OBTAINING NARROW TRANSITION REGION IN STFT DOMAIN PROCESSING USING SUBBAND FILTERS

Meng Guo

Oticon A/S
Kongebakken 9
DK-2765 Smørum
Denmark
email: megu@oticon.com

Bernhard Kuenzle

Bernafon AG
Morgenstrasse 131
3018 Bern
Switzerland
email: bekn@bernafon.com

ABSTRACT

The short-time Fourier transform (STFT) and the inverse short-time Fourier transform are often used in signal analysis, modification, and synthesis. In this work, we focus on the effect of overlapping STFT subbands in relation to signal modification and synthesis. We illustrate that the subband overlapping can impose a negative mixing effect on STFT domain signal processing such as bandlimited frequency shifting. We propose a subband filtering method, by using subband low-pass and high-pass filters in the affected frequency region, to reduce the mixing effect locally without changing the general STFT processing. We show in simulation experiments that our subband filtering method is efficient; the bandwidth of the affected frequency region is typically reduced by a factor of $2-3$, and there is significantly less signal distortion in the affected frequency region.

***Index Terms*—** Short-time Fourier transform, STFT, ISTFT, frequency shifting, transition region, subband filtering.

1. INTRODUCTION

The short-time Fourier transform (STFT) and inverse short-time Fourier transform (ISTFT) are commonly used in signal processing to analyze and modify signals [1], often in the form of modulation filter banks [2]. The STFT is the analysis part and is often based on a sliding-window and discrete Fourier transform (DFT), and the ISTFT is the synthesis part and is typically carried out using the overlap-add method. The STFT and ISTFT processing are very useful in many audio signal processing applications, such as noise reduction, source separation, pitch modification, and echo cancellation.

The STFT divides the time domain signal $x(n)$, where n is the time index, into the STFT domain signals $X(k,l)$ with k and l as the STFT domain time and frequency subband indices; signal processing can then be carried out in the STFT domain, on $X(k,l)$, before the ISTFT is used to reconstruct the time domain signal $y(n)$; ideally, a perfect reconstruction is possible, i.e., $y(n) = x(n)$.

The window functions applied in the STFT and ISTFT have a significant influence on the signal processing applied in the STFT domain [3]. In this work, we focus on the STFT window function and its impact on the processing in the STFT domain. Different frequency subbands l are bandlimited based on the basis window function. Thus, the shaping and the overlap between frequency subbands are highly dependent on the choice of this window function.

Fig. 1 shows an example of the first four frequency subbands, in the STFT domain, with a 32-point Hamming window, and the DFT size of 32. We observe that there are relatively big overlapping areas between different subbands. In the case that we would like to do bandlimited processing, e.g., the processing A for subband 0 and 1 and the processing B for subband 2 and above, effects from both processing clearly appear in the transition region around the cut-off frequency, in this case between the subband 1 and 2 at approximately 0.1π; within this region, the resulting processing is an undesired mixture of the processing A and the processing B with almost equal weight; far away from the cut-off frequency the mixing effect gets negligible, and we obtain approximately either processing A or B.

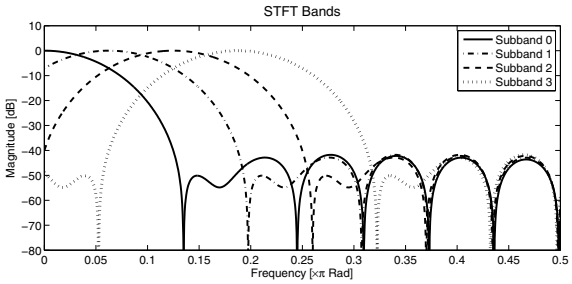

Fig. 1. An example of overlapping STFT frequency subbands.

When carrying out bandlimited processing in the STFT domain, the mixing effect around the cut-off frequency cannot be avoided, due to the design limitations in window functions. Depending on the processing in the transition region, the effect of this mixing of different processing can be unacceptable, as we demonstrate later.

To minimize the mixing effect, the overlap between subbands can be designed to be smaller by using different window functions or other type of filter banks [4,5]. However, very often the STFT processing is a given part of a complete system with already specifically defined properties, e.g., group delay and band overlap, and hence we wouldn't change the STFT processing, nevertheless we would like to locally reduce the mixing effect in the transition region.

An example of this situation is a frequency shifting operation implemented in the STFT domain. The frequency shifting can be useful to improve system stability in closed-loop systems [6–10], but it can also introduce audible artifacts to human listeners [11, 12], especially at the low frequencies, thereby the frequency shifting is usually only applied above a cut-off frequency (around 1 kHz) [13, 14]. However, around this cut-off frequency, audible artifacts can occur due to the mixing effect. In the following, we explain more on the mixing effect, and we propose a concept to reduce this effect by using the frequency shifting operation as an example.

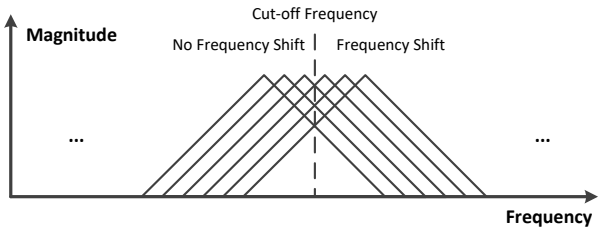

Fig. 2. An illustration of bandlimited frequency shifting in the STFT domain, where it is only applied in the frequency subbands (shown with triangles) with center frequencies above the cut-off frequency.

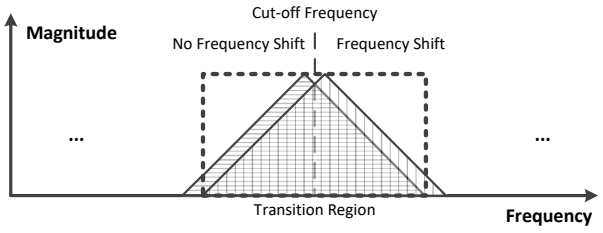

Fig. 3. An illustration of neighbor subbands around the cut-off frequency. The box indicates the relatively broad transition region.

2. AN EXAMPLE BANDLIMITED PROCESSING OF FREQUENCY SHIFTING

In this section, we illustrate the mixing effect in the STFT domain with an example bandlimited frequency shifting operation.

Let $X'(k,l)$ be the frequency shifted version of $X(k,l)$, and in the STFT domain it can simply be obtained as,

$$X'(k,l) = X(k,l)e^{j2\pi f'(l)k\frac{R}{f_s}}, \tag{1}$$

where $f'(l)$ denotes the amount of frequency shifting in Hz at different STFT subbands l, f_s is the time domain sampling rate in Hz, and R is the hop size in the STFT processing.

The frequency shifting is useful for feedback cancellation applications [9], however, it is often used above a cut-off frequency f_c only, as sound quality can be degraded by frequency shifting [13]. This bandlimited processing is illustrated in Fig. 2, and in this case the cut-off frequency f_c is between two STFT subbands, where $f'(l) = c_1$ and c_2 for the subbands l below and above the cut-off frequency, respectively; very often, $c_1 = 0$ and $10 \leq c_2 \leq 25$ for feedback cancellation applications.

In this example case with frequency shifting, the re-synthesized signal contains a mixture of the frequency shifted signal and the original signal without frequency shifting, both with almost equal weight around the cut-off frequency; the result of this can be an audible modulation effect [15], which is undesirable as it can degrade sound quality perceived by human listeners. The modulation effect is illustrated later in Fig. 9(b) in Sec. 4.

Fig. 3 illustrates the affected frequency region referred to as the transition region, for simplicity the illustration is only made for the two closest neighbor subbands.

In Sec. 3, we propose a subband filtering method, in the framework of bandlimited processing in the STFT domain, to obtain a narrower transition region around the cut-off frequency and thereby minimize the mixing effect due to subband overlapping; hence it also reduces the modulation effect in the bandlimited frequency shifting.

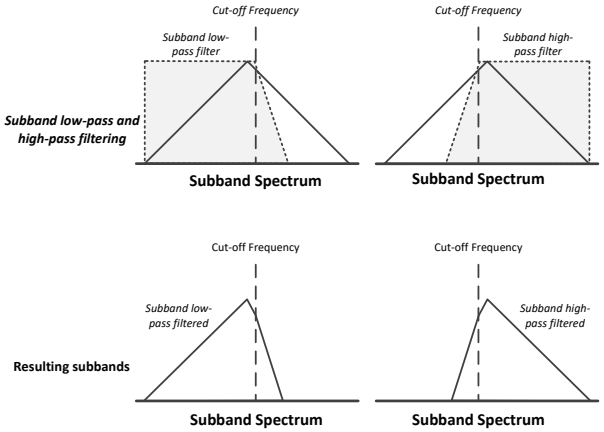

Fig. 4. An illustration of the subband filtering method applied to frequency subbands. In the two upper plots, a subband low-pass filter and a subband high-pass filter are applied to these frequency subbands to create the resulting subbands with steeper slopes, shown in the two lower plots. The designation of low/high-pass filter is defined in comparison to the subband cut-off frequency.

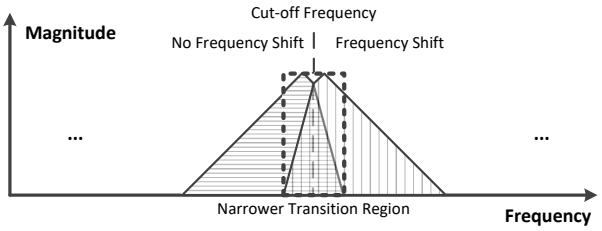

Fig. 5. An illustration of neighbor subbands around the cut-off frequency. The box indicates the narrower transition region.

3. NARROW TRANSITION USING SUBBAND FILTERS

In this section, we demonstrate a subband filtering method to obtain a narrow transition region. The subband filtering is applied to the neighbor subbands around the cut-off frequency, so that the resulting subbands close to the cut-off frequency have steeper slopes than before. The basic idea of this method is illustrated in Fig. 4 for two frequency subbands.

More specifically, the slopes of each subband in the STFT domain are originally defined by the STFT window function (finite impulse response filter), and we make use of additional subband infinite impulse response (IIR) filters to further increase the slopes of the neighbor frequency subbands around the cut-off frequency. Hence, the bandwidth of the transition region is reduced as the result. The subband low-pass and high-pass filters illustrated in Fig. 4 are non-symmetric around the center frequency; this is generally necessary in the STFT domain, only in special cases these filters can be symmetric. Thus, the filters have typically complex valued coefficients.

The increased slopes of subbands lead to a narrower transition region when re-synthesizing the processed signal in the STFT domain. Fig. 5 illustrates the narrow transition region when applying the subband filtering to the neighbor frequency subbands around the cut-off frequency; the new transition region is less than half of the size compared to the original one illustrated in Fig. 3.

It is also possible to apply subband filtering on more subbands near the cut-off frequency to obtain an even bigger improvement.

The subband filtering requires additional computations, however, as it is done in the subband domain with typically much reduced sampling rate, the increase in computational complexity is limited.

Our design of the subband filters is based on spectral rotations of a regular IIR filter with passband from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, and its transfer function $H(z)$ in the pole-zero representation is

$$H(z) = g\frac{(1-q_1z^{-1})(1-q_2z^{-1})\dots(1-q_Mz^{-1})}{(1-p_1z^{-1})(1-p_2z^{-1})\dots(1-p_Nz^{-1})}, \quad (2)$$

where g is a scaling factor, and $q_1, \dots, q_M$ and $p_1, \dots, p_N$ are the M zeros and the N poles of the filter, respectively.

Let $\mathbf{q} = [q_1, q_2, \dots, q_M]^T$ and $\mathbf{p} = [p_1, p_2, \dots, p_N]^T$, the subband low-pass filter zeros $\mathbf{q}^{lp} = [q_1^{lp}, q_2^{lp}, \dots, q_M^{lp}]^T$ and poles $\mathbf{p}^{lp} = [p_1^{lp}, p_2^{lp}, \dots, p_N^{lp}]^T$ are obtained by

$$\mathbf{q}^{lp} = |\mathbf{q}| \odot e^{j(\angle\mathbf{q}+\omega_s-\frac{\pi}{2})}, \quad (3)$$

$$\mathbf{p}^{lp} = |\mathbf{p}| \odot e^{j(\angle\mathbf{p}+\omega_s-\frac{\pi}{2})}, \quad (4)$$

where $|.|$ and $\angle.$ determine absolute value and phase angle of each element in a vector, and $\odot$ denotes element-wise vector multiplication. Furthermore, the normalized subband cut-off frequency ω_s is derived as

$$\omega_s = (\omega_c \cdot R + \pi)\%(2\pi) - \pi, \quad (5)$$

where ω_c is the normalized cut-off frequency for the frequency shifting processing, and $\%$ denotes the remainder operator.

Similarly, we obtain the subband high-pass filter zeros $\mathbf{q}^{hp} = [q_1^{hp}, q_2^{hp}, \dots, q_M^{hp}]^T$ and poles $\mathbf{p}^{hp} = [p_1^{hp}, p_2^{hp}, \dots, p_N^{hp}]^T$, as

$$\mathbf{q}^{hp} = |\mathbf{q}| \odot e^{j(\angle\mathbf{q}+\omega_s+\frac{\pi}{2})}, \quad (6)$$

$$\mathbf{p}^{hp} = |\mathbf{p}| \odot e^{j(\angle\mathbf{p}+\omega_s+\frac{\pi}{2})}. \quad (7)$$

Finally, the subband filter transfer functions $H^{lp}(z)$ and $H^{hp}(z)$ are obtained as

$$H^{lp}(z) = g\frac{(1-q_1^{lp}z^{-1})(1-q_2^{lp}z^{-1})\dots(1-q_M^{lp}z^{-1})}{(1-p_1^{lp}z^{-1})(1-p_2^{lp}z^{-1})\dots(1-p_N^{lp}z^{-1})}, \quad (8)$$

$$H^{hp}(z) = g\frac{(1-q_1^{hp}z^{-1})(1-q_2^{hp}z^{-1})\dots(1-q_M^{hp}z^{-1})}{(1-p_1^{hp}z^{-1})(1-p_2^{hp}z^{-1})\dots(1-p_N^{hp}z^{-1})}. \quad (9)$$

We now illustrate an example of creating subband low-pass and high-pass filters using a regular IIR filter. First, we design a standard low-pass $4th$ order (where $M = N = 4$) Elliptic filter [16], with passband ripple of 0.1 dB, stopband attenuation of 40 dB, and passband cut-off frequency $\frac{\pi}{2}$. Fig. 6 shows the pole-zero plot and the magnitude response of this filter.

Moreover, we consider the case that the subband cut-off frequency is at $\omega_s = -0.25\pi$ using (2)-(9). Figs. 7 and 8 illustrate the rotated pole-zero plots and the shifted magnitude responses of the subband low-pass and high-pass filters, respectively.

4. SIMULATION EXPERIMENTS

In this section, we demonstrate the effect of subband filtering through simulation experiments. In particular, we show the effect by using different frequency shifted signals; the frequency shifting is applied in the subband domain with the DFT size of 128, whereas the sampling rate is 20 kHz for the time domain signal. The STFT and ISTFT windows in this example are both Hamming

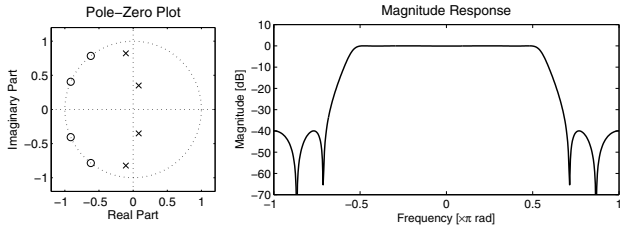

Fig. 6. An example IIR filter with the passband from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$.

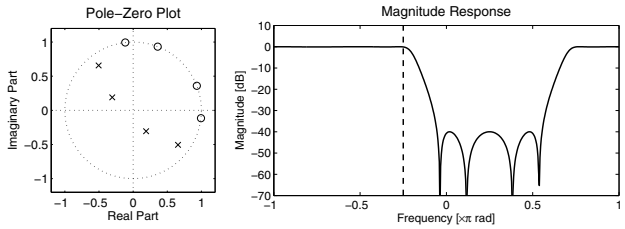

Fig. 7. An example subband low-pass filter, the dashed line shows subband cut-off frequency.

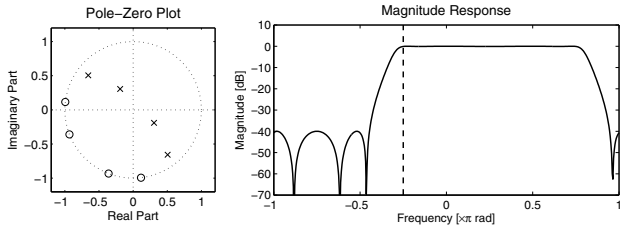

Fig. 8. An example subband high-pass filter, the dashed line shows subband cut-off frequency.

windows with 125 points, and the hop size is $R = 20$. We only apply the frequency shifting of 10 Hz from frequency subband 6, which results in the cut-off frequency $f_c \approx 0.86$ kHz.

As explained in Sec. 2, due to the mixing effect, the frequency region around the cut-off frequency f_c consists of a mixture of the unprocessed and frequency shifted signals. In the following, we demonstrate that using the subband filtering technique presented in Sec. 3 significantly reduce visible/audible modulation artifacts in this frequency region, and we apply the subband filtering in four subbands, two on each side of the cut-off frequency. The applied subband filters are created using the example Elliptic filter as mentioned in Sec. 3. Listening tests shown that the non-linear phase responses of these filters had no perceptual impact in audio applications.

Fig. 9 shows an example of a chirp signal with linear frequency increase from 0.4 kHz to 1.2 kHz. Fig. 9(a) shows the spectrogram of the unprocessed chirp signal, whereas Fig. 9(b) shows the processed chirp signal when bandlimited frequency shifting has been applied in the STFT domain, where the cut-off frequency is $f_c \approx 0.86$ kHz. Fig. 9(c) shows the processed signal when the subband filtering has been applied in addition to the frequency shifting.

We clearly observe that due to the mixing effect, both processed signals shown in Figs. 9(b) and 9(c) suffer from modulation artifacts around the cut-off frequency $f_c \approx 0.86$ kHz. However, the further subband filtering processed signal shown in Fig. 9(c) has much less modulation, and the affected frequency region has a bandwidth of about 100 Hz, which is much narrower compared to the processed signal without subband filtering shown in Fig. 9(b), where the affected frequency region covers more than 300 Hz.

When listening to both processed chirp signals, it is clear that the one with subband filtering has much less audible modulation,

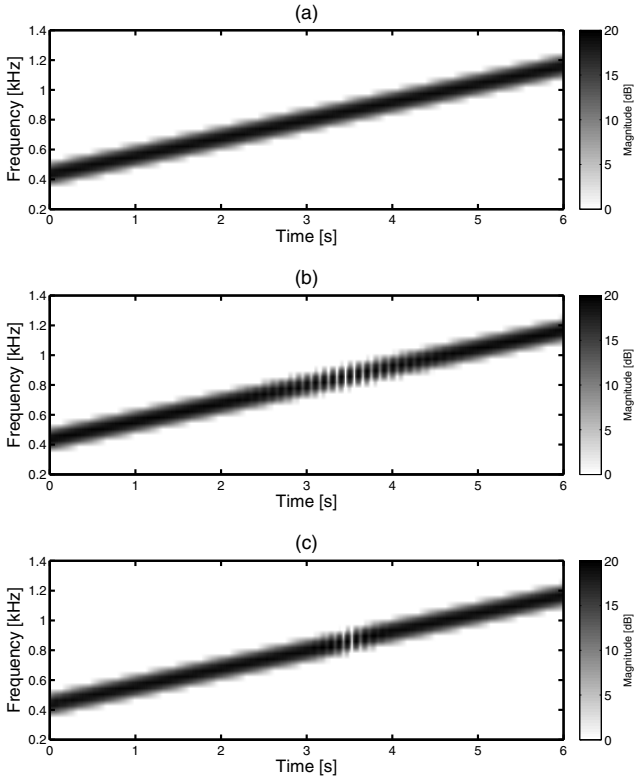

Fig. 9. An example of chirp signal, and frequency shifting is applied above the cut-off frequency of $f_c \approx 0.86$ kHz. The mixing effect is observed around the cut-off frequency. (a) Unprocessed chirp signal. (b) Without subband filtering. (c) With subband filtering.

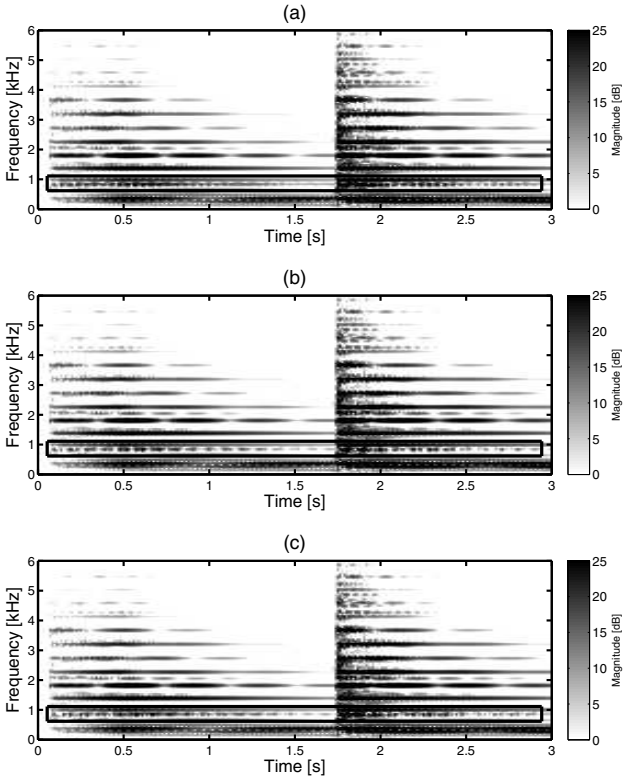

Fig. 10. An example of church bell ringing sound, and frequency shifting is applied above the cut-off frequency of $f_c \approx 0.86$ kHz. The mixing effect is observed around the cut-off frequency indicated by the box. (a) Unprocessed bell sound. (b) Without subband filtering. (c) With subband filtering.

and the affected frequency region is much narrower.

Another sound example of a church bell ringing is illustrated in Fig. 10. Fig. 10(a) shows the spectrogram of the unprocessed recording of the church bell ringing sound. We observe from Fig. 10(b) that the processed signal without subband filtering is clearly different to the unprocessed reference signal around $f_c \approx 0.86$ kHz (within the box), where a modulation over time occurs due to the mixing effect. In Fig. 10(c), we observe that the processed signal with further subband filtering, although not perfect, is closer to the original unprocessed signal shown in Fig. 10(a).

More interestingly, when listening to the processed signal in Fig. 10(b), there is a clearly audible reverberated tail of the bell ringing sound around $0.8 - 0.9$ kHz, where the audible difference between Fig. 10(a) and Fig. 10(c) is very small if at all.

Finally, we demonstrate the mixing effect around the cut-off frequency using the power spectral densities (PSDs) of different white noise sequences. Fig. 11 shows the PSDs of a white noise sequence and both processed versions with frequency shifting.

Without subband filtering, we observe that the mixing effect in the STFT domain leads to a spectral dip of more than 3 dB around the cut-off frequency of $f_c \approx 0.86$ kHz, and the PSD differs from the unprocessed white noise PSD around $0.6 - 1.0$ kHz. With additional subband filtering, the dip is both shallower and narrower at the same time, about half of the size; however, its PSD differs from the unprocessed white noise PSD around $0.3 - 1.3$ kHz, which is a bigger region compared to the processing without subband filtering. Nevertheless, the deviation in the regions of $0.3 - 0.6$ kHz and $1.1 - 1.4$ kHz are very small and less than 0.5 dB, and in practice it does not lead to audible signal distortions.

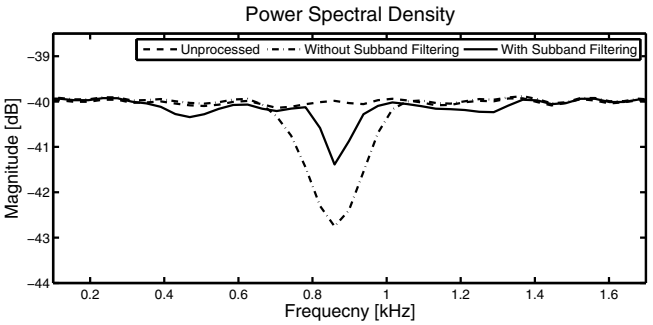

Fig. 11. An example of white noise sequence and subband frequency shifting processing without and with subband filtering.

5. CONCLUSIONS

We demonstrated that the unavoidable overlapping in STFT subbands can create undesired signal distortion referred to as the mixing effect, e.g., the modulation over time when using bandlimited frequency shifting in the STFT domain. We presented a general approach by using subband filtering to reduce the mixing effect locally. We showed through simulation experiments that the proposed subband filtering method can significantly reduce the mixing effect and hence signal distortions, as it typically reduces the bandwidth of the affected frequency region by a factor of $2 - 3$.

6. REFERENCES

[1] M. M. Goodwin, "The STFT, sinusoidal models, and speech modification," *Springer Handbook of Speech Processing*, pp. 229–258, Nov. 2007.

[2] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. Upper Saddle River, NJ, US: Prentice Hall, Oct. 1992.

[3] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[4] A. N. Akansu and M. J. T. Smith, *Subband and Wavelet Transforms*. Berlin, Heidelberg, Germany: Springer, Oct. 1995.

[5] M. M. Goodwin, "Realization of arbitrary filters in the STFT domain," in *Proc. 2009 IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2009, pp. 1–5.

[6] L. N. Mishin, "A method for increasing the stability of sound amplification systems," *Sov. Phys. Acoust.*, vol. 4, pp. 64–71, 1958.

[7] M. R. Schroeder, "Improvement of feedback stability of public address systems by frequency shifting," *J. Audio Eng. Soc.*, vol. 10, no. 2, pp. 108–109, Apr. 1962.

[8] T. van Waterschoot and M. Moonen, "Assessing the acoustic feedback control performance of adaptive feedback cancellation in sound reinforcement systems," in *Proc. 17th European Signal Process. Conf.*, Aug. 2009, pp. 1997–2001.

[9] ——, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, Feb. 2011.

[10] M. Guo, S. H. Jensen, J. Jensen, and S. L. Grant, "On the use of a phase modulation method for decorrelation in acoustic feedback cancellation," in *Proc. 20th European Signal Process. Conf.*, Aug. 2012, pp. 2000–2004.

[11] B. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Bingley, UK: Emerald Group Publishing Limited, Apr. 2003.

[12] C. J. Plack, *The Sense of Hearing*. Hove, UK: Psychology Press, Jun. 2005.

[13] M. Guo, S. H. Jensen, and J. Jensen, "Evaluation of state-of-the-art acoustic feedback cancellation systems for hearing aids," *J. Audio Eng. Soc.*, vol. 61, no. 3, pp. 125–137, Mar. 2013.

[14] M. Guo and B. Kuenzle, "On the periodically time-varying bias and its correction in adaptive feedback cancellation systems with frequency shifting," in *Proc. 2016 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 539–543.

[15] W. M. Hartmann, *Signals, Sound, and Sensation*. Berlin, Heidelberg, Germany: Springer, Jan. 1997.

[16] T. W. Parks and C. S. Burrus, *Digital Filter Design*. Hoboken, NJ, US: Wiley, Aug. 1987.