ROBOTIZATION EFFECT USING PHASE VOCODER PROCESSING

Digital Audio Systems, DESC9137, 2017
Lab Report

Nicolas Ramírez-Guevara SID460038045
Graduate Program in Audio and Acoustics
Sydney School of Architecture, Design and Planning, The University of Sydney

1. INTRODUCTION AND PROBLEM DESCRIPTION

Vocal effects have been among digital processing for many years, whether they are produced for musical purposes, radio, broadcasting or other media. One particular effect that arises among the others is the robotization effect due to its particular sonic characteristics and interesting sound. The robotization effect was very famous in musical contexts during the 60's and 70's mainly because of its implementation in many popular music. The basic principle of the robotic effect relies on the phase vocoder, which splits an input signal into bands before the processing, in order to achieve this, a representation in the frequency domain is necessary which is achieved by using a window function in conjunction with the Short Time Fourier Transform. Although many robot-robot effect devices have been available during the years, none of the allowed for a wide variation of effects thanks to selection of window function, window sizes or hop sizes, which in conjunction create a vast variation in the resulting signal which then can be used for diverse purposes with great success.

2. SPECIFICATION

The desired robot effect is achieved by the means of phase vocoding, which is a digital signal process that splits an input signal into a representation of it in time and frequency. The basic principle consists in analyzing and representing the signal, followed by a transformation of that representation before reconstructing it again into sound. Segments of the signal are created by multiplying the input audio by a window of finite length, which are then transformed to the frequency domain via FFT and then reconstructed via IFFT and represent them in the time. The desired effect is achieved by applying zero phase to every component of the segments of the signal, yielding a fixed pitch in the resulting audio.

A Short Time Fourier Transform (STFT) is prefered when processing audio due to its improved computational requirements. As shown in Figure 1, the STFT takes a time varying spectrum of a signal x(m) with frequency index 0 = *k* = *N* - 1 and time index *n* and represents it by its phase f(*n*, *k*) and magnitude |*X*(*n*, *k*)| components (Zölzer, 2011). The signals is then weighted by a window *h*(*n* - *m*) with finite length at each time index.

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)h(n-m)W_N^{mk}, \quad k = 0,1,\ldots,N-1$$

$$W_N = \mathrm{e}^{-j2\pi/N}$$

$$= X_R(n,k) + jX_I(n,k) = |X(n,k)| \cdot \mathrm{e}^{j\varphi(n,k)}.$$

Figure 1. Mathematical representation of STFT of a signal x(m) (Zölzer, 2011).

The length of the window is directly related to the resolution and quality of the final effect, which is why is key for the implementation of the proposed function. The window perform a fade in and out in the analysis stage of the processing in order to avoid artefacts and undesired modifications in the resulting signal. The performance of different windows relies heavily on its magnitude and phase response. As for this effect the phase response is crucial, allowing window selection provides a wider range of effects with different sonic characteristics. In general terms, the width of the principal lobe of a particular window is inversely proportional to the smoothness of its sides, generating differences in frequency resolution. Higher sides produce good frequency resolution, whereas smooth sides result in a drop in resolution (De Götzen and partners, 2000.)

Finally, the two remaining parameters of the function are as important as the window selection and they are the window size and the hop size. On one hand, the resolution of the STFT is directly related to a higher window size, which ultimately will create a smoother effect and on the other hand, higher hop sizes will yield to a reduction in signal information.

3. IMPLEMENTATION

The implementation of the effect is defined by the function that creates a robotization of a signal as described as follow:

```
robot_effect = my_robot_effect
(In_wave,fs,analysis_step,win_size,opt
ion);
```

The function takes an input signal and defines a hop size (`analysis_step`), a window size (`win_size`) and different options of windows (`option`).

The parameters that modify the signal are described as follow:

`analysis_step:` This section of the function defines the hop size that will determine the resolution of the resulting audio. The recommended values are between 50 and 250 for a signal with higher frequency content and 250 to 1000 for a reduction in audio information with is translated into lower frequency content.

`win_size:` This parameter defines the size of the window that will be applied in the analysis and resynthesis stages of the processing. It is suggested for the user to use a number which is a power of 2, (2^N) to allow maximum computational efficiency when performing the STFT. The recommended values for this parameter are between 2 and 8192. Lower values of window size allows lower resolution in the STFT, which in this case creates interesting and usable effects. With higher values, the resolution is improved and clearer audio is achieved.

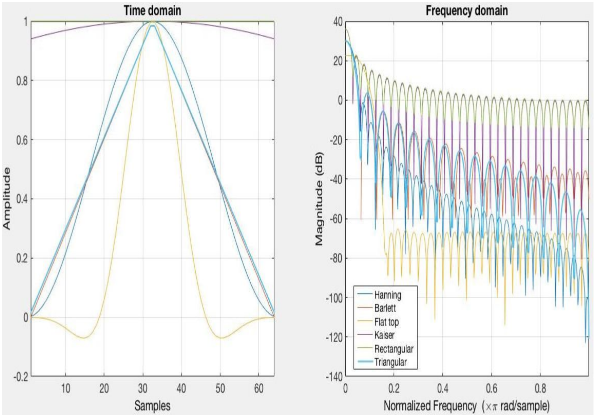

Figure 2. Characteristics of the window functions that the user can select.

`option:` This parameter allows the user to select between 6 different kinds of windows to perform the STFT in the synthesis and resynthesis stages. As depicted in figure 2, the differences in phase and magnitude responses of the windows define their performance as well as the sides and principal lobe. The windows that the user are able to select are:

1. Hanning:
2. Barlett
3. Flat top
4. Kaiser
5. Rectangular
6. Triangular

4. EVALUATION

The robotization effect varies greatly by changing the three parameters described above, and the results are very wide between one and the other. In order to perform an accurate assessment of the performance of the function, the three main parameters will be evaluated separately.

At first, the hop size is evaluated by contrasting values of 50 and 700. A Hanning window with a window size of 2048 samples was maintained during the two evaluations. As stated before, the sort hop size allow greater resolutions, which is reinforced in Figure 3 by the flatter response as compared to a longer hop size, which yields a loss in crucial audio information translated into deeper notches and poor response.

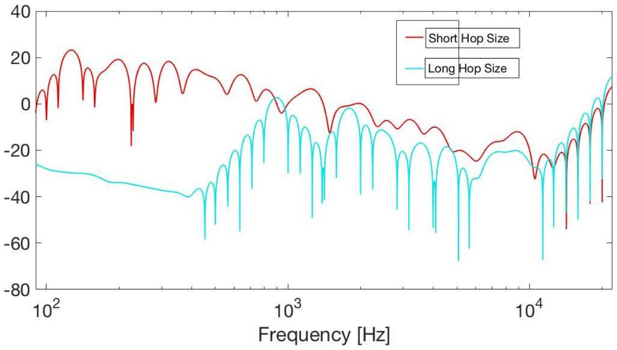

Figure 3. Frequency response comparison of a Hanning window with short and long hop sizes.

Moreover, the window size differences are presented in Figure 4, where the function was tested with a Flat Top window of 500 hop size. It is shown that the spectral characteristics of a small window (64 samples) have lower resolution than a larger window (8192.) As shown in the picture, the response of the short window creates notches and poor response on the middle frequency range, whereas the larger one presents a smoother performance through the whole spectrum.

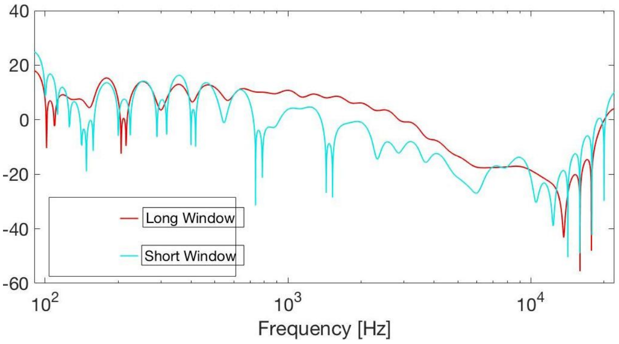

Figure 4. Frequency response comparison of a Flat Top window with 500 hop size, with small and large window size.

Finally, the spectral differences that the selection of the window type presents are very noticeable. In Figure 5, it is illustrated that the performance of a Hanning window is more stable than a Triangle Window. In the picture, it is seen that the frequency response of the Hanning window is flatter as compared to the Triangle window. Both have notches and lobes through the spectrum, however, the triangle notches in higher frequencies are deeper and the lobes at lower frequencies go higher in level.

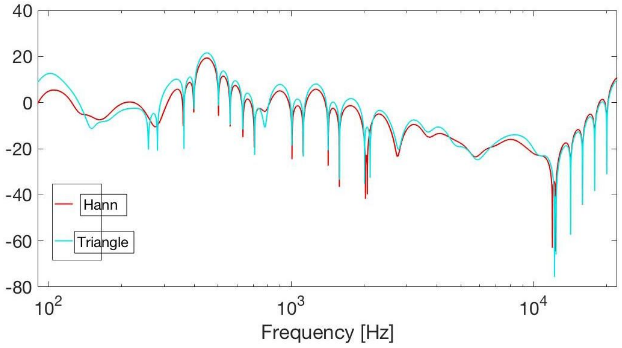

Figure 5. Frequency response comparison of a Hanning and a Triangle window.

5. REFERENCES

De Götzen, Amalia, Nicola Bernardini and Daniel Arfib. 2000. *Traditional (?) Implementations of a Phase Vocoder: The Tricks of the Trade*. Proceedings of the COST G 6 Conference on Digital Audio effects (DAFX 00). Verona, Italy.

Zölzer, Udo. 2011. *DAFX: Digital Audio Effects*. Second Edition. John Wiley and Sons.