's residual is adjusted to match the target speaker's pitch in average value and variance
  - modified residual and spectrum are convolved to render the converted speech
- 2001:
  - predict the target residual from LPC parameters during voiced speech
  - synthesize using a sinusoidal overlap/add system

Example: Kain and Macon 4

Overview




Oytun Tuerk.
New methods for Voice Conversion.
Master Thesis at Bogazici University, Istanbul, 2003.


Avinash Kak.
ML, MAP and Bayesian - The holy trinity of parameter estimation and data prediction.
Purdue University, 2010.


Heiga Zen, Keiichi Tokuda, Alan W. Black.
Statistical parametric speech synthesis.
Speech Communication 51, 2009.


Kain, A. and M. W. Macon.
Design and Evaluation of a Voice Conversion Algorithm Based On Spectral Envelope Mapping And Residual Prediction.
Proceedings of the IEEE ICASSP, 2001.