CLOCK DRIFT ESTIMATION AND COMPENSATION FOR ASYNCHRONOUS IMPULSE RESPONSE MEASUREMENTS

Hannes Gamper

Microsoft Research

ABSTRACT

The impulse response (IR) of an acoustic environment or audio device can be measured by recording its response to a known test signal. Ideally, the same digital clock should be used for playback and recording to ensure synchronous digital-to-analog and analog-to-digital conversion. When measuring the acoustic performance of a hardware device, be it for audio input to a device microphone or audio output from a device speaker, it is often difficult to access the device's audio signal path electronically. Therefore, the device-under-test (DUT) has to act either as a playback or recording device for the IR measurement. However, it may be impossible to synchronise the internal clock of the DUT with the reference clock of the measurement system. As a result, the recorded DUT response may be subject to unknown clock drift which may lead to undesired artefacts in the measured IR. Here, a method is proposed for estimating the drift between a playback and recording clock directly from the recorded response to obtain a drift-compensated IR. Experimental results from IR measurements of a DUT subject to clock drift indicate that the proposed method successfully estimates the drift rate and yields an accurate IR estimate in magnitude and phase.

Index Terms— Sampling rate, clock synchronization, asynchronous IR measurement

1. INTRODUCTION

The behaviour of a linear and time-invariant system can be conveniently described by its transfer function in the frequency domain or its impulse response (IR) in the time domain. A measured IR is useful for analysing and simulating the acoustic properties of a room or a hardware device [1, 2]. While an IR can theoretically be measured by recording the response to a Dirac delta function, in practice this approach may suffer from poor signal-to-noise ratio and reproducibility [3], in particular if an acoustic IR is measured using an impulsive excitation source, for example a pistol shot or balloon pop [4]. A common approach to improve the signal-to-noise ratio of an IR measurement is to use an excitation signal that stretches the energy delivered to the system over time. Examples of such signals include maximum-length sequences (MLS) [5], time-stretched pulses (TSP) [6], and swept sines [7]. After recording the response of a device-under-test (DUT) to the excitation signal, the IR is obtained by applying an appropriate inverse filter. As an example, when using an exponential sine sweep as the test signal, the IR can be obtained by convolving the recorded response with an amplitude-modulated time-reversed version of the excitation signal [8]. This works under the assumption that the playback clock and the recording clock are either one and the same or synchronised. However, clock synchronisation may be difficult or impossible to guarantee if the audio path leading from a DUT microphone (when measuring the DUT input IR) or to a DUT speaker (when measuring the DUT output IR), can not be accessed electronically, and the test signal has to either be played back by the DUT or recorded on it. The lack of clock synchronisation may cause the external reference clock and the DUT clock to drift relative to one another which may result in artefacts in the estimated IR.

Svensson and Nielsen present an overview of the effects of clock drift and other sources of time variance on MLS measurements [9]. Farina describes the effect of a mismatch between playback and DUT clock in a measurement using an exponential sine sweep as resulting in a "skewed" IR [8]. Bryan et al. derived closed-form expressions describing the impact of clock drift on IR measurements [10]. They showed that without clock synchronisation, even professional-grade audio devices may exhibit clock drift which in turn potentially corrupts IR measurements. The authors propose two approaches for estimating impulse responses in the presence of clock drift. The first approach relies on estimating the clock drift rate between the playback and the recording device from an electronic loopback recording of a sequence of impulses or chirps. Given an estimate of the drift rate, it can be compensated for either via resampling or by applying a drift correction filter. The second approach compensates for the clock drift implicitly, by using an electronic loopback recording of the inverse filter used to derive the impulse response from the raw measurement. Both approaches are shown to yield highly accurate results. However, in practical situations, recording an electronic loopback may not be possible if the DUT does not offer electronic audio inputs or outputs.

Building on the work by Bryan et al., a method is proposed for estimating the clock drift between a reference clock and the DUT directly from the recorded response, without the need for an electronic loopback. It is shown that the effect of clock drift is exacerbated when using time averaging to improve the signal-to-noise ratio, but that repeated measurement runs can be used to successfully estimate the clock drift rate and derive a drift-compensated IR estimate. Blind clock drift estimation from multiple asynchronously recorded responses has been shown previously in the context of distributed acoustic sensor arrays [11, 12, 13]. However, rather than evaluating multiple asynchronous recordings, here drift estimation is performed on a single DUT recording in the context of IR measurements, by comparing recorded repetitions of the excitation signal.

2. ASYNCHRONOUS IR MEASUREMENTS

To illustrate the effect of clock drift, the example of sine sweep measurements is discussed. However, the proposed method is not limited to sweep-based IR measurements.

2.1. Excitation signal and inverse filtering

Various test signals for IR measurements have been studied, including impulse trains and pseudo-random noise sequences. For acoustic measurements, exponential sine sweeps are frequently used [8, 10]. When measuring an IR in the presence of acoustic background noise, exponential sine sweeps may provide a better signal-to-noise ratio due to increased signal energy at low frequencies, in line with the typical noise distribution in rooms [10]. An exponential sine sweep has a frequency trajectory that increases exponentially from a start frequency, ω_1, at time $t = 0$ to a stop frequency, ω_2, at time $t = \mathrm{T}$, given as [14]:

$$s(t) = \sin\left(K(e^{\frac{t}{L}} - 1)\right), \tag{1}$$

where

$$K = \frac{\omega_1 \mathrm{T}}{\log(\frac{\omega_2}{\omega_1})} \tag{2}$$

and

$$L = \frac{\mathrm{T}}{\log(\frac{\omega_2}{\omega_1})}. \tag{3}$$

The response of a DUT with unknown linear and time-invariant system behaviour to the sweep signal is given by

$$y(t) = s(t) * h(t) + n(t), \tag{4}$$

where $h(t)$ is the unknown IR and $n(t)$ denotes additive noise. An estimate of the IR can be obtained as

$$\hat{h}(t) = y(t) * \bar{s}(t), \tag{5}$$

where $\bar{s}(t)$ is an appropriate inverse filter. A simple method of obtaining the inverse filter, $\bar{s}(t)$, consists in reversing the test signal, $s(t)$, in time and applying an appropriate amplitude modulation [8]. When generating the test sequence digitally, zeros can be appended to allow the system response to decay after the last output sample of the test sequence. To improve the signal-to-noise ratio of the estimated IR, averaging over multiple measurement runs can be performed, though care has to be taken when averaging over the responses of time varying systems [8]. The simplest way to perform averaging is to repeat the digital test sequence with appended zeros, $s_0[n], n = [0 \cdots \mathrm{N}-1]$, R times, split the recorded signal into R non-overlapping frames of length N and average over those frames.

2.2. Effect of clock drift on IR estimation

Bryan et al. show that the effect of clock drift on the IR obtained from a single sweep is approximately equivalent to filtering the IR with an allpass filter exhibiting the same frequency trajectory as the test signal [10]. However, when averaging over repeated runs, as outlined in Section 2.1, clock drift causes temporal misalignment between successive frames (see Figure 1b). Assuming the DUT clock has drifted relative to the reference clock by D_{run} samples after the first run, the r-th measurement run can be approximated as a delay applied to the first measurement run, $\tilde{y}_0$:

$$\tilde{y}_r(z) \approx \tilde{y}_0(z) z^{-rD_{\mathrm{run}}}, \tag{6}$$

Averaging $\tilde{y}_0$ and $\tilde{y}_r$ results in

$$\frac{1}{2}\left(\tilde{y}_0(z) + \tilde{y}_r(z)\right) \approx \frac{1}{2}\left(\tilde{y}_0(z)(1 + z^{-rD_{\mathrm{run}}})\right), \tag{7}$$

i.e., a feed-forward comb filter [15]. Averaging over several runs is equivalent to applying several feed-forward delay lines in parallel with a combined transfer function

$$C(\omega) \approx 1 + \sum_{r=1}^{R-1} e^{-j\omega r D_{\mathrm{run}}}. \tag{8}$$

It is evident from (8) that even minor clock drift can severely distort an IR measured using sine sweeps when applying time-averaging over multiple runs. An example of the comb filtering effect is shown in Figure 1e.

3. PROPOSED METHOD

3.1. Clock drift estimation

In their experiments, Bryan et al. found that clock drift between two audio devices remained relatively constant when comparing two measurements taken five hours apart, but would vary from one day to another [10]. The clock drift estimation method proposed here assumes that the clock drift between the DUT and an external reference clock is constant for the duration of one entire measurement, which typically takes up to one minute.

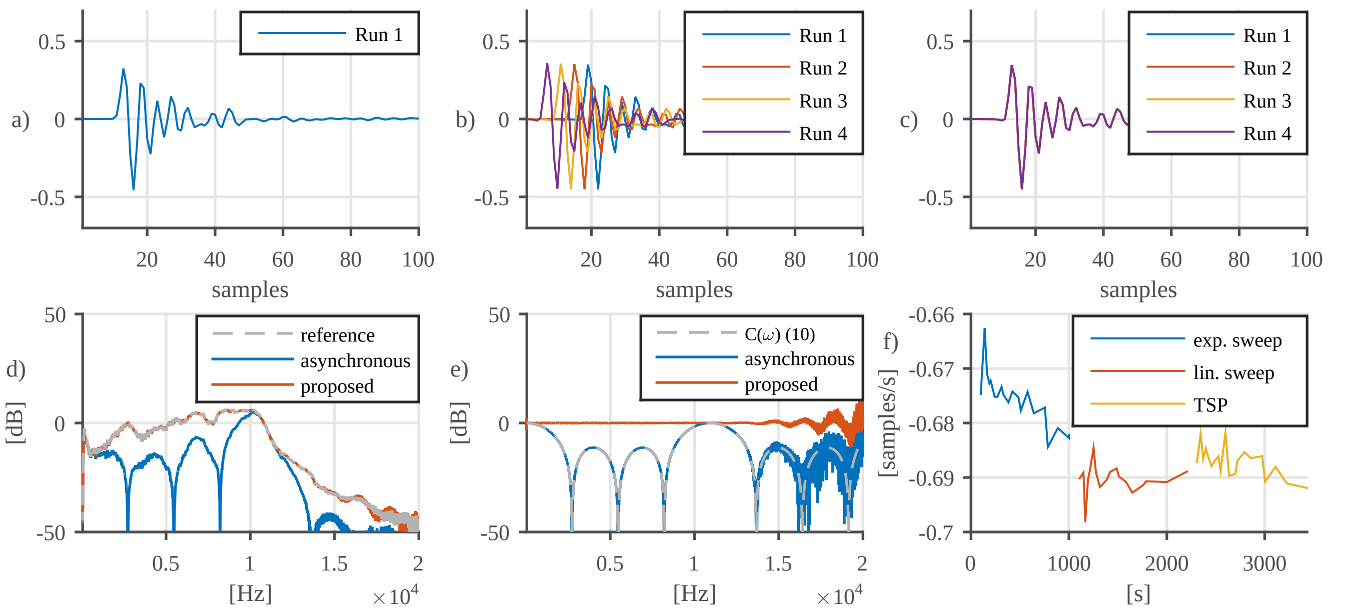

Fig. 1. Temporal and magnitude response of ground-truth reference IR (a, d) and asynchronous IR measurement with four repetitions without clock drift compensation (b, d) and with proposed drift compensation (c, d); e) shows the magnitude response error of the asynchronous IR measurement; f) shows estimated drift rates using various test signals over time.

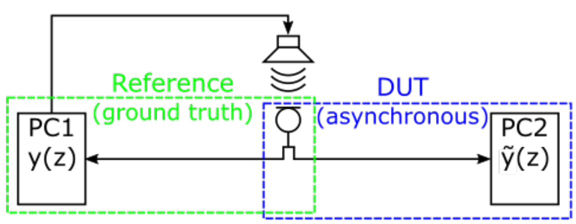

Fig. 2. Experimental setup consisting of clock-synchronous reference system (PC1) and asynchronous DUT (PC2).

The scenario considered here consists of a reference playback clock and a drifting DUT recording clock, as illustrated in Figure 2. Given (6) and assuming a constant clock drift D_{run}, the first and the r-th measurement run are aligned in time by applying an appropriate compensation delay

$$\tilde{y}_0(z) \approx \tilde{y}_r(z) z^{r D_{\text{run}}}. \tag{9}$$

To minimise measurement noise, the individual measurement runs can be convolved with the inverse filter $\bar{s}$, yielding IR estimates subject to clock drift, $\hat{\tilde{h}}_r$. The clock drift compensation in (9) can be applied in the frequency domain via

$$\hat{\tilde{H}}_0(\omega) \approx \hat{\tilde{H}}_r(\omega) e^{j\omega r D_{\text{run}}}. \tag{10}$$

For an overview of various fractional delay implementations and their characteristics see Välimäki and Laakso [16]. The clock drift is estimated via a least-squares minimisation:

$$\hat{D}_{\text{run}} = \underset{\hat{D}_{\text{run}}}{\arg\min} \sum_{r=1}^{R-1} \int_{\omega_1}^{\omega_2} \left| \left(\hat{\tilde{H}}_0(\omega) - \hat{\tilde{H}}_r(\omega) e^{j\omega r \hat{D}_{\text{run}}} \right)^2 \right| d\omega. \tag{11}$$

Note that to evaluate (11), only the recorded DUT response is required, filtered with the excitation signal to reduce noise.

3.2. Clock drift compensation

Given the drift estimate $\hat{D}_{\text{run}}$, the raw recording is resampled to compensate for the clock drift:

$$\hat{y}(t) = \tilde{y}(\hat{\alpha} t), \tag{12}$$

where $\hat{\alpha}$ is the estimated drift rate, given as

$$\hat{\alpha} = \frac{\mathrm{N} + \hat{D}_{\text{run}}}{\mathrm{N}}. \tag{13}$$

Here, the resampling was implemented via windowed sinc interpolation [15]. After resampling, the recording is split into individual measurement runs, $\hat{y}_r(t)$, of length N. The drift-compensated IR estimate is obtained by averaging over the individual measurement runs:

$$\hat{h}(t) = \frac{1}{R} \sum_{r=0}^{R-1} \hat{y}_r(t) * \bar{s}(t). \tag{14}$$

4. EXPERIMENTAL EVALUATION

Experiments were carried out in a quiet office space. The experimental setup, illustrated in Figure 2, consisted of a Fostex 6301B personal monitor and a Beyerdynamic TG-X58 microphone pointing at the centre of the speaker at a distance of about 20 cm. The loudspeaker was driven by a personal computer acting as the reference clock (PC1 in Figure 2). The microphone signal path was split and recorded simultaneously on the reference system to obtain a ground-truth clock-synchronous IR measurement, and a second personal computer acting as the asynchronous DUT (PC2 in Figure 2). The clocks of the two computers were not synchronised.

The reference IR was measured using a single exponential sine sweep from 20 Hz to 20 kHz of length 2^{18} samples, i.e., about six seconds at a sampling rate of 44.1 kHz.

Figure 1a and 1d show the temporal and magnitude response, respectively, of the reference IR. Figure 1b illustrates the temporal misalignment of four repetitions of an exponential sweep with 2^{17} samples. Averaging over these repetitions causes a distortion of the estimated magnitude response, as shown in Figure 1d. The error relative to the reference response, $|Y(f)| - |\tilde{Y}(f)|$, is plotted in Figure 1e. It can be seen that the error caused by averaging over asynchronous measurement runs without compensation matches the comb filter model, $C(\omega)$, given in (8) remarkably well. This indicates that the approximation in (6) is reasonable. After drift estimation and compensation, the individual runs of the asynchronous measurement shown in Figure 1b are correctly aligned in time, as can be seen in Figure 1c. Figure 1d and 1e show that the resulting magnitude response matches the reference response quite well up to about 15 kHz, which corresponds to the rated upper frequency limit of the microphone.

Figure 1f illustrates the clock drift estimated using exponential and linear sine sweeps and TSPs of various lengths and with 2–16 repetitions. The range of drifts estimated over the course of one hour is about ± 0.02 samples per second.

Farina states that due to the possibility of time variance, a single, long exponential sweep is preferable over an average estimate of several shorter sweeps [8]. Figure 3a compares the reference IR to the IR estimated in the presence of clock drift using four repetitions of a short sweep (2^{16} samples, i.e., about 1.5 seconds each) after applying the proposed clock drift compensation as well as a single, much longer exponential sweep (2^{21} samples, i.e., about 48 seconds) without compensation. It can be seen that the clock drift distorts the temporal response of the single asynchronous sweep, but that the proposed method successfully restores the response of the repeated shorter sweeps. Figure 3b and 3c show the magnitude and phase of the equivalent drift response, given as

$$H_{\mathrm{D}}(\omega) = \frac{\hat{\tilde{H}}(\omega)}{H(\omega)}, \tag{15}$$

where $\hat{\tilde{H}}(\omega)$ and $H(\omega)$ denote the estimated and reference IR response, respectively. It can be seen that the drift filters exhibit allpass behaviour, with an essentially flat magnitude response up to about 15 kHz. This is in line with the findings of Bryan et al. [10]. The drift filter equivalent to the proposed method has a phase response that is approximately linear, but not zero, indicating that the proposed method introduces a slight shift in time to the measured IR. This is not surprising, given that the DUT recordings were not synchronised and hence may be shifted by a fraction of a sample relative to the reference recording. However, without clock drift compensation, the single, long exponential sweep measurement introduces a nonlinear phase shift, shown in Figure 3. This may be problematic when using the measured IR for applications including beamforming or acoustic simulation [17].

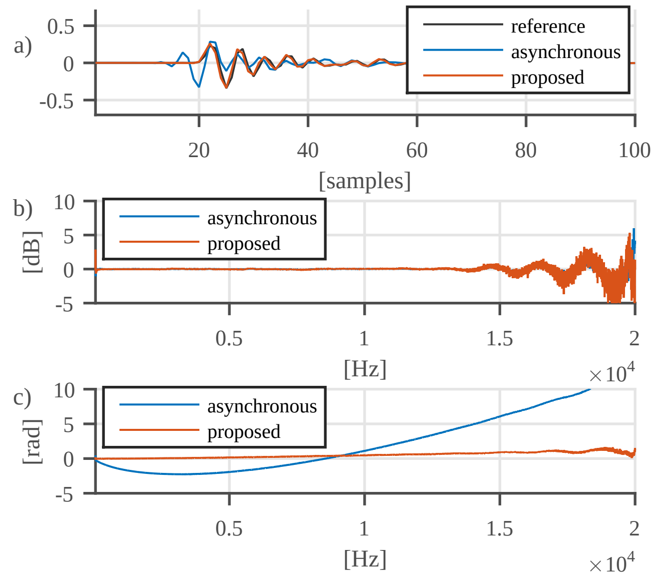

Fig. 3. a) Asynchronous IR estimated with a single, long exponential sweep, and with four repetitions of a short sweep after the proposed clock drift compensation; b) magnitude and c) phase response of the equivalent drift filters, given in (11).

5. CONCLUSION

Clock drift may distort asynchronous IR measurements. Furthermore, when using sine sweeps or TSPs as test signals, temporal averaging may result in comb filtering effects distorting the measured response. The proposed method is shown to reliably estimate clock drift directly from the recorded DUT response and compensate the measured IR, the only requirement being that the excitation signal be played and recorded at least twice. The proposed approach may yield more accurate results than using a single, long sweep measurement without compensation. Initial experiments indicate that the proposed method works with MLS excitation signals too. However, a formal investigation is left for future work.

6. REFERENCES

[1] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. 3148–3161, 2012.

[2] H. Gamper, M. R. P. Thomas, L. Corbin, and I. J. Tashev, "Synthesis of device-independent noise corpora for realistic ASR evaluation," in *Proc. Interspeech*, San Francisco, CA, USA, 2016.

[3] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc*, vol. 31, no. 7/8, pp. 478–488, 1983.

[4] J. Pätynen, B. F. Katz, and T. Lokki, "Investigations on the balloon as an impulse source," *J. Acoust. Soc. Am.*, vol. 129, no. 1, pp. EL27–EL33, 2011.

[5] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, 1989.

[6] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119–1123, 1995.

[7] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. 108th Conv. Audio Eng. Soc.*, Paris, France, 2000.

[8] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Conv. Audio Eng. Soc.*, Vienna, Austria, 2007.

[9] P. Svensson and J. L. Nielsen, "Errors in MLS measurements caused by time variance in acoustic systems," *J. Audio Eng. Soc.*, vol. 47, no. 11, pp. 907–927, 1999.

[10] N. J. Bryan, M. A. Kolar, and J. S. Abel, "Impulse response measurements in the presence of clock drift," in *Proc. 129th Conv. Audio Eng. Soc.*, San Francisco, CA, USA, 2010.

[11] S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *Proc. Int. Workshop on Acoustic Signal Enhancement*, Aachen, Germany, 2012, pp. 1–4.

[12] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Signal Processing*, vol. 107, pp. 185 – 196, 2015.

[13] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 571–582, 2016.

[14] Q. Meng, D. Sen, S. Wang, and L. Hayes, "Impulse response measurement with sine sweeps and amplitude modulation schemes," in *2nd Int. Conf. Signal Process. and Commun. Syst.*, Gold Coast, Australia, 2008.

[15] J. O. Smith, *Physical Audio Signal Processing*, http://ccrma.stanford.edu/ jos/pasp/, accessed Dec 2016, online book, 2010 edition.

[16] V. Välimäki and T. I. Laakso, "Principles of fractional delay filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000, pp. 3870–3873.

[17] H. Gamper, L. Corbin, D. Johnston, and I. J. Tashev, "Synthesis of device-independent noise corpora for speech quality assessment," in *Proc. Int. Workshop on Acoustic Signal Enhancement*, Xi'an, China, 2016, pp. 1–5.