Australian Journal of Basic and Applied Sciences, 5(9): 1599-1605, 2011
ISSN 1991-8178

A Review on Audio Sampling Rates Mismatch and Its Effect on the Acoustic Echo Cancellation of Persenol Computers

Adel Nadhem Naeem and Sureswaran Ramadass

[1]National Advanced IPV6 Center , Universiti Sains Malaysia , 11800, Penang, Malaysia.

Abstract: Acoustic echo cancellation (AEC) is a very important technique for hands-free telephone systems and is considered a complicated process in real-time implementation. Previously, it has only been possible through custom very-large-scale integration (VLSI) processors or digital signal processors (DSP). These processors are specially designed for AEC. However, the computation power of personal computers (PC) has become more powerful; real-time signal processing and AEC in the PC environment has now become possible. Sampling rate offset often occurs between the loudspeaker output (D/A converter) and the microphone input (A/D converter) on PCs using commercial audio hardware. The acoustic echo cancellation algorithm may fail because of sampling rate offset. Sampling rate conversion (SRC) corrects the occurrence of offset.

Key words: Acoustic echo cancellation (AEC), Personal Computers (PC), sampling rate conversion (SRC).

INTRODUCTION

PCs need a digital sound card, microphone, speakers, and an operating system to run acoustic echo cancellation (AEC) using VoIP, conference, or any hands-free telephone system. However, these elements may not always be acceptable. Problems sometimes occur when using AEC on a PC. One of these problems may be the sound I/O device on the PC, in which the input and output may have different clock sources (Robjohns, 2003; Carôt and Werner, 2009), there by each having different sample rate. Such deficiencies may cause echo path changes in two ways (Frunze, 2003; Ding and Havelock, 2011). First, AEC will be affected when playing CD-quality music or any other sound file where the playback sampling rate of the sound file is usually higher than the capture rate (Stokes and Malvar, 2004), Second, the different sampling rates of the D/A converter and the A/D converter of cheap PC audio hardware may increase or decrease delay, thereby causing lost or repeated samples (Pawig *et al.,* 2010).

Acoustic Echo Cancellation (AEC):

Echo is a known problem for telephone communication systems (Storn, 1996). Two kinds of echo can be encountered in telephone communication systems. One is electrical echo caused by the hybrids of a Public Switched Telephone Network (PSTN) exchange, where the subscriber's two-wire lines are connected to four-wire lines. The other is acoustic echo, which usually occurs in hands-free telephone communication systems because of the coupling between the loudspeaker and microphone (Raghavendran, 2003). A general explanation of acoustic echo problem is shown in Figure 1, where there signal (x) enters, is transmitted from the far end by the speakers, and is then bounced back and received by the microphone at the near end side.

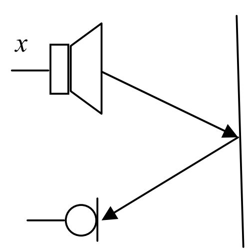

Fig. 1: Illustration of Acoustic Echo Occurrence.

Several methods can be used to eliminate or reduce the acoustic echo (Adrian, 2004):

1. Headsets are the powerful and simplest tools to avoid acoustic echo; however, the increasing use of hands-free gadgets that requires speakers with a separate microphone has made this solution useless (Storn, 1996).
2. System amplification reduces overall amplification, thereby causing the feedback to fade. This solution can result in the low volume quality (Adrian, 2004).

Corresponding Author: Adel Nadhem Naeem, National Advance IPV6 Center, University Sains Malaysia,11800, Penang, Malaysia.
E-mail: adel@nav6.usm.my

3. Echo Suppression can only be used with a half-duplex system, which means only one side can talk at a time. The echo suppressor works by detecting signals on one side and shutting down the microphone on the other side. The speaker signal on the non-active side does not travel back to the active side; thus, it cannot be used with full-duplex telephone systems (Storn, 1996; Adrian, 2004).
4. The concept of Acoustic Echo Cancellation (AEC) is shown in Figure 2. The signal (x) from the far end is transmitted from the speaker. After being bounced back by different surfaces, the echo signal (d) is captured by the microphone, along with the near end signal (s) captured with the noise signal (n). Therefore, to remove the echo signal (d) from the near end signal (s), an adaptive filter that models the room acoustics should be used. The filter should determine the acoustics from a given microphone and speaker signals, after which the filter will calculate an estimated microphone signal from the speaker signal. This estimated microphone signal is then subtracted from the real microphone signal, and the resulting signal no longer contains the speaker signal (acoustic echo) (Storn, 1996; Adrian, 2004).

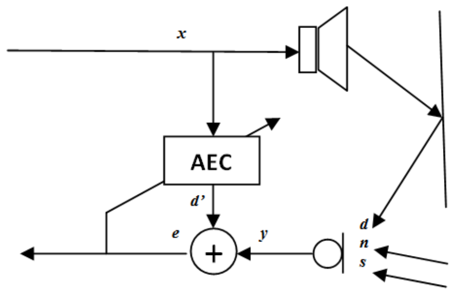

Fig. 2: General concept of AEC work.

Adaptive Filter Used In AEC:

The main types of digital filtering used are finite impulse response (FIR) and infinite impulse response (IIR). IIR and FIR are able to achieve the same performance, but with different coefficients and computation. The filters usually used in AEC are FIR because of instability problem of IIR (Lu, 2007). The main work in the adaptive filter of the AEC is estimating the echo path of the room to obtain a similar signal to the echo signal; the echo path estimating an adaptive update is needed to adapt the environmental change, similar to the movement of people. Convergence speed is crucial in obtaining the best echo path estimate. Several kinds of adaptive algorithms are used in acoustic echo cancellation, including least mean squares (LMS), recursive least squares (RLS), and affine projection algorithm (APA) (Lu, 2007).

Sample Rate Conversion (SRC)

SRC operation takes one audio signal with specific sample rate and changes it to another sample rate. Sampling is converting the continuous-time signal $x(t)$ to discrete-time signal $x[k]$ by taking repeated measurements as a fixed interval to obtain specific time. The interval is called T_s. The sampling rate of $F_s = 1/T_s$, shown in Figure 3. (Franz, 2001; Kappeler and Grünert, 2004; LaValley, 2004)

$$x[k] = x(k.T_i) \quad (1)$$

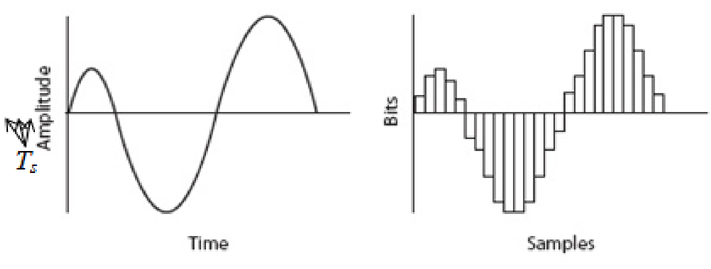

Fig. 3: Conversion continuous-time signal $x(t)$ to discrete-time signal $x[k]$*(Franz, 2001)*.

Another issue of sampling audio signal is bit depth, which defines how many bits are used to describe each sample. When the bit depth becomes higher, the sound quality is better, and the most common bit depths are 8-bit and 16-bit (See Figure 4) (LaValley, 2004).

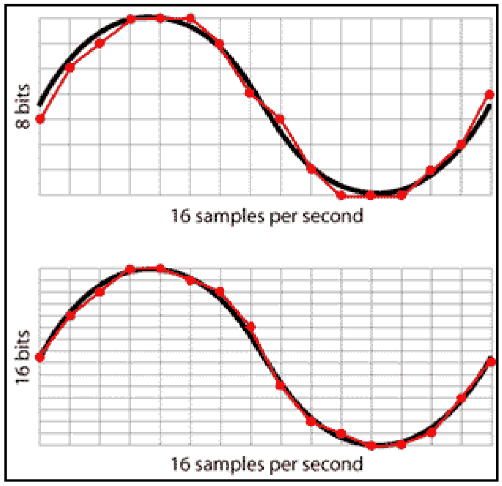

Fig. 4: Illustration of the bit depth effect on the signal (LaValley, 2004).

SRC is used manage the spectrum of the signal. This can be calculated using discrete Fourier transform by limiting the frequency spectrum of the sampled signal to half the sampling rate (Nyquist frequency, which is $F_n = F_s/2$) (Redmon, 2007; Parker and ScienceDirect, 2010).

$$X(e^{i2\pi F_s}) = a_0 + \sum_{k \in Z}^{\infty} x[k] e^{i2\pi F_s} \tag{2}$$

SRC can be used in two ways: for rational factors or for arbitrary ratios. SRC for rational factors can be performed by upsampling (interpolation by factor *L*), downsampling (decimation by factor *M*) or resampling by factor *L/M* (Rothacher, 1995; Kappeler and Grünert, 2004; Wang, 2008):

- Decimation decreases the samples in audio signal. As mentioned, the frequency domain representation of the signal has to be considered, as well as the time domain. Figure 5 shows the old sample rate and the new sample rate after decimation by 2 (decreasing the samples by 2) (Parker and ScienceDirect, 2010).

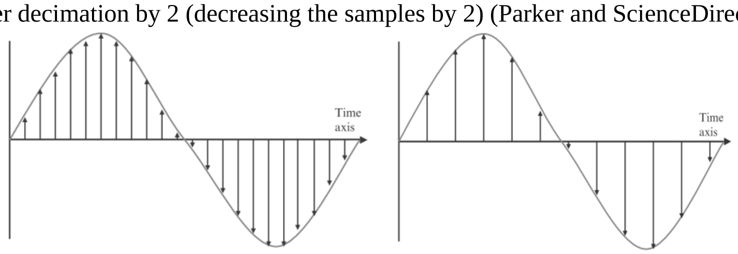

Fig. 5: Decimation digital signal (Parker and ScienceDirect, 2010).

The sample is taken from audio signal with sample rate Fs and converted it into the new signal by removing half the signal samples. The signal itself has not changed; only the sampling rate frequency and corresponding Nyquist rate frequency have changed. If the new Nyquist frequency is larger than the signal frequency, no aliasing will occur.

- Unlike decimation, interpolation increases the samples in audio signal. Figure 6 shows an example of an old sample rate and the new sample rate after interpolation by 2 (increasing the samples by 2) (Parker and ScienceDirect, 2010).

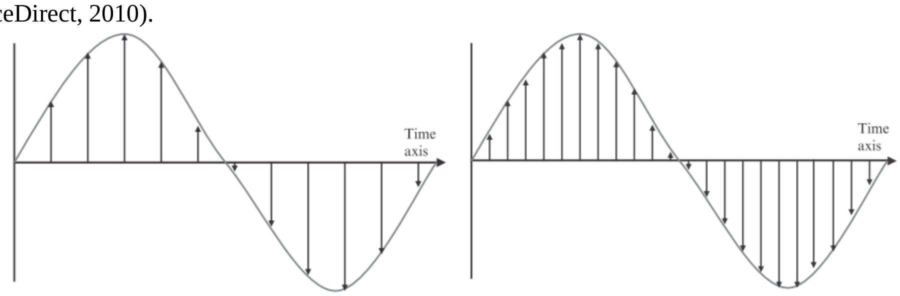

Fig. 6: Interpolation digital signal (Parker and ScienceDirect, 2010).

The sample is taken from audio signal with sample rate F_s and converted into the new signal by doubling the samples in the signal. The signal itself has not changed; only the sampling rate frequency and corresponding Nyquist rate frequency have changed. As long as the new Nyquist frequency is larger than the signal frequency, no aliasing will occur.

- Resampling uses the fractional value of ratio by combining decimation and interpolation to change the sampling rate. For example, to interpolate by a factor of 3 and decimate by a factor of 2, the resampling factor used is 3/2=1.5.

SRC for Arbitrary Ratio:

When the sample rates of audio signal have an arbitrary ratio, SRC for rational factors (interpolation, decimation, or resampling) cannot be used (Evangelista, 2003). Therefore, the audio signal must be changed from a discrete-time signal to a continuous-time signal by interpolating the signal with a large and fixed interpolation factor *L*. Various interpolation methods are possible, such as hold interpolation, linear interpolation, and interpolation with B-splines (Rothacher, 1995; Kappeler and Grünert, 2004).

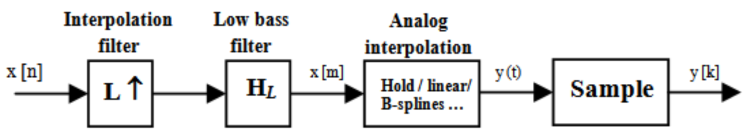

Fig. 7: Illustration of SRC for arbitrary ratio.

The discrete signal *x[n]* shown in Figure 7 enters the interpolation filter to increase the samples with factor *L*. It then shifts to the low-bass filter to prevent aliasing. The new discrete signal *x[m]* rises to go through the analog interpolation (hold, linear, polynomial, or B-spline), and is converted to a continuous signal *y(t)*. Taking this signal as sample, the final discrete signal *y[k]* can be obtained.

Audio Sampling Rates Mismatch And Its Effect on the AEC:

(Stokes and Malvar, 2004) assumed that playing CD-quality music and Internet gaming with voice chat would affect AEC because playback sampling rate will usually be higher than the capture rate. He proposed a solution for subband AEC (Bai *et al.,* 2009) in which the sampling rate conversion uses frequency-domain interpolation to match the transform lengths of the playback and to capture signals (Figure 8).

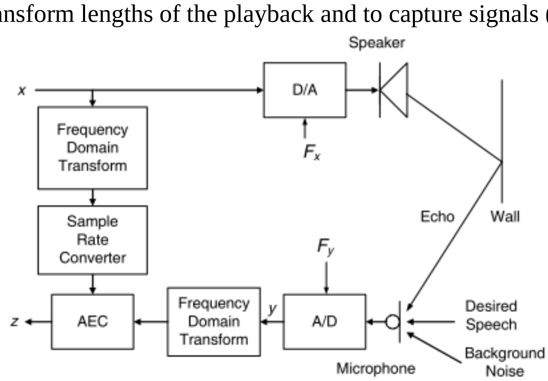

Fig. 8: Solution proposed by Stokes.

The solution proposed by Stokes uses three types of sampling rate conversion (SRC): time domain SRC, exact frequency-domain SRC, and interpolated frequency-domain SRC.

- Time-domain SRC is the simplest way using linear interpolation. It has two disadvantages: (1) it leads to aliasing levels that produce audible distortion and (2) it needs to use a very long poly phase FIR filter to handle the multi rate filtering.
- In exact frequency-domain SRC, the sampling rate conversion occurs after the frequency-domain transform for the signal *x* in a standard frequency domain-sampling. When the transformed length matches the number of points in a frame, the SRC discards the frequency domain coefficients for the bands over the capture sampling rate (i.e., when the playback sampling rate is higher than the capture sampling rate). Moreover, SRC includes zero padding to the frequency domain bands of the transformed playback signal up to the length of the transformed capture signal (i.e., when the playback sampling rate is lower than the capture sampling rate). The main disadvantage of this method is that it requires difficult factorable lengths of FFT, thereby leading to complex and significantly less efficient implementations.

- Interpolated frequency-domain SRC combines a frequency-domain transform with length power of 2 and an SRC using a simple frequency-domain interpolation. This method compares the exact frequency domain and slightly degrades the quality of the AEC algorithm; however, it is more efficient for real-time implementation.

Stokes evaluated the performance of the suggested AEC architecture using frequency-domain interpolation by measuring the numerical accuracy and CPU consumption. He used a capture sampling rate of 16 kHz with typical wideband conferencing, and playback sampling rate of 44.1 kHz with typical system audio or CD-quality music playback. He considered only mono playback because he assumed that stereo playback leads to additional issues. Likewise, the capture signal used was mono.

(Robledo-Arnuncio *et al.,* 2007) studied how can rate mismatches affects acoustic echo cancellation. He explained the occurrence of several problems when using VOIP applications with PC where a sample rates vary among the devices:

- Clock signals generation may have a certain tolerance in their nominal frequency.
- Temperature change can affect the operating frequency of these devices.
- In a portable device, the clock signals used for the audio hardware are often obtained by applying different division factors to a higher frequency clock; therefore, the user may not find the same expected nominal frequencies.

Furthermore, according to Enrique, sampling rate mismatch problems of basic VOIP applications can be solved by simple mechanisms for the resulting gaps or skips, such as during audio silences, or by adding/removing a few samples at a time. However, these solutions are inapplicable for more complex applications, such AEC (Lienhart *et al.,* 2003). He therefore suggested correcting such mismatches in sampling rate as follows:

- Accurately estimate the sampling rate mismatch.
- Process the streams to correct the mismatch.

Enrique used synthetic recordings to perform AEC experiments. In one experiment, three loudspeakers were placed on a conference table with two microphones. The set of room responses was recorded in a controlled room configured to emulate a remote collaboration site, with a reverberation time of 300 ms. Speech signals were used as sources. The speech segments and the room responses were down sampled to 8 kHz.

(Pawig *et al.,* 2010) focused on VOIP over personal computers, particularly when using commercial audio hardware, as sampling frequency offset between the loudspeaker output D/A converter and the microphone input A/D converter often occurs. With sampling frequency offset, echo cancellation algorithms failed to track the correct room impulse response. He suggested an LMS-type adaptive algorithm to estimate the frequency offset and resynchronize the signals using arbitrary sampling rate conversion.

According Pawig, the different sampling rates of the D/A converter and the A/D converter of low-quality PC audio hardware may cause increasing or decreasing delay, which would cause lost or repeated samples, thereby affecting echo and deteriorating its estimation.

Figure 9 illustrates the effect of sampling rate offset in echo path, in the experiment of Pawig using two kinds of offset (i.e., Δf = 0 Hz, Δf = 6 Hz), and fixing the other coefficients (i.e., echo path length M = 300) using NLMS algorithm with step-size $\alpha\ (k)$=0.5 and filter length N=300, The adaptive filter coefficients are shown as $t0=0.1s$, $t1=3s$, and $t2=5s$. Clearly, the offset 6 Hz changed and shifted the peak, affecting the estimation of echo path.

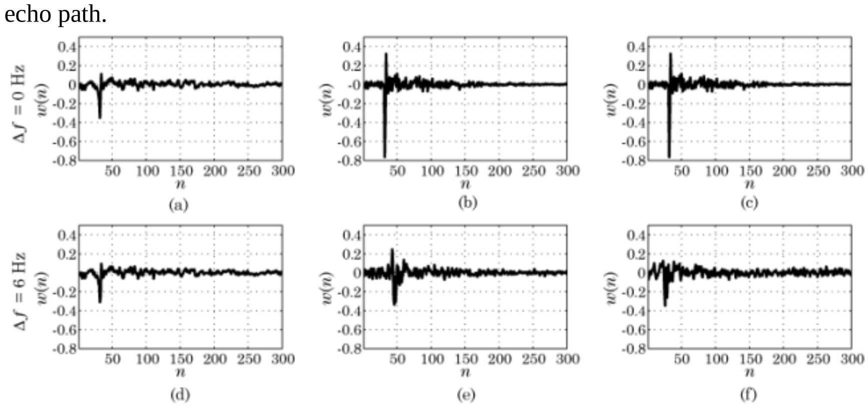

Fig. 9: The effect of sampling rate offset in echo path.

The new design proposed by Pawig is able estimate offset of sampling rate and resample the rates depending on the estimated offset (Figure 10):

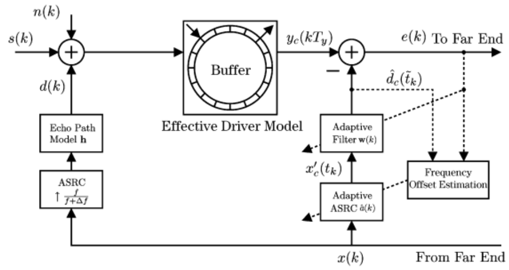

Fig. 10: Design proposed by Pawig for solving sample rate problem with AEC.

Pawig estimated the offset by comparing the near end signal, $y_c(kT_y)$ and the far end filtered version, $\hat{d}_c(t_k)$, and then minimizing the error e(k), where subscript c represents continuous-time versions of the signals T_x and T_y in the sampling periods. However, Pawig used arbitrary sampling rate conversion (ASRC) to correct frequency offset. The ASRC principle is shown in Figure 11:

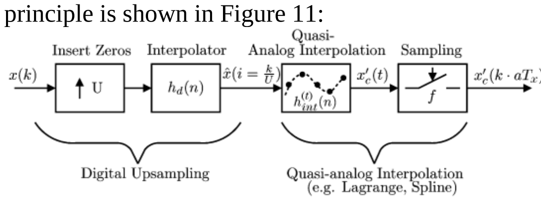

Fig. 11: ASRC in Pawig's design.

Pawig obtained his result by simulating and analyzing the structure shown in Figure 10 using Matlab. The far end input signal *x(k)* was recorded speech by male and female speakers taken from a German audio book. The room impulse responses h were different and generated randomly by:

$$h(i) = \begin{cases} 0.01r(i) & i < \Delta_h \\ r(i)(e^{-1} + 0.1)e^{-\frac{i}{0.15M}} & i \geq \Delta_h \end{cases} \tag{3}$$

Where $r(i)$ is a white noise random variable of Gaussian distribution and variance σ_r^2, and $0<i<M$, M is the desired filter length, and Δ_h is a direct sound delay. The effect of the sampling rate offset Δf was modeled by ASRC.

Table 1: Comparison between works have been done on Sample rate mismatch problem with AEC

Year	Problem	Author	Suggested solution	AEC algorithm	SRC type	Simulator	Motivations
2004	playing audio with voice chat will affect AEC	Stokes	Interpolated Frequency-Domain SRC	subband AEC	Rational SRC	Matlab	Reduce the code complexity and memory requirement
2007	rate mismatches effects AEC	Enrique	Using interpolation to simulate small sample rate mismatches, and analyze how these affect the AEC	NLMS	Rational SRC	Matlab	Quantify the requirement of a rate mismatch correction for AEC
2010	sampling frequency offset between the loudspeaker and the microphone in commercial audio hardware	Pawig	least mean square adaptive algorithm to estimate the frequency offset and resynchronize the signals using arbitrary SRC	LMS	Arbitrary SRC	Matlab	correct a frequency offset that may occurred because of commercial audio hardware

The time-scale jump effects of the audio driver were implemented by a model to check the current sampling rate differences and the relative position of the read/write pointers on a buffer. Whenever the pointers meet, a buffer length was repeated or removed from the signal to simulate the additional time-jump effect.

Conclusion:

Table 1 summary and shows the differences between that have been done to solve the sample rate mismatch problem with AEC. Sample rate mismatch occurs by playing sound during VoIP communication or because of the commercial sound cards. Solving this problem is done by use SRC to resample the audio signals and make them match to use in AEC efficiently.

ACKNOWLEDGMENTS

This research is supported by National Advanced IPv6 Centre (NAV6) UNIVERSITI SAINS MALAYSIA (USM).

REFERENCES

Adrian, A., 2004. Voice over internet acoustic echo cancellation.

Bai, M.R., C.K. Yang and K.N. Hur, 2009. Design and implementation of a hybrid sub-band acoustic echo canceller (aec). Journal of Sound and Vibration, 321(3-5): 1069-1089.

Carôt, A. and C. Werner, 2009. External latency-optimized soundcard synchronization for applications in wide-area networks.

Ding, H. and D.I. Havelock, 2011. Drift-compensated adaptive filtering for improving speech intelligibility in cases with asynchronous inputs. EURASIP Journal on Advances in Signal Processing, 2010: 95. DOI 10.1155/2010/621064.

Evangelista, G., 2003. Design of digital systems for arbitrary sampling rate conversion* 1. Signal processing, 83(2): 377-387.

Franz, D., 2001. Producing in the home studio with pro tools. Berklee Pr Pubns.

Frunze, A., 2003. Echo cancellation demystified. http://www.embeddedstar.com

Kappeler, R. and D. Grünert, 2004. Sample rate converter.

LaValley, D., 2004. Digital audio properties. http://streaming.wisconsin.edu

Lienhart, R., I. Kozintsev, S. Wehr and M. Yeung, 2003. On the importance of exact synchronization for distributed audio signal processing, IEEE: pp: IV-840-843 vol. 844.

Lu, L., 2007. Implementation of acoustic echo cancellation for pc applications using matlab. Stoskholm.

Parker, M. and Science Direct, 2010. Digital signal processing: Everything you need to know to get started. Newnes/Elsevier.

Pawig, M., G. Enzner and P. Vary, 2010. Adaptive sampling rate correction for acoustic echo control in voice-over-ip. Signal Processing, IEEE Transactions on, 58(1): 189-199.

Raghavendran, S., 2003. Implementation of an acoustic echo canceller using matlab. University of South Florida.

Redmon, N., 2007. Sampling in-depth. http://www.earlevel.com

Robjohns, H., 2003. Digital clocking explained.

Robledo-Arnuncio, E., T.S. Wada and B.H. Juang, 2007. On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation. IEEE: pp: 34-37.

Rothacher, F.M., 1995. Sample rate conversion: Algorithm and vlsi implementation. SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH.

Stokes, J.W. and H.S. Malvar, 2004. Acoustic echo cancellation with arbitrary playback sampling rate. IEEE: pp: 153-156: 154.

Storn, R., 1996. Echo cancellation techniques for multimedia applications-a survey. INTERNATIONAL COMPUTER SCIENCE INSTITUTE-PUBLICATIONS-TR.

Wang, F., 2008. A self-tuning flexible sample rate converter using a linear interpolator on the ti dsk 6713 board. In: Information and Electrical Engineering. University of Applied Sciences Hamburg.