

Free Speech Synthesis

Milan Zamazal

Brailcom, o.p.s.

Speech synthesis

- Many possible uses.
- Our target group: the blind and visually impaired users.
- Efficient delivery of computer output to the users.

Speech synthesizer is a very important component of an accessible computing environment.

Software speech synthesis

- The simplest way to get speech output today.
- Hardware synthesis inconvenient: External device, drivers, restricted use, limited customization, maintenance not guaranteed.
- Speech quality is no principal reason for using hardware synthesizers.
- We can support hardware synthesizers if possible, but it is more important to focus on improving free software synthesizers.

The job of a speech synthesizer

Transformation of a written text to sound:

- Analysis of the input text.
- Transformations and enhancements to provide the information in a more compact form (punctuation reading, sound icons, ...).
- Determining output parameters (prosody).
- Synthesis of the sound sample.

The jobs of other components

Speech synthesizer shouldn't care for:

- Audio device output (should be managed by an audio output tool).
- Managing messages and interaction with applications (this is what tools like Speech Dispatcher or gnome-speech are for).

General requirements

- Decent speech quality.
- Multiple language support.
- Extensibility.
- Free Software.

Extensibility

Crucial feature allowing to implement missing particular features.

- Nothing is feature complete.
- New features may be needed for advanced use.
- It's much easier to install extensions without touching the synthesizer code than to apply patches to the source code.

Particular requirements

- Changing speech parameters (rate, pitch, etc.).
- Spelling (short, long).
- Punctuation reading.
- Capital letter signalization.
- Sound insertion.
- Interface to speech output software (Speech Dispatcher, gnome-speech).
- Etc.

Possible to implement in a trully extensible speech synthesizer.

Why free?

- Proprietary software synthesizer may share some disadvantages of hardware synthesizers: restricted use, limited customization, maintenance not guaranteed.
- Easy redistribution, no problem with making accessibility distributions.
- Investments don't get lost, any working power can be utilized.
- Anyone can afford it.
- No dependence on a particular entity, avoiding ethical conflicts.

Stealth danger of dependence on proprietary software components.

What's available

- Festival – reasonable quality, extensible, free, but not much active development
- Flite, based on Festival – quality similar to Festival, but non-extensible, very difficult to support new languages and voices, not much active development
- FreeTTS, based on Flite – quality similar to Festival, not sure about extensibility, the code is free but can't be run on a free software system
- Epos – reasonable quality (but no good free voices available), only limited extensibility

Languages

- Most synthesizers support English.
- Most synthesizers don't support much more than English.
- Problems: language dependent know-how needed, making good new voices is not easy.

The synthesizer

- No synthesizer fully satisfying the requirements is available.
- **We need one!**
- The best match: Festival.
- We use Festival in the Free(b)soft project as the only available choice now.

Festival pros

- Complete TTS system.
- Customizable and extensible.
- Quite good extension language (SIOD).
- Completely free.
- New languages can be added (festival-czech).

Festival cons

- Lagging maintenance.
- Latest speech synthesis research not followed anymore?
- Missing means for practical use.
- No clear TTS framework to uniformly hook up into particular phases of the speech synthesis process.
- SIOD is non-standard.

festival-freebsoft-utils

It is possible to do a lot with Festival:

- Generalized concept of input events, including sound icons and word replacements.
- Punctuation reading modes, spelling, capital letter signalization.
- Index marking.
- SSML support.
- Speech Dispatcher interface.

This makes Festival an advanced speech output tool.

Components of speech enabled system

- application – toolkits – Speech Dispatcher – synthesizer – audio output
- application – toolkits – AT-SPI – gnome-speech – synthesizer – audio output

Synthesizer is the only language dependent part.

Couldn't be some parts (events, signalizations) processed independent of the particular synthesizer?

Questions – the synthesizer

- How to make at least one free, fully functional, feature complete and multilingual speech synthesizer?
- What are the available resources?
- Can Festival be extended to what we need?
- Or is it better to develop a completely new speech synthesizer?
- Should we join our efforts or try to develop more alternatives?

Questions – components

- Should the speech synthesizer job be split into several independent (but language dependent!) components?
- What to use in the foreseeable future in Speech Dispatcher and gnome-speech?
- How about audio output server or library?

Questions – users

- What functionality, not yet provided by Festival + festival-freebsoft-utils, is needed?
- How to support more languages?
- Do the blind and visually impaired users care?

Happy hacking!