

# Speech Synthesis for Minority Languages: A Case Study on Scottish Gaelic

Jeff Berry  
University of Arizona  
jjberry@email.arizona.edu

November 1, 2008

# Why Speech Synthesis?

## Speech synthesis has many applications

- ▶ Accessibility
  1. Visually impaired
  2. Reading disabilities
  3. Speech Impairment
- ▶ Mobile applications (i.e. reading the news)
- ▶ Military training
- ▶ Entertainment
- ▶ More...

# Why Speech Synthesis for Revitalization?

- ▶ Accessibility for native speakers
  1. Native speakers of endangered languages are often in need of assistive technology
- ▶ Increase the prestige of the language
  1. Popular apps to compete with majority languages
  2. Practical apps encourage younger generation to use the language
- ▶ Useful for language teaching
- ▶ Documentation
- ▶ More...

# Approaches to Speech Synthesis

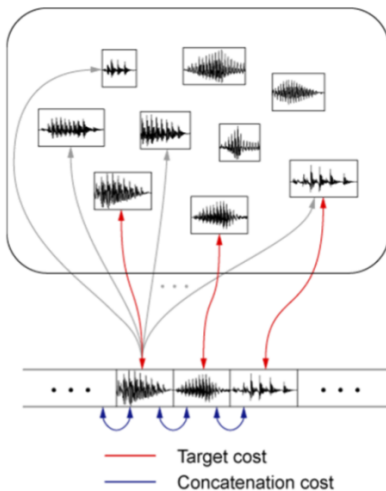
## Parametric synthesis

- ▶ Formant synthesis
- ▶ Articulatory synthesis
- ▶ HMM-based synthesis

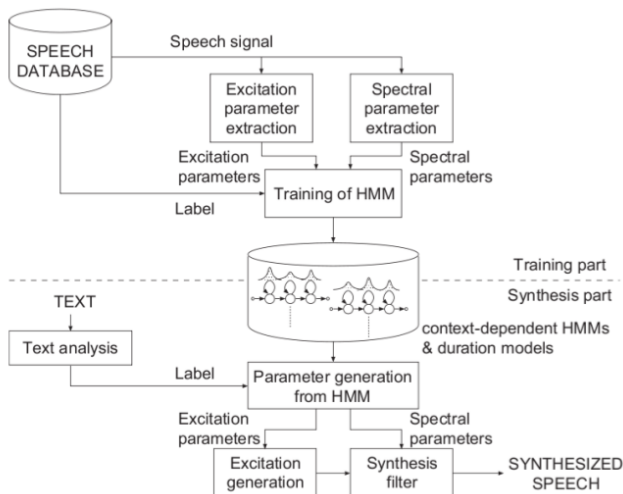
## Concatenative synthesis

- ▶ Diphone synthesis
- ▶ Unit selection synthesis

# Concatenative synthesis



# HMM-based synthesis



# Issues for voice construction in a minority language

## Diphone-based

- ▶ Requires a list of possible diphones
- ▶ Voice talent has to be comfortable reading nonsense prompts to get high-quality diphones
- ▶ Needs a preexisting voice to build prompts (usually English)
- ▶ Results in heavily accented, robotic sounding voice

## HMM-based

- ▶ Needs a speech database to train on
- ▶ Voice talent can read real sentences
- ▶ Results in a more natural sounding voice

# Procedure for diphone voices – Using Festival and Festvox

## Preparation – Generate a native sounding voice

- ▶ Create the diphone list
- ▶ Record a word list that covers the diphone list
- ▶ Label and index diphones
- ▶ Generate nonsense prompts using the extracted diphones

## Create a higher quality voice

- ▶ Record the nonsense prompts
- ▶ Label and index diphones
- ▶ Generate dictionary and letter-to-sound rules
- ▶ Fine-tune and package for distribution

# Procedure for HMM-based voices – HTK/HTS/Festival

## Creating a speech database

- ▶ A set of sentences must be found that gives balanced phonetic coverage of the language (the CMU Arctic database for English has 1132 sentences)
- ▶ Wikipedia makes a good corpus for creating the list – available for 264 languages
- ▶ Use a computer script to select the 1000 most phonetically diverse sentences from Wikipedia

# Procedure for HMM-based voices – HTK/HTS/Festival

## Training the HMMs

- ▶ Recorded sentences have to be labeled, either entirely by hand, or hand corrected after automatic labeling
- ▶ Model type has to be specified, i.e. for triphone, pentaphone; what type of signal processing, etc.
- ▶ Training is done using the tools from Hidden-Markov Tool Kit (HTK/HTS)

# Procedure for HMM-based voices – HTK/HTS/Festival

## Putting everything together

- ▶ Dictionary and letter-to-speech rules have to be implemented for front-end text processing
- ▶ Trained HMMs are used to synthesize the output waveform

# Implications for revitalization

- ▶ A usable voice can be constructed from scratch with a small amount of programming ability
- ▶ Accessibility – helps native speakers interact with technology
- ▶ Practical applications – helps make the language ‘cool’ to use for younger generation
- ▶ Pedagogy

# Thank you

## Special thanks to:

- ▶ Dainon Woudstra
- ▶ Muriel Fisher
- ▶ Andrew Carnie
- ▶ Sunjing Ji
- ▶ Ying Lin

# References



A. Black and K. Lenzo.

**Building Synthetic Voices.**

*Language Technologies Institute, Carnegie Mellon University and Cepstral LLC. Retrieved from <http://festvox.org/bsv>, 2003.*



A. Black and K. Lenzo.

**Multilingual text-to-speech synthesis.**

*Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, 3, 2004.*



D. Crystal.

**Language Death.**

Cambridge University Press, 2000.



M. Krauss.

**The world's languages in crisis.**

*Language, 68(1):1–42, 1992.*



P. Ladefoged, J. Ladefoged, A. Turk, K. Hind, and S.J. Skilton.

**Phonetic structures of Scottish Gaelic.**

*Journal of the International Phonetic Association, 28:1–41, 1998.*



D. MacAuley.

**The Scottish Gaelic language.**

In D. MacAuley, editor, *The Celtic Languages*, pages 137–248. Cambridge University Press, 1992.



M. Wolters.

**A diphone-based text-to-speech system for Scottish Gaelic.**

*Master's thesis, University of Bonn, 1997.*