

# Introduction of Rocaloid3

Rocaloid Development Team

Built on March 7, 2014

## 1 About Rocaloid

Rocaloid is a free vocal synthesis **system**.

Please notice the bold word “**system**”, which differs it from a “software” or “application”: in a broad sense, Rocaloid provides a series of tools and libraries from the very basic DSP libs to synthesis engine and musical score editor. In a narrow sense, what Rocaloid represents in this reference is the back-end of this system.

### 1.1 Purpose

Rocaloid aims at singing voice synthesis rather than speech synthesis(though it is possible to be used as a speech synthesizer).

The description for history of Rocaloid Project can be found on our website:

<http://www.rocaloid.org>

### 1.2 Why Reinventing the Wheel?

In the back-end of Rocaloid, we code everything from scratch. The author Sleepwalking himself is a pure enthusiast in voice signal processing and he does not really care about development cycle. If you are looking for a reason for reinventing the wheel, we would suggest:

- Our wheels are smaller and faster than other wheels(on certain conditions and platforms which are commonly used by our users). e.g. CVEDSP performs 1024-point FFT 1.3 times faster than fftw3 on x86 with SSE.

- Our wheels are interrelated with great consistency. e.g. variable naming convention.
- Our wheels focus on audio signal processing. Surely they are more relevant to our topic.
- We intend to minimize dependencies of Rocaloid(libc only).
- By reverting the wheels we are sure that Rocaloid will be forever free(because we own all of the libs from bottom to top).
- And finally, the developers themselves learn from the project as they reinvent the wheels.

## 2 Current Status

The development of Rocaloid3 started in September, 2013. The core synthesis engine, CVE3, was finished in December. However, a crucial defect was detected in the algorithm and we had to redesign the engine.

Our decision was to abandon CVE3 and move on to CVE3.5, by the way restructuring the whole project.

We have adopted SMS technique for the new synthesis engine. The detailed algorithm will be discussed in Chapter 2.

Currently we are **just about to start** the restructuring.

## 3 About the System

### 3.1 General Diagram

From the users' point of view, Rocaloid reads in a musical score file with melody and lyrics, accesses to a soundfont database, and outputs a .wav sound file.

The soundfont is sampled from a real person(the sound provider). The raw samples undergo preprocesses(e.g. segmentation) and are then converted to sound font format.

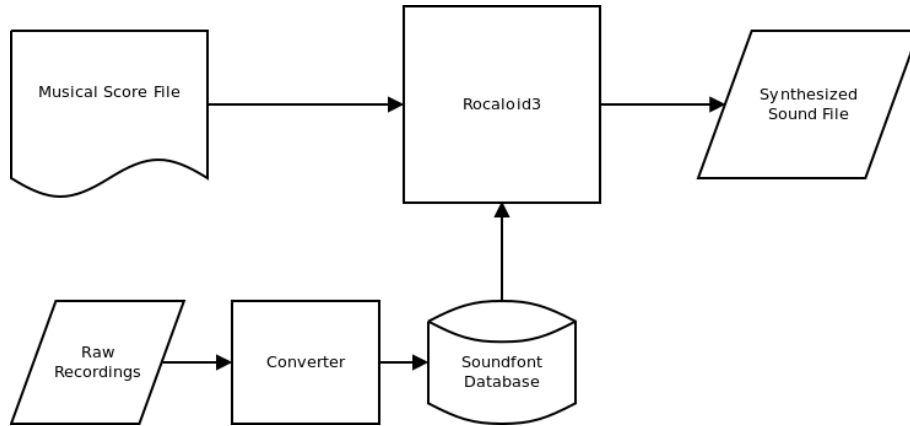


Figure 1: General Concept of Rocaloid3

### 3.2 Inside Rocaloid

After Rocaloid starts up, the musical score file is first converted to a vocal description file. This file describes voice in a phonetic level (lower than musical level). The second step is to convert the voice description file to a soundfont-specific engine script. Many details of voice are generated in this step, such as the exact duration for each phoneme.

Then the CVE synthesis engine is called. CVE directly reads the soundfont-specific script and the soundfont database. The synthesis process may take a while, and finally a .wav file is produced.

### 3.3 Modules

Rocaloid3 is composed of the following modules:

|               |  |
|---------------|--|
| RUtil2        | A tiny library for Objected Oriented Programming and dynamic data structure in C.                    |
| RFNL          | A fast float-point numeric library. Contains basic techniques such as FFT and interpolation methods. |
| CVEDSP2       | A digital signal processing framework that wraps some DSP techniques in interconnectable modules.    |
| CVESMS        | A voice signal processing toolbox based on CVEDSP2.  |
| RFILE3        | A library for file support and input/output.   |
| CVE3.5        | The core synthesis engine.   |
| RParagen      | The soundfont-specific script generator.   |
| CVDBToolChain | A set of tools for building the soundfont database.  |

Beyond these modules, Wavetave is an audio processing sandbox written in Octave Language and C++. All algorithms used by Rocaloid are first tested in Wavetave.

One undocumented module is CVEANN, a small Artificial Neuron Network library initially designed for formant detection and voice parameter generation. However the performance of CVEANN was unsatisfactory and it has been obsoleted.

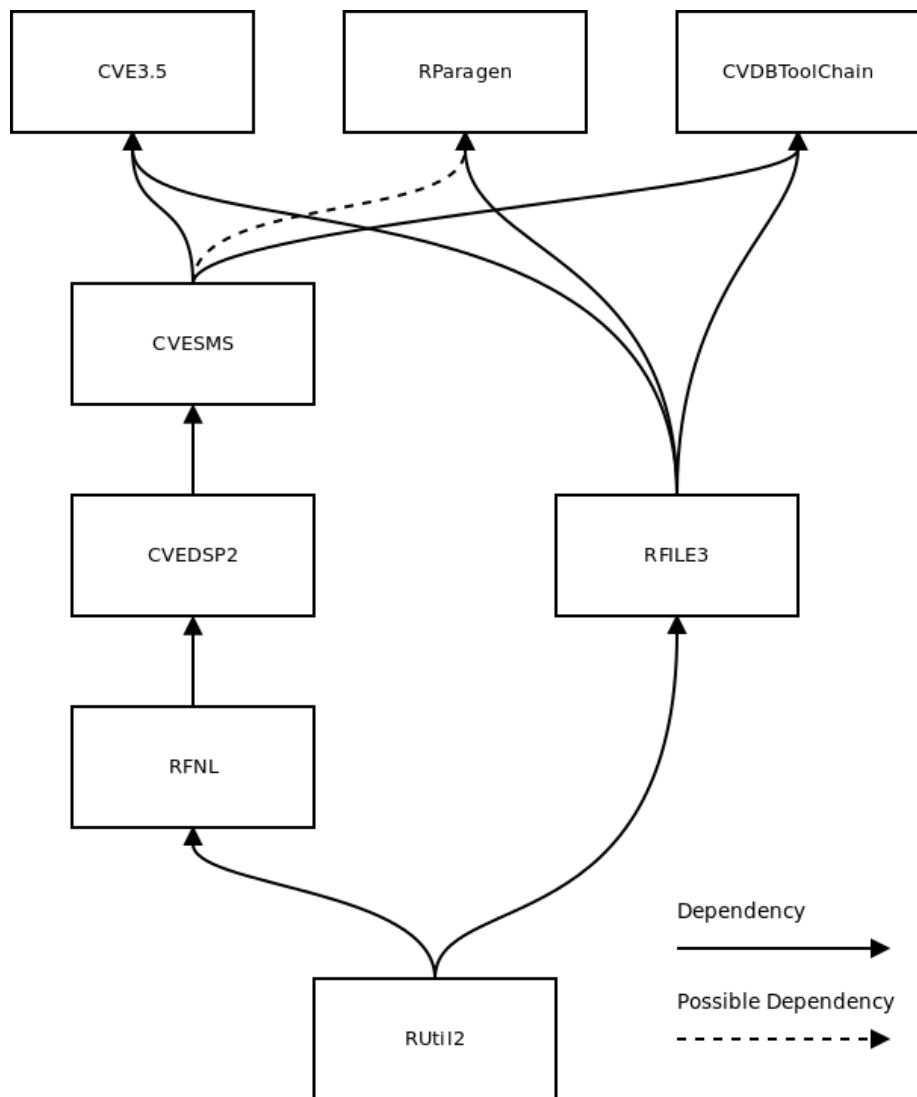


Figure 2: Dependency Chart of Rocaloid3

### 3.4 File Types

- .rvs     **Rocaloid Vocal Script:** Phonetic description of voice.
- .cvs     **CyberVoice Script:** Detailed soundfont-specific description of voice.
- .cvdb    **CyberVoice DataBase:** A soundfont particle which contains a diphone or a single phoneme.

### 3.5 For Developers

We are glad you are willing to contribute to Rocaloid.

To join the development team, please send an email to Sleepwalking and describe yourself:

`sleepwalking@rocaloid.org`

You can also join the IRC channel **#Rocaloid** in Freenode.