Soundtransformations based on the Modulation Power Spectrum

Thomas Mayr[1], Robert Höldrich[2]

[1] *University of Music and Performing Arts Graz, Austria, Email: thomas-karl.mayr@student.kug.ac.at*

[2] *Institute of Electronic Music and Acoustics, Graz, Austria, Email: robert.hoeldrich@kug.ac.at*

Abstract

In this work, the Modulation Power Spectrum (MPS) is used to visualize temporal and spectral modulations of sounds. For this purpose the two-dimensional Fourier transformation (2DFT) of the spectrogram is calculated. In speech signal processing the MPS has already been shown how modifications (filtering) effect the resynthesized sound in this domain. We investigate how transformations like high-, low- and notchfiltering in the MPS-domain modify modulations in various sound materials. Moreover we propose a filter to lower or boost up- or downward frequency glides and introduce distortions like stretching and morphing in the MPS domain. After a manipulation the sound is resynthesized through a spectrogram inversion.

Introduction

Most sound signals contain various temporal and spectral ripples. These fluctuations are called spectrotemporal modulations. In [1] it has been shown which degradations of modulations in a speech signal cause a loss of intelligibility or speaker identification. The spectrogram of a sound is composed not only of isolated temporal or spectral modulations but also of combined versions of them. These combinations range over shorter or longer time durations and frequency regions and constitute important aspects of the perceived timbre, the movement of the formantstructure and the formation of syllables.

Modulations of speech and natural sounds exhibit lowpass characteristics [2] due to the mass inertia of the sound producing mechanism. In time-frequency processing, further lowpass behavior is introduced by windowing operation of the Short-Time Fourier Transform (STFT). The longer the time window, the lower the maximum detectable temporal modulation in the spectrogram.

Modulation Power Spectrum

The magnitude of the two-dimensional Fourier Transform of a spectrogram is called MPS. First the spectrogram is calculated with overlapping Hann-windows. The two-dimensional Fourier Transform decomposes the spectrogram in its Fourier components where the ripples in the spectrogram can be compared to visual gratings [2]. A calculation of the logarithm of the STFT results in a seperation of spectral and temporal modulation terms. Otherwise these terms would be linked in a multiplicative way [3]. Figure 2 shows the decomposition of different ripples after the 2DFT. The top right ripple describes a sound with a fundamental tone with constant frequency and harmonics. Such a ripple is plotted on the ordinate in the MPS which is called the τ-axis and represents spectral modulations (cycl/kHz). On the other hand a sound with amplitude fluctuations (bottom left ripple) is plotted along the abscissa which is called the f_{tmod}-axis and represents temporal modulations (Hz). The 2DFT distinguishes between up- and downward frequency glides of the ripples which represent up- and downward glissandi of the fundamental tone and its harmonics. As the 2DFT results in Fourier pairs, positive spectral modulations are mapped in the second and fourth quadrant while negative spectral modulations are mapped in the first and third quadrant. Due to the point symmetry of real valued ripples it is not necessary to view all four quadrants but only the first and the second.

The resolution of the MPS is given by the length of the windows of the STFT and the hopsize. Longer windows and too high hopsizes decrease the maximum temporal modulation but increase the maximum resolvable spectral modulation (time-frequency tradeoff [4]).

The 2DFT of a sung baritone vowel is depicted in Figure 1. The time span between 0.3 and 0.5 seconds is transformed into the MPS at 4 cycl/kHz (spectral) and 0 Hz (temporal). This is because four partials are within the range of one kHz. Lower fundamental frequencies result in higher spectral modulations and vice versa. Areas where the fundamental frequency increases over a time period are transformed into the second quadrant and areas where the frequency decreases are transformed into the first quadrant.

Filtering in the MPS-Domain

The signal chain to manipulate the MPS is depicted in Fig. 3. After the 2DFT of the log-magnitude of the STFT the MPS can be manipulated with different filtering techniques.

In [1] lowpass, highpass and notch filtering were introduced to investigate if a degradation of modulations causes a loss in intelligibility of speech. We defined a lowpass filtering process with

$$\mathbf{S_m}(t', \omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}(\mathrm{f_{tmod}}, \tau) \circ \hat{\mathbf{F}}_{\mathrm{lp}}(\mathrm{f_{tmod}}, \tau)\right], \qquad (1)$$

where $\mathbf{S_m}$ is the resulting spectrogram, $\hat{\mathbf{S}}$ is the frequency representation of the original spectrogram which is denoted with a hat and $\hat{\mathbf{F}}_{\mathrm{lp}}$ is the filter in the frequency domain respectively. $\mathcal{F}_{\text{-1,-1}}$ denotes the inverse 2DFT.

We introduced other filtering techniques to manipulate musical sounds. For example in order to manipulate a

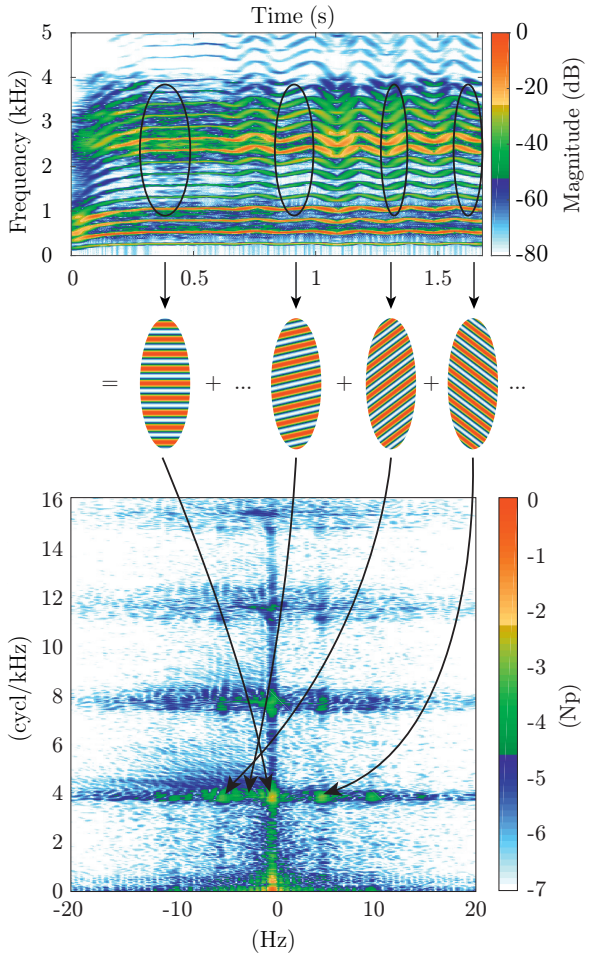

Figure 1: MPS of a sung baritone vovel. Regions of the spectrogram where the fundamental frequency of a harmonic structure doesn't change over time can be found along the τ-axis. The exact position along the τ-axis arises from the number of partials in the range of one kHz. In this example the fundamental frequency at around 250 Hz results in 4 cycles/kHz. The energy at ± 5 Hz (temporal) arises from the Fourier transform along the time axis and represent the periodicity over time.

frequency vibrato a Gaussian filter (Eq. 2) is used to lower or boost either the up- or downward motion of the fundamental frequency, independent of other tonal parameters.

$$\mathbf{S_m}(t',\omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}} \circ \hat{\mathbf{F}}_{\text{gauss}}^{\mu_f,\mu_\tau,\beta}\right]. \tag{2}$$

where $\circ$ denotes the Hadamard product and the two-dimensional Gaussian filter is defined by

$$\hat{\mathbf{F}}_{\text{gauss}}^{\mu_f,\mu_\tau,\beta} = 1+(\beta-1)\cdot\exp\left(-\frac{1}{2}(\mathbf{x}-\mu_f)^T\Sigma^{-1}(\mathbf{x}-\mu_\tau)\right). \tag{3}$$

μ_f describes the mean value of temporal modulations and μ_τ the mean value of the spectral modulations. β is a factor to lower or boost the amplitude of a Gaussian bell. A multiplication in the frequency domain where β is higher than 1 causes an expansion of the values in the resulting spectrogram and a factor $\beta < 1$ a compression

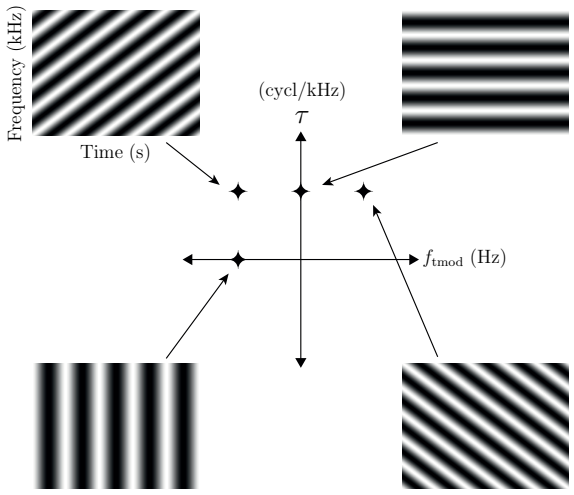

Figure 2: Definition of the modulation spectrum. A 2DFT decomposes a spectrogram in its Fourier components. The four pictures represent ripples in a spectrogram of a sound. After the 2DFT these ripples will be mapped into the two-dimensional plane of the Fourier spectrum. The abscissa represents temporal modulations (Hz) and the ordinate represents spectral modulations (cycl/kHz).

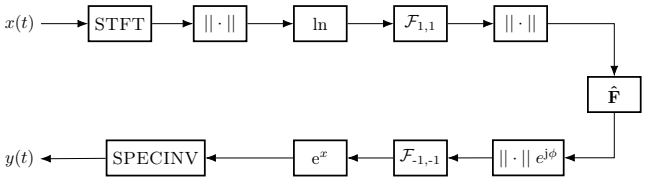

Figure 3: Signal chain and filtering process in the MPS-Domain. First the spectrogram of a sound is calculated. After the 2DFT a filter manipulates the magnitudes of the frequency representation of the spectrogram. After the inverse 2DFT and the inversion of the amplitude a spectogram inversion calculates the resultant sound.

of the values in the spectrogram. An addition of a Gaussian bell in the MPS-domain will lower or boost values in the spectrogram in a linear way.

To combine the formant structure of one sound with the parital tone structure of another sound we do lowpass filtering of the first sound to extract the formant structure and add the highpass partial tone structure of another sound. This results in an imprint of the formant structure of the first sound onto the other sound which we call a **morphing** process which is denoted with:

$$\mathbf{S_m}(t',\omega) = \mathcal{F}_{\text{-1,-1}}\left[\hat{\mathbf{S}}_1 \circ \hat{\mathbf{F}}_{\text{hp}} + \hat{\mathbf{S}}_2 \circ \hat{\mathbf{F}}_{\text{lp}}\right]. \tag{4}$$

A *distortion* where the MPS is stretched along the τ-axis causes a displacement of the formant structure on the one hand and also moves the partial tone structure to higher or lower regions. This is compareable to the sound of a pitchshifter. Another way to manipulate the formant structure independent of the harmonic struture is to seperate a sound into a lowpass and highpass version with the same cutoff-frequencies on both the temporal

and spectral axis. After streching the highpass filtered part towards higher values in the MPS along the τ-axis the formant structure remains the same but the fundamental frequency in the spectrum moves down to lower frequencies. This is depicted in Fig. 4.

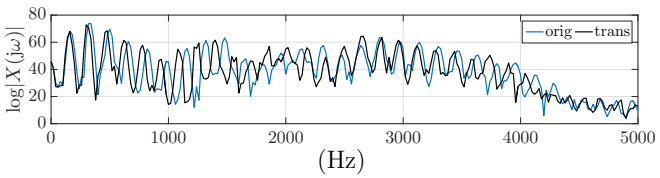

Figure 4: Spectrum of a sound after stretching the partial tone structure of the MPS along the τ-axis. The formant structure remains.

After a manipulation we get the new spectrogram through an inverse two dimensional Fourier transform and inverting the amplitude with e^x. The resynthesized sound is reached through the Griffith and Lim algorithm which is described in [5].

Tetration

As a future work to expand the signal chain from Fig. 3 we use tetration [6] to continously switch between the logarithm- and exponential function. Functions for natural tetration were implementend in MATLAB from [7] and [8] to manipulate the magnitudes of the STFT. Eq. 5 is referred as iterated exponential where tet$(\cdot)$ is the tetration with base e and ate$(\cdot)$ the inverse tetration.

$$\exp^c(z) = \mathrm{tet}\Big(c + \mathrm{ate}(z)\Big) \tag{5}$$

The factor c denotes how often the exponential is iterated and this allows a seemless warping between the logarithm ($c = -1$), the linear function ($c = 0$) and the exponential ($c = 1$), depicted in Fig. 5.

Conclusion and outlook

We presented some new filtering techniques for the MPS to manipulate the spectrotemporal modulations of a sound. A combined plot shows both spectral and temporal modulations. To reach such a plot a 2DFT of a spectrogram is calculated. In this frequency representation of the spectrogram we manipulate the spectrotemporal modulations through filtering techniques. As a new way to manipulate the vibrato of a sound independent of other parameters we proposed a filter which lowers or boosts up- or downward frequency glides in the spectrogram. For resynthesizing the sound after the manipulation we used the Griffth and Lim algorithm.

As a future work more filtering techniques and manipulations in the MPS domain should be investigated. New filter shapes could extend the filtering process. Moreover it should be explored what happens if the stopband of a filter is not reduced to zero.

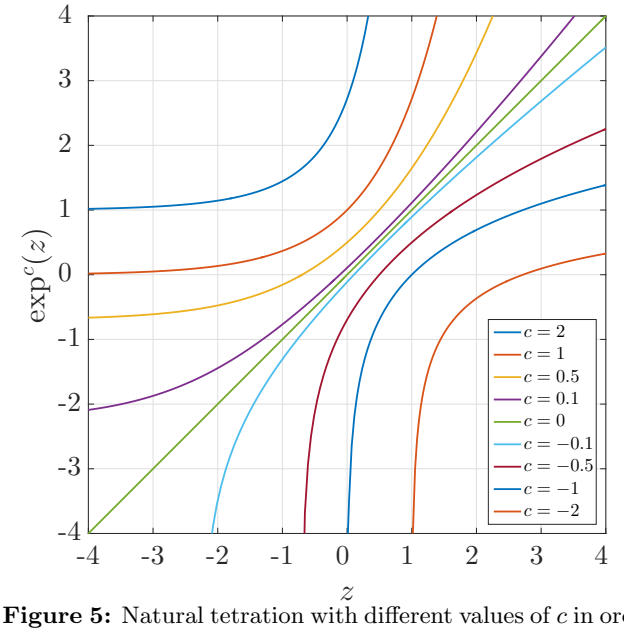

Figure 5: Natural tetration with different values of c in order to switch between the logorithm and the exponential function continously.

References

[1] T. M. Elliott and F. E. Theunissen, *The Modulation Transfer Function for Speech Intelligibility.* PLoS Comput Biol 5(3): e1000302. doi:10.1371 journal.pcbi.1000302, 2009.

[2] N. C. Singh and F. E. Theunissen, *Modulation spectra of natural sounds and ethological theories of auditory processing.* Department of Psychology and Neuroscience Institute, University of California, Berkeley, 3210 Tolman Hall, Berkeley, California 94720-1650, 2003.

[3] T. M. Elliott, L. S. Hamilton, and F. E. Theunissen, *Acoustic structure of the five perseptual dimensions of timbre in orchestral instrument tones.* Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, 2012.

[4] J. E. Wilhjelm, *Bandwidth Expressions of Gaussian Weighted Chirp*, vol. 25. 1993.

[5] D. W. Griffin, Jae, S. Lim, and S. Member, *Signal estimation from modified short-time fourier transform.* 1984.

[6] D. Kouznetsov, *Tetration as special Function.* Institute for Laser Science, University of Electro Communications, researcher 1-5-1 Chofugaoka, Chofushi, Tokyo, 182-8585, Japan, 2010.

[7] "fsexp.cin is routine for the fast evaluation of natural tetration." last accessed 11.12.2016.

[8] "fslog.cin is routine for the fast evaluation of natural arctetration." last accessed 11.12.2016.