SPEECH IMPROVEMENT BY NOISE REDUCTION BASED ON A CONTINUOUS FOURIER TRANSFORMATION (CFT)

Michael Walker
ALCATEL
Research & Innovation
Lorenzstr. 10
D-70435 Stuttgart, Germany
Michael.Walker@alcatel.de

Abstract

Noise reduction is needed in many applications. Hands-free speaking devices in mobile and public telephones or new services using speech recognition make this function mandatory.
In many applications, the speech signal is corrupted by unknown noise. Different noise reduction algorithms [Hayk] [Mart] [Pud] are known using spectral subtraction or spectral shaping. For all these methods, a transformation from time into frequency domain is necessary.
The human ear transforms the time signal into a frequency dependent location according to psycho-acoustic rules [ZWI] . The distribution of the critical band rates and the corresponding time response for the perception of a frequency is non-linear because the frequency resolution of the human ear decreases with increasing frequency.

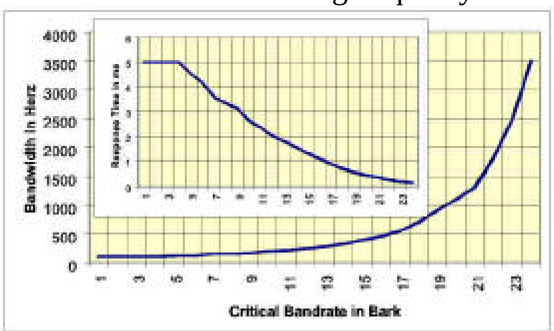

Fig.1 Bandwidth and time response of the critical bands

Hence, the bandwidth between the frequency bins increases with the critical band rate, resulting in a frequency dependent response time (Fig.1).

In this paper a new algorithm realising a Continuous Fourier Transformation (CFT) will be presented. This new method allows to distribute the frequency bins almost freely in accordance with the natural response time.
A reasonable computational complexity can be achieved by subsampling the lower frequencies. With a frequency resolution from 10 to 3000Hz, very comfortable response times between some ms and less than 1ms can be achieved.

The application of the CFT for a noise reduction algorithm, based on the Wiener filter shows strongly reduced musical tones. Compared with the FFT a higher noise reduction can be achieved.

Due to the high degree of freedom in determining system dimensions like number of frequencies, or the function of the frequency distribution, the CFT can be used for many other signal processing algorithms, e. g. echo cancellation in the frequency domain, speech coding, speech synthesis, hearing aids, e. t. c.

State of the art

Signal processing algorithms for noise reduction and echo cancellation use different methods for the transformation from time into frequency domain. The Fast Fourier Transformation (FFT) is the most efficient algorithm which uses a block wise computation with some weak properties such as an equidistant frequency spacing, a very poor response time and the windowed computation with overlap add (Fig. 2).

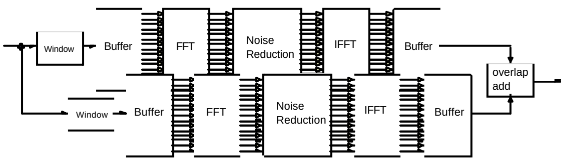

Fig. 2: Blockwise computation with overlap add

The FFT is a block wise computed frequency analysis according to (1)

$$X(n) = \frac{1}{K}\sum_{k=0}^{K-1} x(k)\cdot e^{\frac{-j2pnk}{K}} \qquad (1)$$

The number of samples in the time domain K must be equal to the number of frequencies N in the frequency domain. This leads to an equidistant distribution of the frequency bins with a constant bandwidth and response time for each frequency bin. The response time for a 256 point FFT (at 8kHz sampling rate) is 16ms, nearly ten times higher than the average response time of the human ear with 1,9ms. In this example the frequency resolution is 32Hz, not enough for the lower frequency domain below 300Hz.

With the FFT no adaptation to a non-linear frequency distribution is feasible. The long delay time of the FFT blocks impairs the communication.

The IFFT uses phase and magnitude of a frequency bin from the whole block length. In our example, each 16 ms a fixed and new value is available, leading to audible switched musical tones.

Non-linear Distribution of the frequencies

A simple approach for a non-linear frequency distribution could use a band pass filter solution based on 24 FIR filters for the realisation of the BARK transformation. About 300 coefficients corresponding to the strong masking behaviour of the human ear are necessary for each filter, resulting in a poor response time of 18,75ms and a high computational power of >40 MIPS at 8kHz sampling rate.

New algorithms deal with improved band pass filter techniques or modified discrete Fourier Transformation in order to achieve a better approximation to the properties of the human ear with more efficiency [Drei].

The modification of a discrete Fourier transformation to a non-linear distribution of the frequency bins requires a very high computational power. As $N \neq K$ requires N different summations, each frequency has to be computed with its own number of samples (2).

$$X(n) = \frac{1}{K(n)}\sum_{k=0}^{K(n)} x(k)\cdot e^{\frac{-j2pnk}{K(n)}} \qquad (2)$$

This approach has been developed by [Kap] using the advantages of the Goertzel algorithm. Combined with the polyphase filter technique, this method is used to reconstruct the human ear perception for quality judgements of audio signals and provides 24 critical bands.

For noise reduction, the number of frequencies to be subtracted is determined by the number of distortions within a critical band. For speech improvement, the required frequency resolution depends on the complexity of the distortion. Assuming the signal complexity is limited and the number of frequency bins generated by noise is not more than 3 in each critical
band, the necessary number of frequencies

Continuous Fourier Transformation

With a Continuous Fourier transformation no overlap add and windowing is necessary, leading to a very low complexity for a noise reduction system.

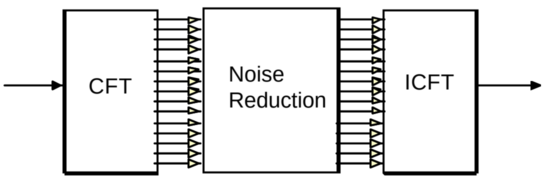

Fig.3 CFT based Noise reduction.

The total release from the block wise computation can be achieved, if the summation is substituted by a low pass filter (LP(n)) of certain order (3).

$$X(n) = LP(n)\left(x(k) \cdot e^{-j2pnk/K}\right) \qquad (3)$$

This filter is computed continuously with the sampling rate and not block wise. Therefore, the frequency bins are able to follow exactly the time behaviour of the time signal. In the easiest case, frequency bins can be computed with a first order recursive filter according to (4).

$$\begin{aligned} X(n,k) = a(n) \cdot \left(x(k) \cdot e^{-j2pnk/K}\right) \\ + b(n) \cdot X(n,k-1) \qquad (4) \end{aligned}$$

With k/K the phase of a discrete frequency is indicated, where K determines the maximum resolution.

within the whole audible frequency range is about 24*(3+1) =96 frequency bins or within a telecommunication channel (<4kHz) 17*(3+1) = 68 frequency bins. The computational effort for a block wise computed non linear distributed frequency analysis is immense (>60MIPS at 8kHz) considering the required frequency resolution.

K may also be interpreted as the time interval of any continuous generated frequency. Tests with a separate sine wave generation showed no limitations.

A very easy determination of the frequency distribution function is feasible with this algorithm. The order and type of this filter can freely be determined. An economically reconstruction of a BARK transformation requires an elliptic recursive filter more than 8^{th} order. However tests with this design showed, that for high quality noise reduction, a higher frequency resolution is needed as above mentioned. Tests with logarithmic distributed frequency bins (up to 75 frequency bins at 8kHz sampling rate) showed strong improvements.

The bandwidth of each frequency bin depends on the distance to the neighbouring frequency and can be determined according to (5).

$$fg(n) = \frac{B(n)}{2} = \frac{X(n+1) - X(n-1)}{4} \qquad (5)$$

The reconstruction of the time signal is realised by the ICFT according to (6).

$$x(k) = \sum_{n=0}^{N-1} X(n) \cdot e^{j2pnk/K} \qquad (6)$$

The computational power for the analysis and synthesis with the CFT is about 22N,

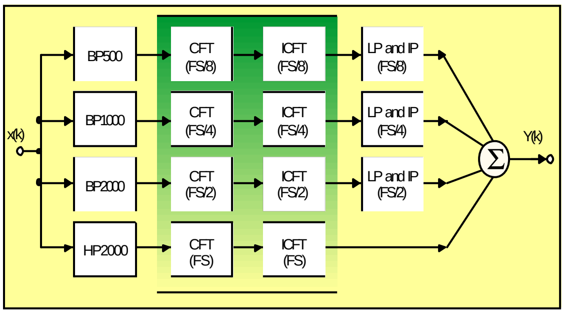

Fig. 4 Sub sampling technique with CFT

whereas this effort can strongly be reduced using subsampling techniques.
As shown in Fig. 4, the time signal has been split into 4 subbands. With a logarithmic scaled frequency distribution the effort for 75 frequency bins is now only 22N/3,75+70, corresponding to a computational power of 4,1MIPS at 8kHz sampling rate.

Conclusion

With the CFT a new transformation method with most natural properties has been found. Fig. 5 shows the comparison results between FFT/IFFT and CFT/ICFT applied with the Wiener filter for spectral subtraction.

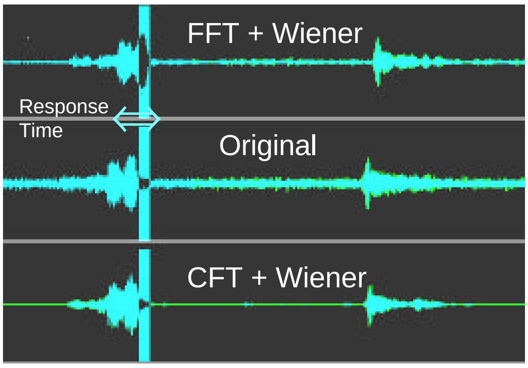

Fig. 5 Comparison FFT-CFT Noise reduction

As with the FFT strong musical tones are generated, the noise reduction is limited. The FFT introduces an undesired delay, whereas the response time of the CFT seems nearly not perceptible.

The essential features of the new CFT are:

- No additional time delay
- Fast time response
- No musical tones
- No overlap and add
- Free transfer function (MEL, BARK Log....)
- Optimal adaptation to desired spectrum feasible
- Adequate computational effort

References

[Drei] P. Dreiseitel, H. Puder: "Speech Enhancement for Mobile Telephony Based on Non-uniformly spaced Frequency Resolution", Proceedings EUSIPCO-98, Isle of Rhodes, Greece, September 1998

[E&M] Y. Ephraim and D. Malah, Speech Enhancement Using a Minimum Mean-Sqare Error Log. - Spectral Amplitude Estimator“IEEE Transactions on Acoustics, Speech and Signal Processing , Vol. ASSP – 33, No.6, April 1985

[Hayk] S. Haykin: „Adaptive Filter Theory“, 3rd Edition, Prentice-Hall, Upper Saddle River, New Jersey, 1996

[Kap] Rolf Kapust„Qualitätsbeurteilung codierter Audiosignale mittels einer BARK – Transformation“,
Dissertation, Technische Fakultät der Universität Erlangen –1993

[Mart] R. Martin„Spectral Subtraction Based on Minimum Statistics“,
Proc. EUSIPCO-94, Edinburgh, pp 1182-1185, 1994

[Pud] H. Puder: "Single Channel Noise Reduction Using Time-Frequency Dependent Voice Activity Detection", International Workshop on Acoustic Echo and Noise Control, Pocono Manor, PA, USA, 1999.

[Zwi] E. Zwicker„Psychoakustik“
Springerverlag 1982, ISBN 3-540-11401-7