Speech Synthesis

V 1.1, March 23, 2018
Martin Hagmüller, hagmueller@tugraz.at
Signal Processing and Speech Communication Laboratory
http://www.spsc.tugraz.at/
Graz University of Technology
Inffeldgasse 16c, 8010 Graz, Austria

Abstract

This is the second part of the Speech Communication Laboratory. Its goal is to provide basic insight into the principles of speech synthesis. It mainly covers prosodic manipulation using source-filter methods and PSOLA. The tasks are performed in Matlab, Praat, and Audacity.

Equipment:

- PC with a sound card installed
- MatLab, Praat, Audacity
- MatLab scripts and source files (these you can download from our website) https://www.spsc.tugraz.at/sites/default/files/file/scl/lpc_synth_2008.zip
- Headsets and necessary cables.

Before you start...

1. You are expected to write a detailed report about your work. The report should be prepared on the fly, as you proceed with the tasks and handed over at the end of the laboratory.
2. Do not forget to answer the questions asked in experiments and those asked by the lab assistant. Your answers should be provided in full, with all the necessary explanations and, if required, plots and tables. Try to explain what you observe! If you can't, do not hesitate to ask your lab assistant!

Experiment 1 – Commercial systems evaluation

There are many commercial text-to-speech (TTS) systems available and they are already used in everyday life. Navigation software and mobile phones, among others are applications where TTS has become standard.

(1.1) Listen to commercial text-to-speech demos
Listen to the following TTS demos:

1. CereProc – http://www.cereproc.com/de/ (Check out the voice 'Leopold')
2. acapela – http://www.acapela-group.com/text-to-speech-interactive-demo.html
3. Google – https://translate.google.at/
4. Amazon – https://aws.amazon.com/de/polly/

A more comprehensive overview of German TTS systems can be found at:
http://ttssamples.syntheticspeech.de/

(1.2) Evaluation Write down problems and weaknesses of the different systems. Use the following sentences:

"An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht. Dabei war eigentlich immer sehr schönes Wetter gewesen."

"Dr. A. Smithe von der NATO (und nicht vom CIA) versorgt z.B. - meines Wissens nach - die Heroin seit dem 15.3.00 tgl. mit 13,84 Gramm Heroin zu 1,04 DM das Gramm."

"Die Manpowerdiskussion wird gecancelt, du kannst das File vom Server downloaden."

Consider the following criteria:

- Inteligibility
- Correct pronunciation of foreign language words, acronyms, ...
- Naturalness (speech sounds, pitch contour, rhythm)
- Other issues?

Experiment 2 – Prosodic Manipulation - Source filter model

One of the most important parameters to make synthetic speech sound natural is natural prosody. Prosody describes all features, that are not limited to a phone, but involves longer periods, such as a phrase. The most obvious parameter is the fundamental frequency contour, which evolves over a sentence. Another parameter is the rhythm of a speech utterance, i.e. the duration of phones and phrases, that add e.g. accents in a sentence.

The source-filter model is used to separate the excitation and the vocal tract filter. Pitch and time-scale modification (TSM) are then applied on the excitation signal, only.

(2.1) Source-Filter Model - Unvoiced excitation

1. Start MatLab
2. Run `lpc_noise.m` and `lpc_noise_dur.m` How is the TSM implemented?
3. Use different values for `TSM` in the matlab file, e.g. 0.3, 2, 8.
4. Describe why at certain `TSM` values problems occur.

(2.2) Source-Filter Model - Voiced excitation

1. Run `lpc_pulse.m` and `lpc_pulse_dur.m` How is the TSM implemented?
2. Use different values for `TSM` in the matlab file, e.g. 0.3, 2, 8.
3. Describe why this approach works better for long time stretch values.
4. Point out where the weaknesses of this method are specially audible. Find a reason for this.

Experiment 3 – Prosodic Manipulation - PSOLA

Pitch synchronous overlap-add (PSOLA), is a method that works with single pitch cycles. A window is centered around a pitch cycle maximum and the signal parts are then rearranged according the a new pitch contour or duration.

(3.1) PSOLA Pitch modification with Praat

1. Start `Praat` . Record your own voice or load a pre-recorded utterance

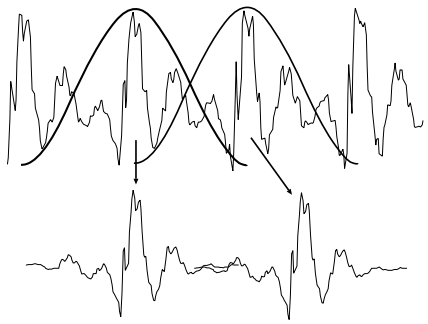

Figure 1: PSOLA Pitch Modification

2. Resynthesize the signal: **To Manipulation...** → **Edit**
3. Simplify the pitch contour: **Pitch** → **Stylize pitch...** with 2 Hz resolution
4. Change the pitch contour, so that a statement is turned into a question. Add pitch points if necessary (**Pitch** → **Add pitch point at cursor**)

Experiment 4 – Concatenative Speech Synthesis

Most state-of-the-art speech synthesis systems concatenate elements of pre-recorded speech. One element size is a diphone, i.e. the transition from one phone to the next.

(4.1) Diphone Concatenation

1. Think of a word to synthesize.
2. Start `Audacity` .
3. Record a different! utterance, that includes all diphones of the word you want to synthesize.
4. Once you have the word put together, you may adjust the pitch contour in Praat.