Work plan

Reading the original audio

The program uses the name of the file containing the recorded sound as input data. Supported files are Waveform Audio (.wav) format with one audio channel (mono) and 16-bit depth. There are no restrictions on the sampling frequency.

Construction of a spectrogram of an audio signal

The spectrogram of an audio signal is the dependence of the sound spectrum on time. The spectrogram is constructed using the discrete Fourier transform.
The original signal is divided into "frames" (segments of fixed length), to each of which the Fourier transform is applied separately. From the result of the transform, the dependence of the intensities of individual sound frequencies on time is calculated.

The main parameters of the described process are the frame length and the length of the signal fed to the Fourier transform input. These values may differ due to the fact that a part of the original signal (frame) may be supplemented with zeros to increase the frequency resolution of the spectrogram. To increase the time resolution of the spectrogram, frame overlapping is used - dividing into frames in such a way that each subsequent frame begins before the previous one ends.

Additive sound synthesis

Additive synthesis involves creating sound by adding sound waves of a fixed frequency. The main parameters of the synthesis algorithm are the frequencies of the waves, their phases, and the functions of changing the amplitudes of the waves.

Recreating sound

The sound reconstruction algorithm iteratively reconstructs the original sound. At each iteration, the optimal parameters of a single sound wave are selected using a genetic algorithm. Not only this wave, but also all the sound waves obtained at the previous steps of the algorithm participate in the sound synthesis. The reconstruction algorithm ends when at the next step it is not possible to select a sound wave that allows reducing the difference between the synthesized sound and the target one.

Genetic Algorithm Operation

To solve the problem of minimizing the difference between sounds, a genetic algorithm is used.

The main parameters of the algorithm are the population size and the stopping criterion. The following condition is used as the latter: the algorithm stops if, over a certain number of generations, the fitness of the fittest genotype has not increased by the expected minimum acceptable value.

Objective function

The objective function of the genetic algorithm is the function of the difference between two sounds.
Formally, this function calculates, based on the spectrograms of the input sound signals, a certain real number reflecting the magnitude of the difference between the sounds.

Parameters contained in the genotype

The parameters contained in the genotype must be the parameters used in additive sound synthesis: frequency, phase, and amplitude change function. All parameters are stored in real numbers.

Crossbreeding

The selection of genotypes for crossbreeding is carried out using an algorithm known as "Tournament Selection". Several genotypes are randomly selected from the entire population, among which the fittest is selected and used as an ancestor. When crossing, two ancestral genotypes give rise to two offspring genotypes. The offspring genotypes depend on the ancestral genotypes as follows: each offspring gene is a normally distributed random variable with a mathematical expectation equal to the arithmetic mean of the corresponding ancestral genes, and a standard deviation equal to half the difference in the ancestral gene values.

When generations change, only a part of the entire population, determined by the degree of crossing, participates in crossing.

Mutation

The main parameter of mutation is the mutation degree — a value showing the probability of mutation of one genotype in a new generation. Accordingly, the number of mutations occurring during generational change depends on the mutation degree. Mutation of an individual gene is implemented by replacing the initial value with a random variable that has a normal distribution with a mathematical expectation equal to the initial value and a standard deviation equal to half the difference between the maximum and minimum permissible values of the gene.

Recording results

As a result, the program writes the synthesized sound to a file for the possibility of comparing it with the original and assessing the quality of the work. In addition, the program writes a description of the sound synthesis process in a separate file in the Csound sound synthesizer description language. The latter allows you to use the result of the program to change the sound by duration and frequency (timbre) without distortion.