Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Sound and Vibration 294 (2006) 981–1003
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Leakage-free identification of FRF's with the discrete time Fourier transform

J. Antoni*

Laboratoire Roberval de Mécanique, UMR CNRS 6066, Centre de Recherche de Royallieu, BP 20529, 60205 Compiègne Cédex, France

Received 17 June 2004; received in revised form 21 November 2005; accepted 19 December 2005
Available online 9 March 2006

Abstract

The discrete time Fourier transform (DTFT) has for long been used for non-parametric measurement of frequency response functions (FRF). However, beyond the appreciable simplicity and computational efficiency of the method, its use has been severely criticised when applied to stationary random signals because of the inherent spectral leakage it induces—e.g. in the so-called H_1, H_2, H_3, H_v, etc. estimators. This problem is one major reason which has motivated recent research on alternative identification methods, always at the price of increased complexity. This paper aims at demonstrating that, contrary to the common belief, solutions do exist for designing low-bias, or even unbiased, i.e. leakage free, FRF estimators based on the DTFT of stationary random signals. One such solution was proposed 25 years ago by Rabiner and has surprisingly remained unknown by mechanical engineers concerned with system identification. We first give a brief review of the principles which underlie Rabiner's method. We next present original analytical results, which exactly specify under which conditions Rabiner's estimator is unbiased, and we also provide the expression of its variance. These results then lead to a number of practical guidelines. The actual validity of the method is finally illustrated on some examples, and compared with other more classical estimators.

1. Introduction

Getting accurate frequency response function (FRF) measurements from experimental data is a persistent topic of research due to its prime importance in many physical fields. A vast literature currently exists on the subject, which tackles the problem either in the time-domain or in the frequency-domain, and either from a parametric or a non-parametric approach. This paper is concerned with the frequency-domain non-parametric approach which has gained considerable popularity among engineers because of its simplicity, its availability on most commercial data analysers, and its direct interpretation in physical terms. A thorough discussion of this approach can be found in classical references such as Refs. [1–4]. More recently, the non-parametric frequency-domain approach has regained new interest as a pre-processing step before more sophisticated parametric techniques are used. For a modern treatment of this topic, see Ref. [5].

In brief, the problem at hand is the following: given some input measurement $\{x[n]\}_{n=0}^{N-1}$ and output measurement $\{y[n]\}_{n=0}^{N-1}$ to a linear and time-invariant system, find the best possible measurement of the system

*Tel.: +33 3 4423 4525; fax: +33 3 4436 7314.
E-mail address: jerome.antoni@utc.fr.

0022-460X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2005.12.037

impulse response (IR) $h[n]$, or equivalently in the frequency-domain, of the FRF $H(f) = \mathscr{F}\{h[n]\} = T_s\sum_{n\in\mathbb{Z}} h[n]\mathrm{e}^{-\mathrm{j}2\pi fnT_S}$, with T_s denoting the sampling period and $\mathscr{F}$ the discrete time Fourier transform (DTFT). The problem usually faces extra complexity as the input and output measurements are corrupted by extraneous noise and undesirable disturbances. Depending on the nature of the input excitation to the system, several estimators of $H(f)$ have been developed which are optimal according to some statistical criteria. The classical approach as described in Refs. [3,4] is to consider a stationary random excitation and then to derive the optimal FRF as the function which minimises the mean square of the error $y[n] - T_s\sum_k h[k]x[n-k]$ for some assumed length L of the IR. This yields the well-known $H_1(f) = S_{yx}(f)/S_{2x}(f)$ and $H_2(f) = S_{2y}(f)/S_{yx}(f)$ estimators when only output noise or only input noise is present, respectively, with $S_{yx}(f)$ the cross-spectrum between signals y and x, and $S_{2y}(f)$ and $S_{2x}(f)$ the auto-spectra of signals y and x. When both input and output noises are present, dedicated estimators also exist such as $H_3(f)$ (instrumental variable), $H_v(f)$ (maximum-likelihood), $H_\alpha(f)$ (periodically correlated input) [6], which all involve more elaborate functions of $S_{yx}(f)$, $S_{2y}(f)$, and $S_{2x}(f)$. In this respect, the non-parametric frequency-domain estimation of FRF's boils down to a spectral analysis issue, i.e. the issue of estimating the spectral density $S_{vu}(f)$ for any two signals v and u from finite-length measurement $\{v[n]\}_{n=0}^{N-1}$ and $\{u[n]\}_{n=0}^{N-1}$. This is a classical but ill-posed problem which has received considerable attention in the past [2,3,7]. It is in particular a standard result that any (non-parametric) estimate of $S_{vu}(f)$ suffers from a difficult trade-off between systematic errors (bias) and stochastic errors (variance). As will be recalled in the next section, the systematic errors result from the use of the DTFT on finite-length data which produces significant "boundary effects" known as leakage, but also from voluntary (and necessary!) smoothing of the spectral estimates in order to reduce the stochastic errors. As a consequence, this problematic compromise finds its exact replica with FRF's measurements. One solution to alleviate the aforementioned difficulty is to force the use of specific excitation signals which can produce unbiased spectral estimates. This is the case for random bursts since the DTFT of transient signals does not suffer from leakage [1], and also for periodic signals as long as the DTFT is evaluated on an integer number of periods. This latter solution has recently gained a lot of interest [5]. However, neither random bursts nor periodic signals may be systematically available in practice and there are many instances when one has no other solution than adopting the classical assumption of stationary random excitations.

Contrary to the common belief, (nearly) unbiased spectral estimates, and hence FRF measurements, can be devised even with finite-length measurements of stationary random signals. Surprisingly, this fact has rarely been recognised despite the existing literature on the subject [8,9]. It seems to have remained completely unknown from the engineering community concerned with vibrations and modal analysis, where yet it would be of obvious importance. It is the object of this paper to partially fill in this gap. The idea we are reporting herein is originally due to Rabiner and dates back to the late 1970s [8,10]. More than simply digging out old material, we give to Rabiner's method a simpler and more comprehensive formulation which provides original interpretations and potential directions for future research. We also demonstrate that Rabiner's method is computationally efficient and we address its statistical performance. The paper is organised as follows. In Section 2, we first introduce Rabiner's method from an empirical point of view as a correction to the classical Welch's method. In Section 3, we then prove rigorously under which conditions it yields an unbiased estimator and we derive a closed-form expression for its variance. Finally, we provide in Section 4 some practical guidelines so as to optimise the parameters involved in Rabiner's method, and in Section 5 we illustrate its use on actual data.

2. The principle of Rabiner's method

2.1. Bias induced by classical spectral estimates

There are two customary spectral estimators used in practice. The first one, historically the most ancient, is known as "Blackman and Tuckey" (B&T) or "lag-window" spectral estimator and is defined as the DTFT[1] of

[1]For the sake of simplicity, we will use the normalised frequency $f/T_s \to f$ from here onwards, unless specifically stated otherwise. Practically, this amounts to setting $T_s = 1$.

the weighted empirical cross-correlation $\widehat{R}_{vu}[k]$ between signals v and u, viz [3]:

$$\widehat{S}_{vu}^{(B)}(f) = \sum_{k=-N+1}^{N-1} g[k] \cdot \widehat{R}_{vu}[k] \mathrm{e}^{-\mathrm{j}2\pi fk}, \tag{1}$$

where

$$\widehat{R}_{vu}[k] = \frac{1}{N-|k|} \sum_{n=\max(0,-k)}^{\min(N-1,N-1-k)} v[n+k]u[n]^*, \quad |k|<N \tag{2}$$

and $\{g[k]\}_{k=-N+1}^{N-1}$ is a suitably chosen lag-window ($g[k] \geqslant 0$, $g[-k] = g[k]$, $g[0] = 1$). The expression of B&T's estimator follows closely from the theoretical definition of the cross-spectral density as given by the Wiener-Khintchine theorem [3], but unfortunately it is not computationally efficient when N is large. The second customary spectral estimator is due to Welch [11] and is commonly referred to in the literature as the "weighted overlapped segment averaging" (WOSA) method. The Welch's estimator is defined as

$$\widehat{S}_{vu}^{(W)}(f) = \frac{R_{2w}(0)^{-1}}{I} \cdot \sum_{i=0}^{I-1} V_{N_w}^{(i)}(f) U_{N_w}^{(i)}(f)^*, \tag{3}$$

where

$$V_{N_w}^{(i)}(f) = \mathscr{F}\{w_i[n].v[n]\} = \sum_{n=i\Delta}^{i\Delta+N_w-1} w_i[n].v[n] \mathrm{e}^{-\mathrm{j}2\pi fn} \tag{4}$$

is the so-called "short-time DTFT" of $v[n]$, and similarly for $U_{N_w}^{(i)}(f)$. In Eq. (4), $w_i[n] = w[n - i\Delta]$ is the shifted version of a N_w-long data-window $\{w[n]\}_{n=0}^{N_w-1}$ that selects a segment of the measured signals $\{v[n]\}_{n=0}^{N-1}$ anchored on time $i\Delta$. The increment Δ is set between 1 and N_w so as to allow possible overlap between adjacent segments. For N-long signals, there are $I = \lfloor (N - N_w)/\Delta \rfloor + 1$ (where $\lfloor x \rfloor$ stands for the greatest integer smaller than or equal to x) segments of data averaged together in Eq. (3) such as to decrease the variance of Welch's estimator by approximately a factor I. Note finally that Welch's estimator must be normalised by the energy $R_{2w}(0) = \sum_n w[n]^2$ of the data-window so that it has the correct calibration.

Welch's estimator has a number of advantages over B&T's estimator and therefore has become a standard in modern data analysers and signal processing toolboxes. As a matter of fact, the corresponding algorithm:

- can be efficiently implemented by means of several short FFT's of fixed size,
- does not require excessive memory allocation and therefore can handle very long signals,
- can be implemented in a recursive real-time architecture which virtually does not impose any limit to the number of averages I,
- is robust against outlying and non-stationary data since contaminated segment can easily be disregarded from the summation.

The B&T's and Welch's estimators both have their bias[2] expressed as [7]:

$$b\left\{\widehat{S}_{vu}(f)\right\} \stackrel{\text{def}}{=} \mathbb{E}\left\{\widehat{S}_{vu}(f)\right\} - S_{vu}(f) \simeq \tfrac{1}{2} \cdot B^2 \cdot S''_{vu}(f), \tag{5}$$

where $S''_{vu}(f)$ is the second derivative of the theoretical spectral density $S_{vu}(f)$ and $B^2 = \int_{-1/2}^{+1/2} f^2 \Phi(f) \mathrm{d}f$ is a measure of the frequency resolution or "spectral bandwidth". For instance, $\Phi(f) = \mathscr{F}\{g[k]\}$ in B&T's method, whereas $\Phi(f) = |\mathscr{F}\{w[n]\}|^2$ in Welch's method. Therefore, using B&T's or Welch's spectral estimators in the

[2]For expression (5) to hold, it is implicitly assumed that the IR length is shorter than the time concentration of the lag-window g in B&T's estimator, and than the window length N_w in Welch's estimator.

estimation of a FRF yields the normalised bias (H_1 formulation):

$$\frac{b\left\{\widehat{H}_1(f)\right\}}{H(f)} \simeq \frac{b\left\{\widehat{S}_{yx}(f)\right\}}{S_{yx}(f)} - \frac{b\left\{\widehat{S}_{2x}(f)\right\}}{S_{2x}(f)}. \tag{6}$$

Although this expression involves the difference of two terms, it will generally not be zero. For example let us consider the case where the excitation x has a relatively flat spectrum; then the following approximation holds:

$$\frac{b\{\widehat{H}_1(f)\}}{H(f)} \simeq \tfrac{1}{2} \cdot B^2 \cdot H''(f), \tag{7}$$

which clearly evidences that large relative bias is produced in the vicinity of the resonances of the FRF. The only option to reduce the bias in expression (6) is to decrease the bandwidth B. In Welch's method, this is achieved by enlarging the window length N_w and thus diminishing the number of averages $I = \lfloor (N - N_w)/\Delta \rfloor + 1$. As well known this strategy results in a higher variability of the spectral estimates $\widehat{S}_{vu}(f)$ and consequently of $\widehat{H}(f)$. The original idea of Rabiner's method is to achieve a significant bias reduction in Welch's method *without* changing the window length N_w and thus without reducing the number of averages I.

2.2. A strategy to reduce the bias in Welch's method

Welch's spectral estimator implicitly assumes that the measured signals can be decomposed into a series of *independent* short-time segments, such that subsequent spectral analysis can be performed individually on each of these segments and the results averaged all together. This is illustrated in Fig. 1 in the case of a rectangular data-window with zero overlap ($\Delta = N_w$). That this strategy yields unbiased results makes sense in the case of white noise and is theoretically verified from Eq. (6) in which $S''_{vu} = 0$. The extension to coloured signals, however, does not hold true because in this case adjacent segments are *not mutually independent* and therefore cannot be analysed separately. Instead, a proper exploitation of the former idea must actually rely on two precautions:

(1) the correlation between adjacent segments must be taken into account in the spectral analysis,
(2) the summation of the spectral estimates obtained on adjacent and possibly overlapping segments ($\Delta \leqslant N_w$) must be equivalent (in the mean square sense) to the spectral estimate of the whole signal.

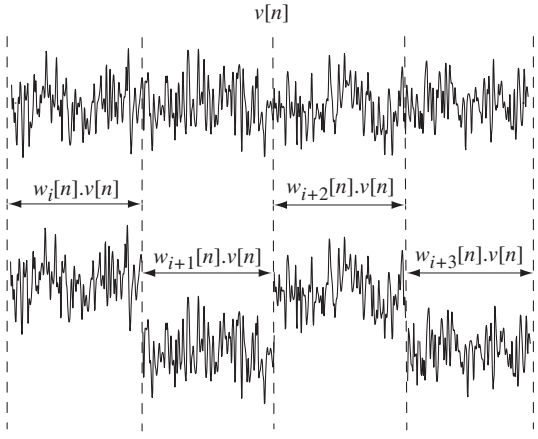

Fig. 1. Welch's method divides the signal into (possibly overlapping) segments of equal lengths.

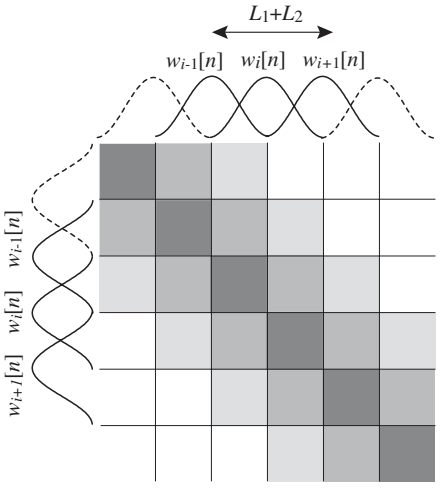

Fig. 2. Correlation patterns between adjacent windowed segments. Gray and white boxes indicate correlated and uncorrelated segments, respectively.

Condition 1 suggests the generalised form

$$\widehat{S}_{vu}(f) = \frac{\theta}{I}\sum_{i=0}^{I-1}\sum_{j=0}^{I-1} V_{N_w}^{(i)}(f)U_{N_w}^{(j)}(f)^* \tag{8}$$

instead of Eq. (3) (where the scaling factor θ is to be determined), whereas condition 2 places a restriction on the choice of the data-window $w[n]$ such that Eq. (8) is a valid estimate of $S_{vu}(f)$. Intuitively, this must the case if:[3]

$$\sum_{i=-\infty}^{\infty} w_i[n] = C, \tag{9}$$

where C is a constant, since then $\sum_{i=1}^{I} w_i[n]v[n] = C \cdot v[n]$ and similarly for $u[n]$ so that Eq. (8) simply becomes

$$\widehat{S}_{vu}(f) = C^2\frac{\theta}{I}\cdot V_N(f)U_N(f)^* \tag{10}$$

where $V_N(f) = \mathscr{F}\{v[n]\} = \sum_{n=0}^{N-1} v[n]\mathrm{e}^{-\mathrm{j}2\pi fn}$ is the DTFT of the whole signal and similarly for $U_N(f)$. Within a scaling factor, this is nothing else than the classical cross-periodogram, a quantity which despite having a significantly reduced bias ($N \gg N_w$) is unfortunately an unsuitable spectral estimator because it is not consistent (i.e. its variance does not decrease with the number N of available data) [4]. The principle of Rabiner's method is to preserve the small bias implied by Eq. (10) (and even to cancel it) while still decreasing its variance at the same time; this is achieved by taking advantage that most of the terms in the summation (8) can be disregarded because they are theoretically zero. More specifically, all products $V_{N_w}^{(i)}(f)U_{N_w}^{(j)}(f)^*, i \neq j$, involving the DTFT of non-overlapping windows $w_i[n]$ and $w_j[n]$ more distant than the cross-correlation length between v and u are zero, as illustrated in Fig. 2 (in Fig. 2 the cross-correlation length of v and u is denoted by $L_1 + L_2$ for reasons to become clear later). Removing these terms from the summation (8) then preserves the expected value of $\widehat{S}_{vu}^{(R)}(f)$ while decreasing its stochastic variability. This produces the refined estimator:

$$\widehat{S}_{vu}^{(R)}(f) = \frac{\theta}{I - q_2 - q_1}\sum_{i=q_1}^{I-q_2-1}\sum_{q=-q_1}^{q_2} V_{N_w}^{(i+q)}(f)U_{N_w}^{(i)}(f)^*, \tag{11}$$

[3]Condition (9) will be formally proved in Section 3.1.

where $0 \leqslant q_1, q_2 < I$ are set as small as possible so that *only* statistically correlated windows $w_{i+q}[n]$ and $w_i[n]$ are kept in the summation.[4] Eq. (11) defines Rabiner's spectral estimator as originally introduced in Ref. [8], although we have arrived to it herein from a more intuitive approach. It now remains to prove rigorously our intuition that the so-defined $\widehat{S}_{vu}^{(R)}(f)$ is unbiased and statistically consistent.

3. Performance analysis of Rabiner's estimator

Rabiner's original approach only focused on finding an unbiased estimator $\widehat{H}_1(f) = \widehat{S}_{yx}(f)/\widehat{S}_x(f)$ of the FRF $H(f)$. This indirectly led him to proposing Eq. (11) as a spectral estimator for $S_{yx}(f)$ and $S_{2x}(f)$, which statistical properties he actually did not study. The unbiasedness and consistency of Eq. (11) were later demonstrated in Ref. [12]. In this section, we provide a simpler and yet more comprehensive approach to analysing the bias and the variance of Rabiner's spectral estimator. Into addition, our formulation provides enlightening graphical interpretations.

3.1. Bias analysis

Let $D_M(x) = \sin(M\pi x)/\sin(\pi x)$ be the (non-normalised) Dirichlet kernel. Then the following result holds true (see proof in Appendix A):

Lemma 1. *The expected value of Rabiner's spectral estimator* (11) *is given by*

$$\mathbb{E}\left\{\widehat{S}_{vu}^{(R)}(f)\right\} = \int_{-1/2}^{+1/2} S_{vu}(f-\lambda)\Phi(\lambda)\,\mathrm{d}\lambda, \tag{12}$$

with

$$\Phi(\lambda) = \theta \cdot |W(\lambda)|^2 \cdot \mathrm{e}^{-\mathrm{j}\pi\lambda\Delta(q_2-q_1)} \cdot D_{q_2+q_1+1}(\lambda\Delta). \tag{13}$$

In the above, $\Phi(\lambda)$ plays the role of a smoothing spectral window on the true cross-spectral density $S_{vu}(f)$. In order to investigate the effect of $\Phi(\lambda)$, let us rewrite Eq. (13) in the time-lag domain:

$$\mathbb{E}\left\{\widehat{S}_{vu}^{(R)}(f)\right\} = \sum_{k=-N+1}^{N-1} \phi[k] \cdot R_{vu}[k]\mathrm{e}^{-\mathrm{j}2\pi fk}, \tag{14}$$

where

$$\phi[k] = \mathscr{F}^{-1}\{\Phi(\lambda)\} = \theta \cdot R_{2w}[k] * \Pi_{(-q_1\Delta;q_2\Delta)}[k]. \tag{15}$$

Without surprise the inverse DTFT of $\Phi(\lambda)$ is a time-lag window $\phi[k]$ which tappers the cross-correlation function $R_{vu}[k]$ just as $g[k]$ does in B&T's method—see Eq. (1). A closer look at Eq. (15) reveals that the structure of $\phi[k]$ is that of a convolution of the autocorrelation $R_{2w}[k] = \sum_n w[n+k]w[n]$ with the rectangular window $\Pi_{(-q_1\Delta;q_2\Delta)}[k] = \sum_{q=-q_1}^{q_2} \delta[k - q\Delta]$. This is illustrated in Fig. 3, where it is shown that such a structure allows the synthesis of a piecewise flat window. At this stage, it becomes clear that Rabiner's method defines an unbiased spectral estimate provided that $\phi[k] = 1$ for any $k \in [-L_1; L_2]$ with $-L_1$ and L_2 the lower and upper bounds of the interval spanned by $R_{vu}[k]$, since then:

$$\mathbb{E}\left\{\widehat{S}_{vu}^{(R)}(f)\right\} = \sum_{k=-N+1}^{N-1} 1 \cdot R_{vu}[k]\mathrm{e}^{-\mathrm{j}2\pi fk} = S_{vu}(f). \tag{16}$$

[4]Note that product terms $V_{N_w}^{(i)}(f)U_{N_w}^{(i)}(f)^*$ for $i = 0, \ldots, q_1 - 1$ and $i = I - q_2, \ldots, I - 1$ are purposely disregarded from the summation so as to make the estimator *exactly* unbiased instead of nearly unbiased. To see why, consider for instance the term $V_{N_w}^{(0)}(f)U_{N_w}^{(0)}(f)^*$; removing the bias it is inducing would require using segments $-1, -2, \ldots, -q1$ which are not available from the data record.

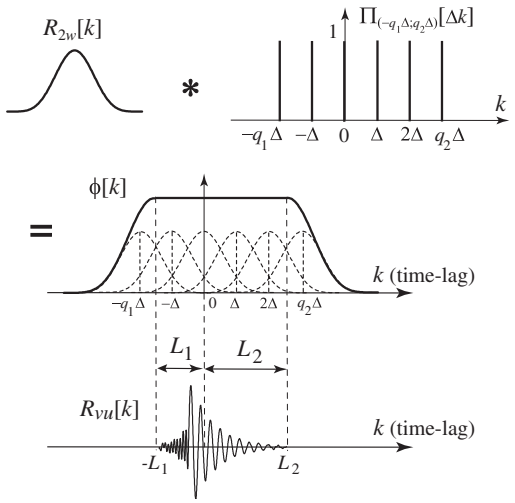

Fig. 3. Structure of the lag-window $\phi[k]$ in Rabiner's estimator, which tapers the cross-correlation function $R_{vu}[k]$. The resulting estimator is unbiased provided that the flat summit of $\phi[k]$ completely covers the support $[-L_1; L2]$ of $R_{vu}[k]$.

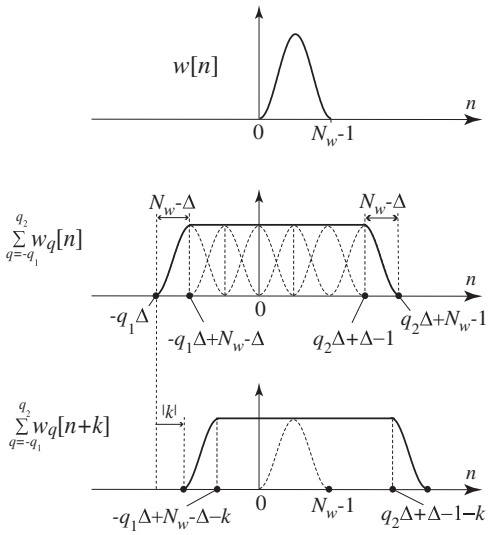

Fig. 4. Condition for the lag-window $\phi[k]$ to have a flat submit: $\sum_{q=-q_1}^{q_2} w_q[n+k]$ must be a constant on an interval large enough so that it completely covers the support $[0; N_w - 1]$ of $w[n]$ for any $-L_1 \leqslant k \leqslant L_2$.

After some trivial manipulations on Eq. (15), the condition that $\phi[k] = 1$ in $[-L_1; L_2]$ becomes:

$$\phi[k] = \theta \cdot \sum_{q=-q_1}^{q_2} R_{2w}[k - q\Delta] = \theta \cdot \sum_n w[n] \sum_{q=-q_1}^{q_2} w_q[n+k] = 1 \quad \forall k \in [-L_1; L_2]. \tag{17}$$

As illustrated in Fig. 4, a sufficient requirement for Eq. (17) to hold true is that $\sum_{q=-q_1}^{q_2} w_q[n+k]$ is a constant on an interval large enough so that it completely covers the support of $w[n]$, i.e. $\forall n \in [0; N_w - 1]$ and $\forall k \in [-L_1; L_2]$, since then $\theta \cdot \sum_{q=-q_1}^{q_2} w_q[n+k] \cdot \sum_n w[n]$ is the product of three constants. This requirement is achieved by imposing that (i) adjacent data windows sum up to a constant:

$$\sum_{i=-\infty}^{\infty} w_i[n] = C \tag{18}$$

and (ii) that:

$$q_i \geqslant \left\lceil \frac{L_i + N_w}{\Delta} \right\rceil - 1, \tag{19}$$

where $\lceil x \rceil$ stands for the smallest integer greater than or equal to x.

Theorem 1. *Let $-L_1$ and L_2 be the lower and upper bounds of the support set of the cross-correlation $R_{vu}[k]$ (i.e. $R_{vu}[k] = 0,\ \forall k \notin [-L_1; L_2]$) and $I = \lfloor (N - Nw)/\Delta \rfloor + 1$ the number of available segments. Then Rabiner's spectral estimator*

$$\widehat{S}_{vu}^{(R)}(f) = \frac{(\sum_{q=-q_1}^{q_2} R_{2w}[q\Delta])^{-1}}{I - q_2 - q_1} \sum_{i=q_1}^{I-q_2-1} \sum_{q=-q_1}^{q_2} V_{N_w}^{(i+q)}(f) U_{N_w}^{(i)}(f)^* \tag{20}$$

is unbiased provided that (i) *the data windows are chosen according to condition* (18) *and* (ii) $q_i \geqslant \lceil (L_i + N_w)/\Delta \rceil - 1$, $i = 1, 2$.

Remarks.

(1) The scaling factor $\theta = (\sum_{q=-q_1}^{q_2} R_{2w}[q\Delta])^{-1}$ is actually equal to $(\sum_n w[n])^{-1}$ as soon as the conditions of Theorem 1 are fulfilled.
(2) The so-defined estimator is a generalisation of Welch's estimator in the sense that Eq. (20) reduces to Eq. (3) when $q_1 = q_2 = 0$. However, Welch's estimator is unbiased only with white noise as previously mentioned. Indeed, in that case $L_1 = L_2 = 0$ which implies with condition (19) that $q_i \geqslant 0$, $i = 1, 2$.
(3) Note that conditions (18) and (19) automatically imply that $\widehat{S}_{2u}^{(R)}(f)$ is an unbiased spectral estimator as well, for the support set of $R_{2u}[k]$ is completely included in that of $R_{vu}[k] = (h * R_{2u})[k]$.

It is noteworthy that, as opposed to the previous references on the subject, Theorem 1 gives a simpler and yet more comprehensive set of conditions for Rabiner's estimator to be unbiased. In particular our formulation (20) does not require any condition on the length N_w of the data-window and N_w is even allowed to be smaller than the cross-correlation length $L_2 + L_1$; in such a case, unbiasedness is simply reached by increasing the value of q_i according to Eq. (19). To our knowledge, this degree of freedom on N_w is unique in non-parametric spectral analysis and has never been recognised before. It is a very useful property which, for example, allows implementing Rabiner's estimator with short fixed-length FFT's while still providing unbiased results; this is a drastic potential improvement of the technique currently used in commercial spectral analysers.

3.2. Variance analysis

Using the same approach as for the bias, the following results holds for the variance[5] of Rabiner's spectral estimator (see proof in Appendix B):

Theorem 2. *The normalised variance of Rabiner's spectral estimator is*:

$$\begin{aligned} \mathrm{Var}\left\{ \widehat{S}_{vu}^{(R)}(f) \right\} &= \frac{(\sum_{q=-q_1}^{q_2} R_{2w}[q\Delta])^{-2}}{(I - q_2 - q_1)^2} \int \int_{-1/2}^{+1/2} S_{2v}(f - \lambda_1) S_{2u}(f - \lambda_2) \cdot \Phi(\lambda_1, \lambda_2) \cdot \mathrm{d}\lambda_1 \, \mathrm{d}\lambda_2 \quad \textit{with } \Phi(\lambda_1, \lambda_2) \\ &= D_{q_2+q_1+1}^2(\lambda_1 \Delta) \cdot D_{I-q_2+q_1}^2[(\lambda_2 - \lambda_1)\Delta] \cdot |W(\lambda_1)|^2 \cdot |W(\lambda_2)|^2, \quad 0 < |f| < \tfrac{1}{2}. \end{aligned} \tag{21}$$

[5]It is implicitly assumed throughout this paper that the number I of segments is large in Welch's estimator; therefore our variance analysis does not account for transient effects due to windowing as described in Ref. [13].

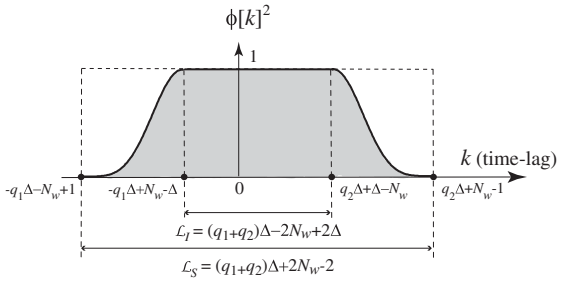

Fig. 5. The variance of Rabiner's spectral estimator is proportional to the energy $\mathscr{E}_\phi$ of the lag-window $\phi[k]$ as illustrated by the shaded area. The latter is lower and upper bounded by the areas of two rectangles of lengths $\mathscr{L}_I$ and $\mathscr{L}_S$, respectively.

This formula seems in accordance with that of Ref. [12]. It proves that Rabiner's estimator is consistent since its variance tends to zero as the number N of available data, and hence of averages I, becomes large; yet it is difficult to interpret it in this form. Relying on the usual assumptions of large samples (I large) and smooth spectral densities, the following more practical formula approximately holds true (see proof in Appendix B):

Corollary 1. *The normalised variance of Rabiner's spectral estimator is asymptotically*:

$$\frac{\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}}{S_{2v}(f)S_{2u}(f)} \simeq \frac{\mathscr{E}_\phi}{\Delta(I-q_2-q_1)} \quad \text{with } \mathscr{E}_\phi=\sum_k \phi[k]^2, \quad 0<|f|<\tfrac{1}{2}. \tag{22}$$

This asymptotic formula better evidences that the variance is proportional to the energy $\mathscr{E}_\phi$ of the lag-window $\phi[k]$. Its interpretation is illustrated in Fig. 5. If required, the energy $\mathscr{E}_\phi$ can be computed from the following expression:

$$\mathscr{E}_\phi=(q_2+q_1+1)\frac{\sum_{q=-q_1-q_2}^{q_1+q_2}(R_{2w}*R_{2w})[q\Delta]}{(\sum_{q=-q_1}^{q_2}R_{2w}[q\Delta])^2}, \tag{23}$$

which involves the convolution of R_{2w} with itself (see Appendix B). Although formulae (22) and (23) now allow straightforward numerical computation of the variance, it is still difficult to assess qualitatively how the parameters $w[n]$, N_w, Δ and q_i will affect the obtained results. Turning back to Fig. 5, one can see that $\mathscr{E}_\phi$ is lower bounded by the area of a rectangle of unit height and of length $\mathscr{L}_I=(q_2+q_1)\Delta-2N_w+2\Delta$, and upper bounded by the area of a rectangle of unit height and of length $\mathscr{L}_S=(q_2+q_1)\Delta+2N_w-2$. Hence, a reasonable approximation for $\mathscr{E}_\phi$ is the average of these two areas.

Corollary 2. *The asymptotic normalised variance of Rabiner's spectral estimator can be approximated by*

$$\frac{\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}}{S_{2v}(f)S_{2u}(f)} \simeq \frac{(q_2+q_1+1)}{(I-q_2-q_1)}, \quad q_i\neq 0, \quad i=1,2, \quad 0<|f|<\tfrac{1}{2}. \tag{24}$$

Several simulations have confirmed that the above remarkably simple approximation is quite sharp. Furthermore it has the following important interpretations:

(1) the normalised variance is approximately independent of the shape of the data-window $w[n]$, provided that $w[n]$ satisfies condition 9.
(2) the normalised variance is a crescent function of q_2 and q_1, indicating that Rabiner's spectral estimator has always more variability than its Welch's counterpart [12]; this is the price to pay for

achieving unbiasedness. In order to minimise the variance of Rabiner's estimator while still keeping unbiasedness as required by Theorem 1, q_2 and q_1 should therefore be set such that $q_i = \lceil (L_i + N_w)/\Delta \rceil - 1$, $i = 1, 2$.

Finally, using the large sample approximation $\left(I - q_2 - q_1\right)^{-1} \simeq I^{-1} \simeq N/\Delta$ and plugging $q_i \simeq (L_i + N_w)/\Delta - 1/2$, $i = 1, 2$ into Eq. (24) (possibly allowing q_i to be non-integer for the sake of simplicity), the normalised variance can finally be re-expressed as

$$\frac{\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}}{S_{2v}(f)S_{2u}(f)} \simeq \frac{1}{N}(L_2 + L_1 + 2N_w), \quad q_i \neq 0, \quad 0 < |f| < \tfrac{1}{2}. \tag{25}$$

This final result clearly evidences that *the normalised variance of Rabiner's estimator is essentially a crescent function of the autocorrelation length* $(L_1 + L_2)$ *and the window length* N_w.

4. Some practical guidelines

We are now in a position to give a complete algorithm for the measurement of FRF's by means of Rabiner's method. We consider here the typical case where $\{x[n]\}_{n=0}^{N-1}$ is the input to a linear time-invariant system and $\{y[n]\}_{n=0}^{N-1}$ is a noisy measurement of the output; hence $\widehat{H}_1(f) = \widehat{S}_{yx}(f)/\widehat{S}_{2x}(f)$ is the optimal estimator to be used. This is however without loss of generality since the same procedure will apply in exactly the same way to other estimators, e.g. H_2, H_3, H_v, H_a, etc.

4.1. Algorithm for FRF measurement

(1) Estimation of $S_{yx}(f)$:
  (a) Get a first guess of the support set $[-L_1; L_2]$ spanned by the cross-correlation $R_{yx}[k]$.
  (b) Choose a data-window $\{w[n]\}_{n=0}^{N_w-1}$ that satisfies $\sum_{i=-\infty}^{\infty} w[n - i\Delta] = C$ for a given length N_w and time increment $\Delta > 0$.
  (c) Set the values of q_2 and q_1 as small as possible while still keeping unbiasedness as required by Theorem 1, i.e. $q_i = \lceil (L_i + N_w)/\Delta \rceil - 1$, $i = 1, 2$.
  (d) Compute the discrete Fourier transform

$$Y_{N_w}^{(i)}(f_k) = \sum_{n=i\Delta}^{i\Delta+N_w-1} w[n - i\Delta] \cdot v[n] \mathrm{e}^{-\mathrm{j}2\pi(kn/N_{\mathrm{FFT}})}, \quad i = 0, \ldots, I-1 \tag{26}$$

  on the frequency grid $f_k = k/N_{\mathrm{FFT}}$, $k = 0, \ldots, N_{\mathrm{FFT}} - 1$, by using the FFT algorithm evaluated on $N_{\mathrm{FFT}} \geqslant N_w$ points. Repeat the computation for $X_{N_w}^{(i)}(f_k)$.
  (e) Get an estimate $\widehat{S}_{yx}^{(R)}(f_k)$ of $S_{yx}(f)$ at frequency $f = f_k$ by using Eq. (11).

(2) Estimation of $S_{2x}(f)$:
  Repeat steps (1a)–(1e) after replacing y with x. Use *all the same parameters* as for $S_{yx}(f)$.

(3) Estimation of $H(f)$:

$$\widehat{H}_1^{(R)}(f_k) = \widehat{S}_{yx}^{(R)}(f_k)/\widehat{S}_{2x}^{(R)}(f_k). \tag{27}$$

4.2. Setting the algorithm parameters

4.2.1. The autocorrelation length

The support set $[-L_1; L_2]$ spanned by the cross-correlation $R_{yx}[k]$ is ideally obtained from physical consideration or, if not possible otherwise, from inspection of the inverse DTFT of the raw cross-periodogram: $\mathrm{FFT}^{-1}\{\mathrm{FFT}\{y[n]\} \cdot \mathrm{FFT}\{x[n]\}^*\}$ where it is assumed that $N \gg L_1 + L_2$. It is important that L_1

and L_2 be not over-estimated such as to keep the FRF variance as low as possible. This assumes implicitly that $h[n]$ is a *finite-length impulse response* (FIR). In the case of *infinite-length impulse responses* (IIR), L_1 and L_2 should be truncated to some reasonable values after which the effective length of the IR can be regarded as negligible.

4.2.2. The data-window

Typical data-windows that (approximately or exactly) satisfy $\sum_{i=-\infty}^{\infty} w[n-i\Delta_k]=C$ for some $\Delta_k>0$ are the rectangular, Hanning and Hamming windows with $N_w/\Delta=2,3,4,\ldots$ the Blackman window with $N_w/\Delta=3,4,5,\ldots$ the Parzen window with $N_w/\Delta=2^k$, $k=1,2,3,\ldots$ and the Bartlett (triangular) window with $N_w/\Delta=2k$, $k=1,2,3,\ldots$. As demonstrated in Section 3.2, the performance of the algorithm virtually does not depend on the window type, provided that condition (19) is fulfilled. This is true for FIR's, but not necessarily for IIR's, the reason being that a IIR must be gently tapered to zero in Eq. (14) in order to be reasonably approximated by a FIR. Indeed, for $\phi[k]$ to have smooth tails that minimise boundary effects, *smooth data-windows should be used*. Hence the Hanning, Blackman or Parzen windows should be favoured as opposed to the Hamming, Bartlett, and rectangular (worst option) windows.

4.2.3. The time increment

The time increment $1\leqslant\Delta\leqslant N_w$ strongly affects the computational burden of the algorithm. If Δ is set too small, then not only is the number I of segments to be processed enlarged, but also the number of cross-terms (q_1+q_2+1). Therefore, we recommend to set Δ as small as allowed by the type of data-window, that is for instance $N_w/2$ (50% overlap) with Hanning and Parzen windows, and $N_w/3$ (67% overlap) with a Blackman window.

4.2.4. The data-window length

It is clear from the conclusions of Section 3.2 that the data-window length N_w should be set as small as possible in order to minimise the estimation variance. As discussed in Section 3.1, there is no limit in theory to setting N_w arbitrarily small, because Rabiner's estimator can handle the case $N_w<L_1+L_2+1$ (window length $<$ impulse response length). However, the practical limit is dictated by the computational demand which increases considerably as N_w becomes small because the number (q_1+q_2+1) of cross-terms in Eq. (11) is then enlarged.

4.2.5. The FFT length

The choice of the FFT length strongly depends on whether the user wishes an estimation of the IR (time-domain analysis) or of the FRF (frequency-domain analysis). Should the proposed algorithm be used for estimating an IR, viz $\widehat{h}[n]=\mathrm{FFT}^{-1}\{\widehat{H}(f_k)\}$, then it must be that $N_{\mathrm{FFT}}\geqslant L_1+L_2$ in order to avoid time-aliasing in $\widehat{h}[n]$. Should the proposed algorithm be used for estimating a FRF, e.g. to be subsequently used in frequency-domain parametric identification, then the only requirements are (i) $N_{\mathrm{FFT}}\geqslant N_w$ and (ii) $N_{\mathrm{FFT}}\geqslant 1/\Delta f$ where Δf is a user specified frequency resolution.

4.3. Performance of Rabiner's FRF estimator

4.3.1. Bias

Referring back to Eq. (6), it is now straightforward to check that Rabiner's FRF estimator $\widehat{H}_1^{(R)}(f)$ is unbiased provided that both $\widehat{S}_{yx}^{(R)}(f)$ and $\widehat{S}_{2x}^{(R)}(f)$ are unbiased.

4.3.2. Variance

The normalised variance of $\widehat{H}_1^{(R)}(f)$ is [3,4]

$$\frac{\mathrm{Var}\left\{\widehat{H}_1^{(R)}(f)\right\}}{|H_1(f)|^2}\simeq\left[\frac{\mathrm{Var}\left\{\widehat{S}_{yx}^{(R)}(f)\right\}}{|S_{yx}(f)|^2}+\frac{\mathrm{Var}\left\{\widehat{S}_{2x}^{(R)}(f)\right\}}{S_{2x}(f)^2}-2\mathscr{R}\left\{\frac{\mathrm{Cov}\left\{\widehat{S}_{yx}^{(R)}(f);\widehat{S}_{2x}^{(R)}(f)\right\}}{S_{yx}(f)S_{2x}(f)}\right\}\right]. \tag{28}$$

Using the results derived in Section 3.2, it is readily checked that the above expression is minimised provided that the *same parameters* are used in the estimation of $\widehat{S}_{yx}^{(R)}(f)$ and $\widehat{S}_{2x}^{(R)}(f)$. Then,

$$\frac{\mathrm{Var}\left\{\widehat{H}_1^{(R)}(f)\right\}}{|H_1(f)|^2} \simeq \frac{\mathscr{E}_\phi}{\Delta(I-q_2-q_1)} \cdot \frac{1-\gamma_{yx}^2(f)}{\gamma_{yx}^2(f)}, \quad 0<|f|<\tfrac{1}{2} \tag{29}$$

with $\gamma_{yx}^2(f)$ the squared-magnitude coherence between signals y and x.

5. Examples of application and comparisons

This section illustrates the practical use of Rabiner's method for the measurement of FRF's. The first example provides an illustration of the theoretical results derived in the former sections. The second example is concerned with the measurement of the mechanical impedance of a lightly damped structure together with the extraction of its modal parameters, such as commonly encountered in experimental modal analysis.

5.1. Example 1

This first example deals with the measurement of a running average filter of $L=30$ coefficients, i.e.

$$y[n] = \sum_{k=0}^{29} \frac{1}{30} x[n-k] + b[n], \tag{30}$$

where $b[n]$ accounts for additive measurement noise. This is a difficult task to carried out in the frequency-domain, because the FRF of the running average filter exhibits a series of zeros (see Fig. 7) which rapidly get smeared out due to the excessive leakage introduced by Welch's spectral estimator. The inverse DTFT of the so-measured FRF then suffers from considerable distortion when compared with the expected box-shape of the theoretical result.

The following experiments were carried out. A set of input–output signals were simulated according to model (30), the input x being a white Gaussian noise, and the output y having a signal-to-noise ratio of 10 dB. The signal length was $N=5000$ samples. These details are resumed in Table 1. Welch's and Rabiner's estimators were then implemented with a 32-sample Hanning window (remember that having N_w larger than the IR length $L=30$ is necessary for Welch's algorithm only), and 50% overlap; thus $\Delta=(1-50/100)N_w=16$. The FFT length was set to $N_{\mathrm{FFT}}=512$ frequency lines so as to keep a fine frequency resolution, but smaller values could have been used just as well provided that $N_{\mathrm{FFT}} \geqslant N_w$ as indicated in Section 4.2.4. The setting of parameters q_1 and q_2 is now purposely detailed. First, we knew that the IR of interest was (i) causal and (ii) was excited by white noise; point (i) then implied that the cross-correlation $R_{yx}[k]$ was also causal and point and (ii) that it had the same length as the IR. Hence $L_1=0$ and $L_2=L-1=29$ (see Fig. 3) and straight application of Eq. (19) gave the conditions

$$q_1 \geqslant \left\lceil \frac{L_1+N_w}{\Delta} \right\rceil - 1 = \left\lceil \frac{0+32}{16} \right\rceil - 1 = 1 \tag{31}$$

Table 1
Parameter settings for example 1

Signal length	$N=5000$ samples
Window type	Hanning
Window length	$N_w=32$
Overlap	50%
FFT length	$N_{\mathrm{FFT}}=512$ lines
Additive noise type	Gaussian
Input signal-to-noise ratio	∞
Output signal-to-noise ratio	10 dB

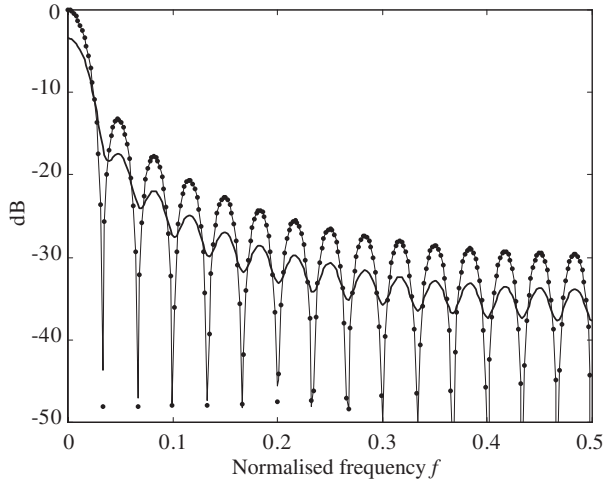

Fig. 6. Measured FRF (averaged over 50 experiments). Comparison of Rabiner's (thin line) and Welch's (thick line) estimates with the true FRF (dotted line). Note that Rabiner's estimate is perfectly superposed on the true FRF over the whole frequency range.

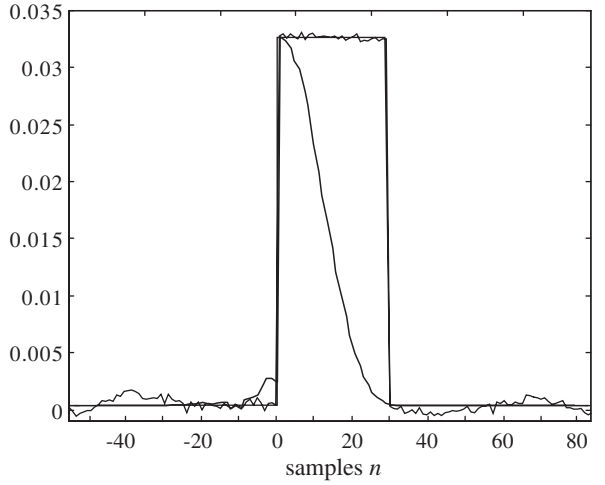

Fig. 7. Measured IR (1 experiment). Comparison of Rabiner's (thin line) and Welch's (thick line) estimates with the true IR (dotted line). Note that Rabiner's estimate is perfectly superposed on the true FRF.

and

$$q_2 \geqslant \left\lceil \frac{L_2 + N_w}{\Delta} \right\rceil - 1 = \left\lceil \frac{29 + 32}{16} \right\rceil - 1 = 3 \tag{32}$$

for Rabiner's estimate to be unbiased. Eventually, the smallest possible values $q_1 = 1$ and $q_2 = 3$ were retained so as to minimise variability as explained in Section 4.1.

The measured FRF and FIR obtained with these parameter settings are displayed in Figs. 6 and 7, respectively. As expected, Welch's estimate exhibits excessive leakage in the frequency-domain, which seriously deteriorates the expected structure of the FRF. This results in a highly distorted IR in the time-domain. On the other hand, Rabiner's estimator was perfectly able to recover the theoretical FRF, and hence the correct IR. As explained in Section 3.1, this is independently of the data-window length N_w. However, the price of Rabiners' method versus Welch's method is a higher variance. Using either Eq. (22) or (24) of

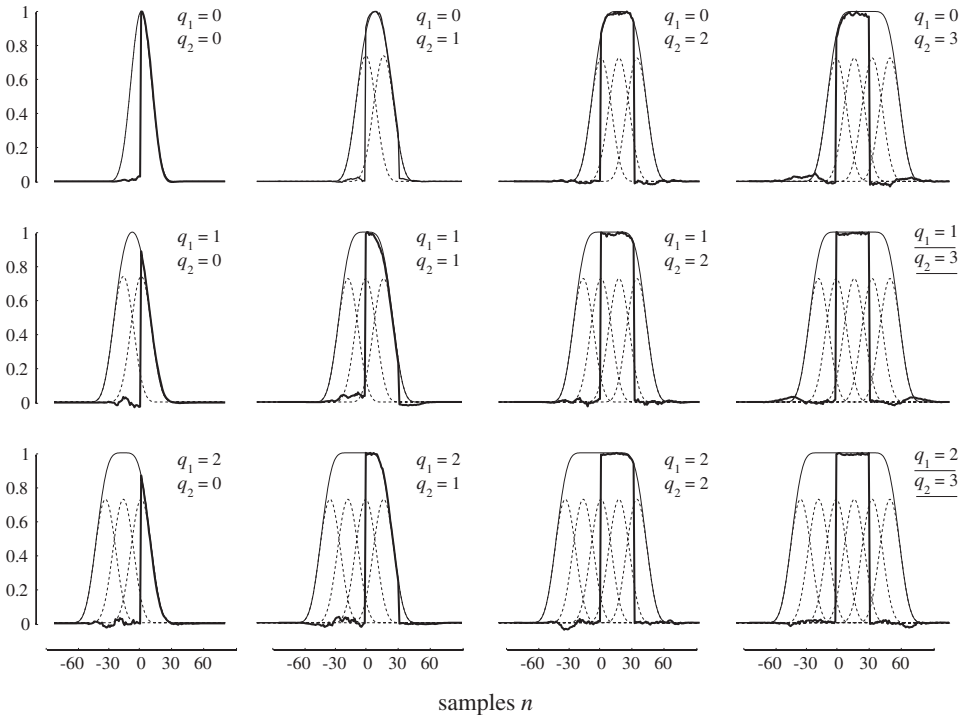

Fig. 8. Measured IR's (thick lines) for different values of the pair $\{q_1; q_2\}$. The results are superposed with the corresponding lag-windows $\phi[k]$ (thin lines) as obtained from translating and summing the auto-correlation function $R_{2w}[k]$ (dotted lines) of the data window $w[n]$.

Section 3.2, the variance of Rabiner's method was found to be 0.0163. This is to be compared with 0.0031 for Welch's method, for which only Eq. (22) can be used, that is a factor five difference.

Incidentally, the simplicity of the IR considered in the present example allows an easy illustration of the mechanism of Rabiner's method. Fig. 8 displays the measured IR's for several possible values of the pair $\{q_1; q_2\}$. It is seen that, as explained in Section 3.1, the auto-correlation function $R_{2w}[k]$ is translated q_1 times to the left and q_2 times to the right by steps of Δ samples. The summation of these translates then produces the lag-window $\phi[k]$, with a flat summit if $q_1 + q_2$ is large enough. The condition to get an unbiased IR is then that the flat part of $\phi[k]$ completely covers the support of $h[n]$, that is the interval $[0; 29]$ in the present case. This is seen to occur for $q_1 \geqslant 1$ and $q_2 \geqslant 3$ in Fig. 8, in accordance with results (35) and (36).

Finally, the same experiment was further repeated for different window lengths N_w. Results were also compared with the so-called "improved FRF" method proposed by Schoukens et al. which, as far we known, is the only alternative introduced so far that can cope with the present issue [9,5]. The "improved FRF" method is based on recognising that short-time DTFT's lead to transient effects, see Fig. 9, which can be compensated for in a recursive manner. In order to briefly explain the rationale of the method, let $e[n] = y[n]w_i[n] - h * (x[n]w_i[n])$ be the transient produced by the difference of the ith output segment with the system response to the ith input segment, as illustrated in Fig. 9. In Welch's method, it is erroneously assumed that $e[n] \equiv 0$ and it is actually where leakage comes from. Nevertheless, if $e[n]$ was known, Welch's FRF measurement could be corrected by extracting from signal e all the information still correlated with x. This is the idea of the "improved FRF" method which recursively estimates e from the current IR estimate:

$$e[n] = y[n] - \widehat{h}^{(r)} * x[n], \quad \widehat{H}^{(r+1)} = \widehat{H}^{(r)} + \widehat{S}_{ex}^{(W)} / \widehat{S}_{2x}^{(W)}. \tag{33}$$

Note that this algorithm is computationally demanding as it involves an inverse DTFT and a filtering operation at each recursion until convergence is reached. Into addition, its exact performance analysis is difficult and so far no general analytical results have been derived in this direction.

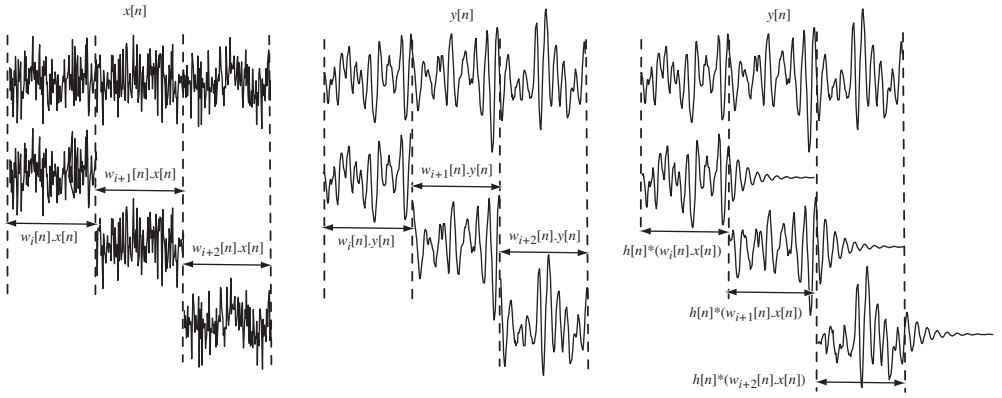

Fig. 9. Principle of the "improved FRF" method: the division of the signal into short-time segments implicitly produces transient effects which can be compensated for given the knowledge of the IR.

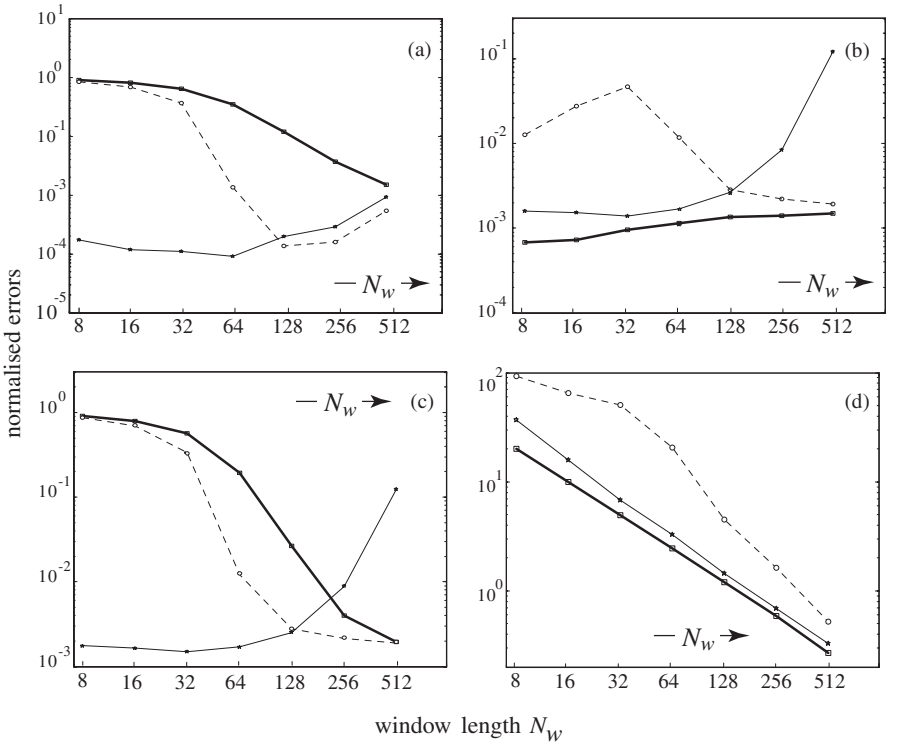

Fig. 10. Comparison of Welch's (thick lines), Rabiner's (thin lines) and "improved FRF" (dotted lines) methods in terms of (a) the normalised squared bias; (b) the normalised variance; (c) the normalised mean square error (NMSE); and (d) the computational time.

The comparisons between Welch's, Rabiner's and the improved FRF methods are displayed in Fig. 10(a–d) in terms of the normalised bias, the normalised variance, and the normalised mean square error (MSE), respectively.[6] These statistics have been obtained over 200 runs and have also been frequency-averaged. It can be verified from these figures that:

(1) Welch's method requires excessively long windows before it can significantly reduce its induced leakage, and hence its MSE.

[6]Note that: normalised squared bias + normalised variance = normalised MSE.

Table 2
Parameter settings for example 2

Signal length	$N = 100,000$ samples
Window type	Hanning
Window length	$N_w = 512$
Overlap	50%
FFT length	$N_{\mathrm{FFT}} = 2048$ lines
Additive noise type	Gaussian
Input signal-to-noise ratio	∞
Output signal-to-noise ratio	10 dB

(2) As expected Rabiner's method has a very small bias. However, its variance increases markedly with N_w and therefore its MSE becomes larger than that of the other methods when $N_w \gg L$. This is consistent with the results derived in Section 3.2: again, we recommend using short data-windows with Rabiner's method.
(3) With respect to bias the "improved FRF" method shows a rather similar behaviour than Rabiner's method for large values of N_w, yet it does not perform as well for reasonable values of N_w. The "improved FRF" method actually requires N_w being significantly larger than L (say $N_w \geqslant 4L$) before reaching satisfying results,[7] an observation that is consistent with the conclusion of Ref. [9]. This plus the fact that it requires the most computational time impose some limitations to the latter method when used with long or infinite-length IR as will be the case in the next example.

5.2. Example 2

This second example deals with the frequency-domain measurement of a lightly damped IIR. This is a case of actual importance since it is commonly encountered in the experimental modal analysis of mechanical structures. In this situation, Rabiner's estimator cannot be made exactly unbiased because the infinite-length IR must be approximated by an equivalent finite-length IR; as explained in Section 2.2 if q_1 and/or q_2 were made as large as possible, then the corresponding spectral estimate would reduce to the classical periodogram and would not be consistent. Nevertheless, even with under-estimated q_1 and q_2, Rabiner's method still provides a drastic improvement over Welch's method which imposes that $q_1 = q_2 = 0$, and also over the "improved FRF" method which poorly copes with IIR [9]. This is illustrated here on a synthetic mechanical impedance with three natural frequencies at 0.1, 0.19 and 0.22 (normalised frequency) with damping ratios of 1.60%, 0.84% and 0.73%, respectively. The corresponding FRF was simulated using the model

$$H(z) = \frac{1 - 1.5533z^{-1} + 2.4826z^{-2} - 1.5224z^{-3} + 0.9606z^{-4}}{1 - 2.7018z^{-1} + 4.9726z^{-2} - 5.7292z^{-3} + 4.8737z^{-4} - 2.5953z^{-5} + 0.9415z^{-6}}, \tag{34}$$

with $z = \mathrm{e}^{\mathrm{j}2\pi f}$. A set of input–output signals was generated according to this model, the input x being a white Gaussian noise, and the output y having a signal-to-noise ratio of 10 dB. The signal length was $N = 100,000$ samples. These details are resumed in Table 2. Contrary to the previous example, we are here in a situation where we did not know a priori the IR (effective) length L. Consequently, it had to be "guessed" from the empirical cross-correlation $\hat{R}_{yx}[k]$ as suggested in Section 4.2.1. Fig. 11 displays this quantity measured directly from $\mathrm{FFT}^{-1}\{\mathrm{FFT}\{y[n]\} \cdot \mathrm{FFT}\{x[n]\}^*\}$ with a FFT length of 2^{17} lines. It can be seen from this figure that the cross-correlation is (i) causal (as expected from a physical system excited by white noise) and (ii) has an effective length of about 500 samples (after this limit all non-zero values seem to pertain to estimation noise). This preliminary analysis therefore suggested that $L_1 = 0$ and $L_2 \simeq 500$. Since for Welch's estimator the customary practice is to set $N_w \geqslant L_1 + L_2 + 1$, we chose $N_w = 512$. The same N_w was used with Rabiner's

[7] It is remarkable that the "improved FRF" method achieves a better MSE than Rabiner's method for long windows (as compared to the system time-constant), and vice versa for short windows. Work is currently in progress so as to provide practical guidelines for deciding when one method should be preferred to the other.

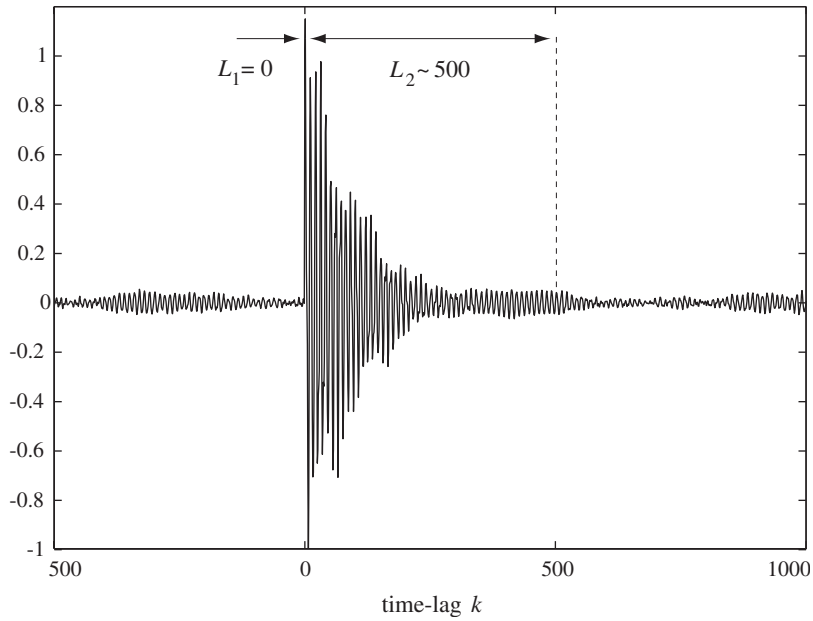

Fig. 11. Measured cross-correlation function $R_{yx}[k]$, from which the values of L_1 and L_2, and consequently q_1 and q_2, can be guessed.

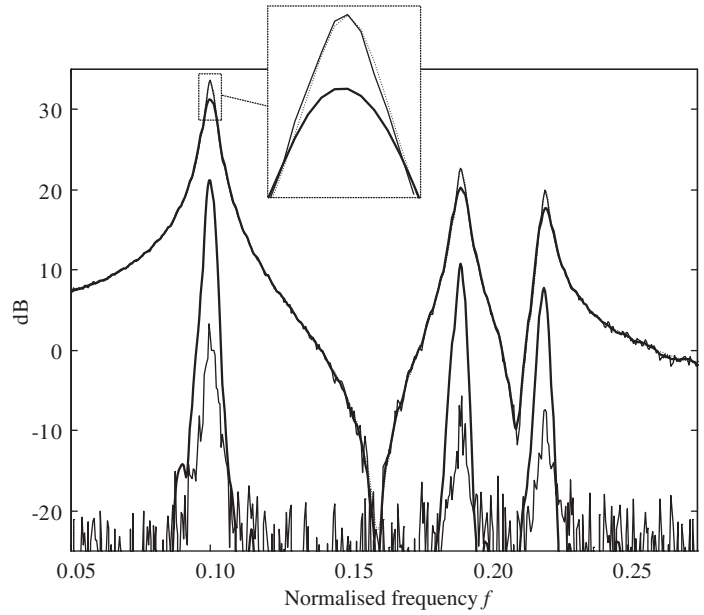

Fig. 12. Measured FRF's of a lightly damped mechanical structure with three poles at $f = 0.1$, 0.19 and 0.22 (Rabiner: thin line, Welch: thick line, theoretical: dotted line). The figure also shows the resulting bias (absolute magnitude in dB) which, according to Eq. (7), has strong maxima at the pole locations.

estimator, but once again smaller values could have been used—and are even recommended for the reasons mentioned in Section 3.2. A Hanning window was used with 50% overlap so that $\Delta = (1 - 50/100)N_w = 256$, and FFT's were run with $N_{\mathrm{FFT}} = 2048$ frequency lines. Straight application of Eq. (19) then gave the

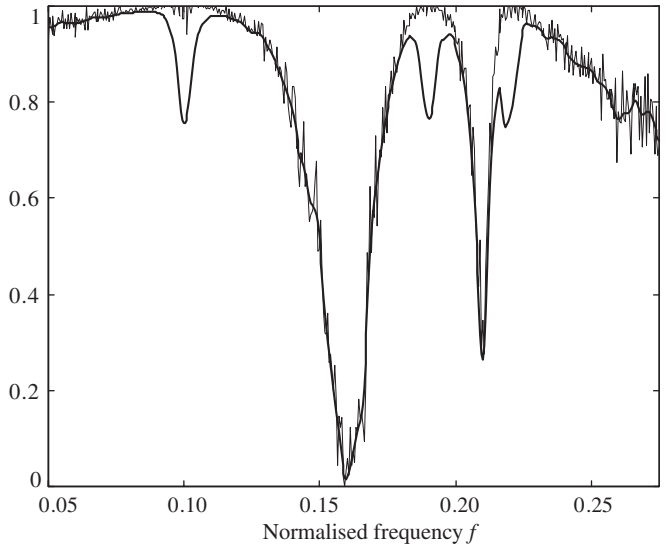

Fig. 13. Measured coherence function $\gamma^2_{yx}(f)$ (Rabiner: thin line, Welch: thick line). Note that the sharp dips at the pole locations in Welch's estimate are artifacts due to leakage; they are suppressed in Rabiner's estimate. Having an accurate measurement of the coherence function is of importance in the evaluation of the FRF variance as given by Eq. (29).

Table 3
Comparison of true and estimated natural frequencies

	f_1	Std	f_2	Std	f_3	Std
True	0.1		0.19		0.22	
Welch	0.10001	0.2 E-4	0.18999	0.2 E-4	0.21991	0.2 E-4
Rabiner	0.10000	0.2 E-4	0.19000	0.2 E-4	0.22000	0.2 E-4

conditions

$$q_1 \geqslant \left\lceil \frac{L_1 + N_w}{\Delta} \right\rceil - 1 = \left\lceil \frac{0 + 512}{256} \right\rceil - 1 = 1 \tag{35}$$

and

$$q_2 \geqslant \left\lceil \frac{L_2 + N_w}{\Delta} \right\rceil - 1 = \left\lceil \frac{500 + 512}{256} \right\rceil - 1 = 3. \tag{36}$$

The measured FRF's are displayed in Fig. 12 together with the resulting bias (absolute magnitude in dB). In accordance with Eq. (7) of Section 3.2, the bias was found to have strong maxima in the vicinity of the poles locations where the effect of leakage was the most severe. Note that in the case of an IIR, Rabiner's estimator could not completely remove the bias, however its superiority over Welch was still clearly verified. The variance was found to be 0.0130 for Rabiner and 0.0025 for Welch, i.e. a factor five difference. Again, this increased variability is the price to pay for a low-bias estimation. Although this is not necessarily what the engineers aims at, there are however important applications where it is of prime importance to favour unbiasedness. One such application is modal parameter identification as illustrated below.

Following the customary approach in modern experimental modal analysis, the measured FRF's were subsequently fed into a parametric frequency-domain method in order to extract the modal parameters. In such a situation, knowledge of the estimation variance as given by Eq. (29) is of prime importance as it allows the use of very efficient identification techniques, such as the frequency-domain maximum likelihood (FDML)

Table 4
Comparison of true and estimated damping ratios

	ζ_1 (%)	Std (%)	ζ_2 (%)	Std (%)	ζ_3 (%)	Std (%)
True	1.60		0.84		0.73	
Welch	2.16	0.02	1.14	0.01	0.97	0.01
Rabiner	1.60	0.02	0.84	0.01	0.73	0.01

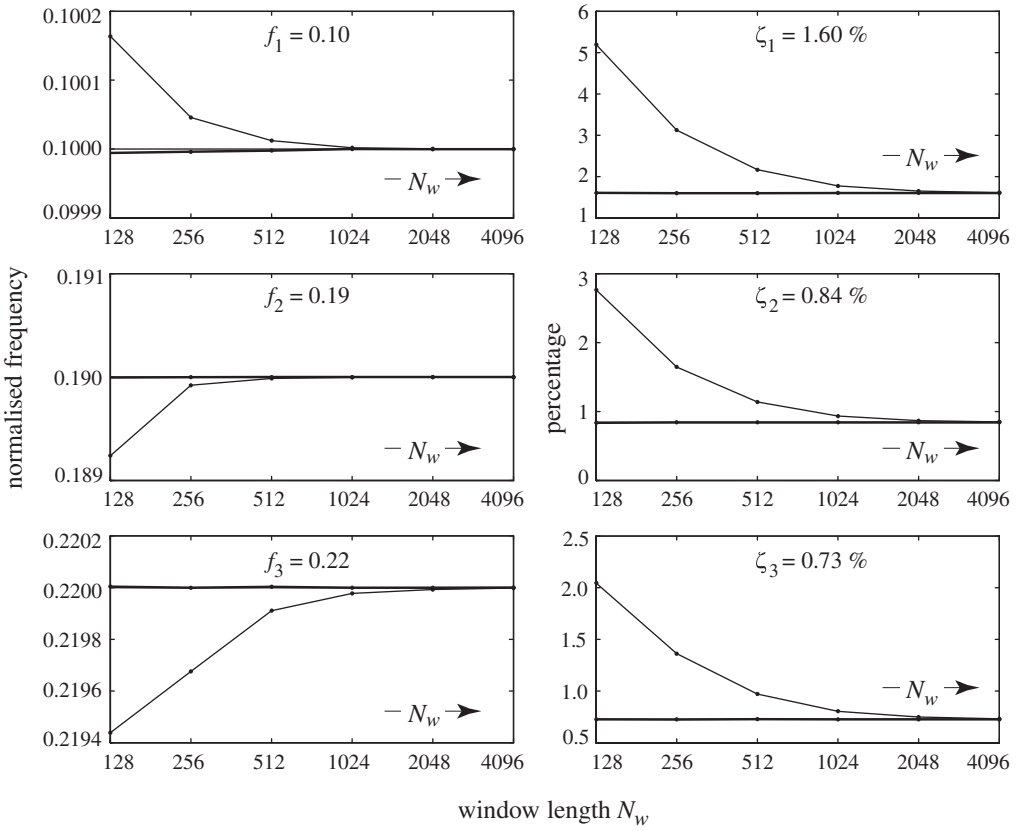

Fig. 14. Estimated natural frequencies f_1, f_2, f_3 and damping ratios $\zeta_1, \zeta_2, \zeta_3$ as functions of the window size N_w. Results are averaged over 200 experiments. The true values are indicated in each subplot. The thick and thin lines refer to Welch's and Rabiner's estimates, respectively.

algorithm [5,14] (Fig. 13). Some results of the FDML algorithm with $N_w = 512$ are displayed in Tables 3 and 4, in terms of the averaged estimated natural frequencies and damping ratios together with their standard deviations as obtained from 200 experiments. Similar results are displayed in Figs. 14 and 15 for different values of the window length N_w. As expected, it is seen that Rabiner's method produces estimates with an excellent accuracy whatever the window length, whereas Welch's method can only achieve a similar quality for very long windows[8] (say $N_w \geqslant 4L$). This is particularly true for the estimation of damping ratios, which are known to be a major problem in non-parametric frequency-domain identification; they are systematically

[8]Incidentally, this experiment makes it clear that Welch's method can achieve nearly as accurate results as Rabiner's method provided that long enough windows are used. However, doing so is not always possible. For instance if a sampling frequency 10 times as great was used to analyse the system of example 2, a window length of about 40 960, i.e. as many lines in the frequency domain, would be necessary with Welch's method to get accurate damping ratios, a value that is beyond the capability of many commercial spectral analysers.

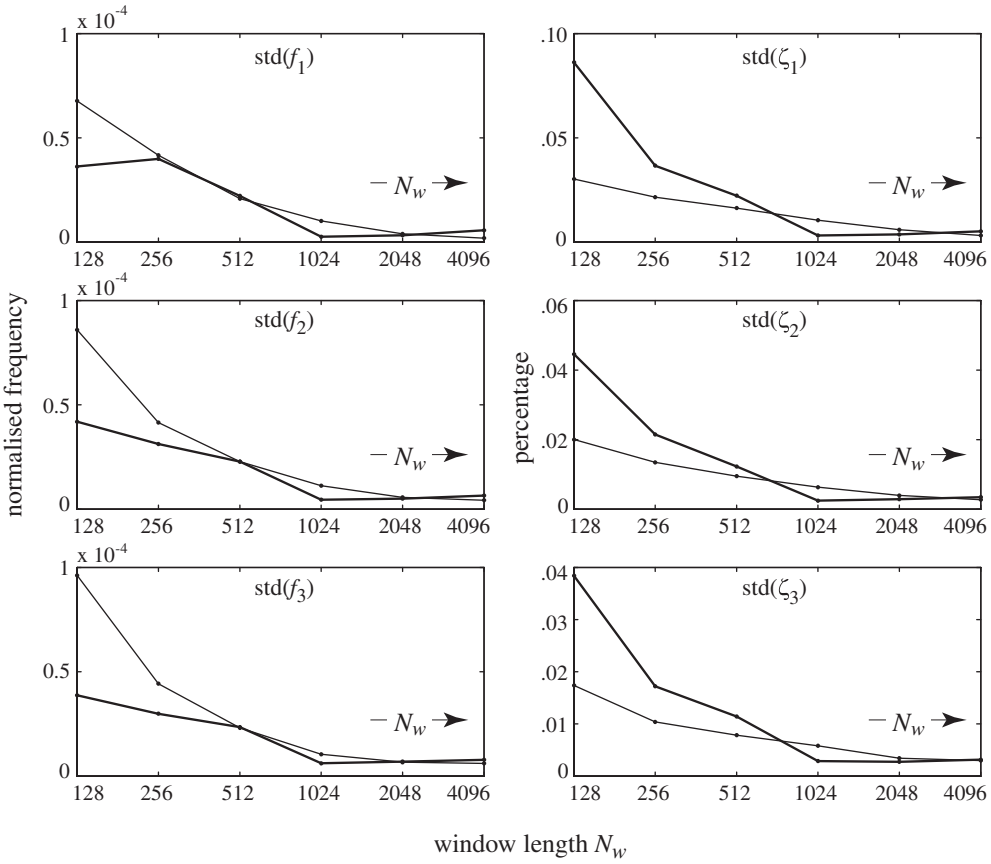

Fig. 15. Standard deviations over 200 experiments of the natural frequencies f_1, f_2, f_3 and damping ratios ζ_1, ζ_2, ζ_3 as functions of the window size N_w. The thick and thin lines refer to Welch's and Rabiner's estimates, respectively.

over-estimated in Welch's method. Very interestingly, Fig. 15 also shows that when it comes to the identification of modal parameters, the standard deviation of Rabiner's method is only slightly inflated as compared to Welch's method, and the difference even tend to disappear when increasing the window length. This is in clear contrast with what was previously observed with the FRF variance, e.g. in Fig. 10(b). The exact reason of this appealing behaviour remains to be investigated.[9] In any case it provides one more argument in favour of Rabiner's method. As a conclusion to this experiment, the accuracy of Rabiner's method plus the fact that it is accompanied with an easy formula for the computation of its variance, e.g. as required in maximum likelihood identification, make it a foremost candidate to be used in subsequent modal-parameters identification algorithms.

6. Conclusion

The aim of this paper was a reviving of the so-called Rabiner's method. Rabiner's method is dedicated to the frequency-domain non-parametric measurement of frequency response functions and has the major, and rather unique, advantage of producing leakage-free estimates even though it is based on the discrete-time Fourier transform. More than giving a mere review of the method, this paper has presented it through a simple

[9]One possible explanation is that the standard deviation of the estimated modal parameters is a function of both the standard deviation *and* the bias of the measured FRF due to the highly nonlinear character of the parametric identification process; the reduced bias of Rabiner's FRF may then result in estimated modal parameters with a compensated variability.

and comprehensive formulation which has lead to new insights and some original analytical results. In particular, we have established the exact conditions under which Rabiner's estimator is unbiased. An interesting finding is that unbiasedness can be achieved whatever the length of the data-window, be it longer or shorter than the impulse response. This provides potential improvement of the technique currently used in commercial spectral analysers with fixed-sized short FFT's. Another useful finding is a sharp approximation of the estimator variance which can be readily used for a fast evaluation of the estimates accuracy. Knowledge of the estimation variance is also of first importance when parametric identification is to be applied subsequently on the measured frequency response responses as is customary in modern experimental modal analysis. These results have finally led to a number of practical guidelines so as how to optimally set the algorithm parameters. When concerned with the bias issue, the superiority of Rabiner's estimator has been verified on realistic examples, and in particular in the case of lightly damped (infinite-length) impulse responses such as those commonly encountered in the vibration analysis of mechanical structures. In such situations it provides accurate estimates of the damping ratios which cannot be achieved by other classical spectral estimators. One disadvantage of Rabiner's estimator is that it suffers from a higher variability than its competitors. This may well be the reason why other researchers in the field have not accepted it. However, we argue that this is no more a problem nowadays thanks to the huge capacity of current data acquisition system which allows recording very long time signals, in consequence of which the variance of Rabiner's algorithm can easily be pulled down to very small values.

We hope this correspondence will give a new breath to a technique whose contribution, we think, had simply been forgotten for many years, although it deserved real attention due to its high accuracy and computational efficiency.

Appendix A. Bias

Our analysis is based on Cramér's spectral decomposition of the two stationary random signals v and u, viz:

$$v[n] = \int_{-1/2}^{+1/2} \mathrm{e}^{\mathrm{j}2\pi fn}\,\mathrm{d}V(f) \tag{A.1}$$

and similarly for $u[n]$, where $\mathrm{d}V(f)$ and $\mathrm{d}U(f)$ are spectral increments such that

$$\mathbb{E}\{\mathrm{d}V(f)\,\mathrm{d}U(f-\lambda)^*\} = S_{vu}(f)\delta(\lambda)\,\mathrm{d}f\,\mathrm{d}\lambda, \quad |\lambda| < \tfrac{1}{2}. \tag{A.2}$$

Substituting Cramér's spectral decomposition into the definition of $V_{N_w}^{(i+q)}(f)$ and $U_{N_w}^{(i)}(f)$ then yields

$$V_{N_w}^{(i+q)}(f) = \int_{-1/2}^{+1/2} \mathrm{d}V(f-\lambda)W_{i+q}(\lambda), \tag{A.3}$$

where $W_{i+q}(\lambda) = \mathscr{F}\{w_{i+q}[n]\}$ and similarly for $U_{N_w}^{(i)}(f)$. In turn, Eq. (11) then becomes

$$\widehat{S}_{vu}^{(R)}(f) = \frac{\theta}{I-q_2-q_1} \sum_{i=q_1}^{I-q_2-1} \sum_{q=-q_1}^{q_2} \int\int_{-1/2}^{+1/2} \mathrm{d}V(f-\lambda_1)\,\mathrm{d}U(f-\lambda_2)^* W_{i+q}(\lambda_1)W_i(\lambda_2)^*. \tag{A.4}$$

Taking the expected value and making use of property (A.2) of the spectral increments gives

$$\mathbb{E}\left\{\widehat{S}_{vu}^{(R)}(f)\right\} = \frac{\theta}{I-q_2-q_1} \sum_{i=q_1}^{I-q_2-1} \sum_{q=-q_1}^{q_2} \int_{-1/2}^{+1/2} S_{vu}(f-\lambda)W_{i+q}(\lambda)W_i(\lambda)^*\,\mathrm{d}\lambda. \tag{A.5}$$

The last step is to note that $W_{i+q}(\lambda) = W(\lambda)\mathrm{e}^{-\mathrm{j}2\pi\lambda(i+q)\Delta}$ with $W(\lambda) = \mathscr{F}\{w[n]\}$ so that, finally

$$\sum_{i=q_1}^{I-q_2-1} \sum_{q=-q_1}^{q_2} W_{i+q}(\lambda)W_i(\lambda)^* = \left|W(\lambda)\right|^2 \cdot (I-q_2-q_1) \cdot \mathrm{e}^{-\mathrm{j}\pi\lambda\Delta(q_2-q_1)} \cdot D_{q_2+q_1+1}(\lambda\Delta) \tag{A.6}$$

from which Eqs. (12) and (13) immediately follow.

Appendix B. Variance

By making use of expression (A.3) and assuming that the measurements are approximatively Gaussian for the sake of simplicity, it readily comes that ($0<|f|<\frac{1}{2}$):

$$\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}=\left(\frac{\theta}{I-q_2-q_1}\right)^2\cdot\sum_{\substack{-q_1\leqslant q,s\leqslant q_2\\ q_1\leqslant r,m<I-q_2}}\mathbb{E}\left\{V_{N_w}^{(m+q)}(f)V_{N_w}^{(r+s)}(f)^*\right\}\mathbb{E}\left\{U_{N_w}^{(r)}(f)U_{N_w}^{(m)}(f)^*\right\}. \tag{B.1}$$

From the previous analysis of the bias, $\mathbb{E}\left\{V_{N_w}^{(k)}(f)V_{N_w}^{(l)}(f)^*\right\}$ is known to depend only on the index difference $k-l$. Therefore, setting $j=q-s$ and $i=r-m$ in Eq. (B.1) gives

$$\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}=\left(\frac{\theta}{I-q_2-q_1}\right)^2\cdot\sum_{m,i}\mathbb{E}\left\{U_{N_w}^{(m+i)}(f)U_{N_w}^{(m)}(f)^*\right\}\sum_{q,j}\mathbb{E}\left\{V_{N_w}^{(m+q)}(f)V_{N_w}^{(m+q+i-j)}(f)^*\right\}. \tag{B.2}$$

The last sum in the above equation expands as

$$\sum_{q,j}\mathbb{E}\left\{V_{N_w}^{(m+q)}(f)V_{N_w}^{(m+q+i-j)}(f)^*\right\}=\int_{-1/2}^{+1/2}S_{2v}(f-\lambda_1)|W(\lambda_1)|^2\sum_{q=-q_1}^{q_2}\sum_{j=q-q_2}^{q+q_1}\mathrm{e}^{-\mathrm{j}2\pi\lambda_1(j-i)\Delta}\,\mathrm{d}\lambda_1, \tag{B.3}$$

where $\sum_{q=-q_1}^{q_2}\sum_{j=q-q_2}^{q+q_1}\mathrm{e}^{-\mathrm{j}2\pi\lambda_1(j-i)\Delta}=\mathrm{e}^{\mathrm{j}2\pi\lambda_1 i\Delta}\cdot D_{q_2+q_1+1}^2(\Delta\lambda_1)$ which still depends on index i. The middle sum times $\mathrm{e}^{\mathrm{j}2\pi\lambda_1 i\Delta}$ expands as

$$\sum_{m,i}\mathrm{e}^{\mathrm{j}2\pi\lambda_1 i\Delta}\cdot\mathbb{E}\left\{U_{N_w}^{(m+i)}(f)U_{N_w}^{(m)}(f)^*\right\}=\int_{-1/2}^{+1/2}S_{2u}(f-\lambda_2)|W(\lambda_2)|^2\sum_{m=q_1}^{I-q_2-1}\sum_{i=q_1-m}^{I-q_2-1-m}\mathrm{e}^{-\mathrm{j}2\pi(\lambda_2-\lambda_1)i\Delta}\,\mathrm{d}\lambda_2, \tag{B.4}$$

where $\sum_{m=q_1}^{I-q_2-1}\sum_{i=q_1-m}^{I-q_2-1-m}\mathrm{e}^{-\mathrm{j}2\pi(\lambda_2-\lambda_1)i\Delta}=D_{I-q_2-q_1}^2(\Delta[\lambda_2-\lambda_1])$. Putting all results together then gives Eq. (21). Furthermore if I is assumed large, then $D_{I-q_2-q_1}^2(\Delta x)\simeq(I-q_2-q_1)\delta(x)/\Delta$ in the vicinity of $x=0$. Hence,

$$\mathrm{Var}\left\{\widehat{S}_{vu}^{(R)}(f)\right\}\simeq\frac{S_{2v}(f)S_{2u}(f)}{\Delta(I-q_2-q_1)}\int_{-1/2}^{+1/2}\theta^2\cdot D_{q_2+q_1+1}^2(\lambda\Delta)\cdot|W(\lambda)|^4\,\mathrm{d}\lambda, \tag{B.5}$$

where $S_{2v}(f)$ and $S_{2u}(f)$ have been brought out from the integral because they are assumed smooth functions of frequency. Note next that $\theta^2\cdot D_{q_2+q_1+1}^2(\lambda\Delta)\cdot|W(\lambda)|^4=|\Phi(\lambda)|^2$ as defined in Eq. (13). Therefore,

$$\int_{-1/2}^{+1/2}|\Phi(\lambda)|^2\,\mathrm{d}\lambda=\theta^2\sum_k\phi[k]^2=\theta^2(q_2+q_1+1)\sum_{q=-q_1-q_2}^{q_1+q_2}(R_{2w}*R_{2w})[q\Delta], \tag{B.6}$$

where $(R_{2w}*R_{2w})[q\Delta]=\sum_k R_{2w}[k+q\Delta]R_{2w}[k]$ and $\theta=(\sum_{q=-q_1}^{q_2}R_{2w}[q\Delta])^{-1}$. Inserting the latter results into Eq. (B.5) then yields Eqs. (22) and (23).

References

[1] B. Randall, *Frequency Analysis*, third ed., Bruel & Kjaer, 1987.

[2] R. Blackman, J. Tukey, *The Measurement of Power Spectra*, Dower Publications, New York, 1958.

[3] G. Jenkins, D. Watts, *Spectral Analysis and its Applications*, Emerson-Adams Press, 1998.

[4] J. Bendat, A. Piersol, *Random Data—Analysis and Measurement Procedures*, second ed., Wiley Interscience, New York, 1986.

[5] R. Pintelon, J. Schoukens, *System Identification—A Frequency Domain Approach*, IEEE Press, New York, 2001.

[6] J. Antoni, P. Wagstaff, J.-C. Henrio, H_α—a consistent estimator for frequency response functions with input and output noise, *IEEE Instrumentation and Measurement* 53 (2) (2004) 457–465.

[7] D. Percival, A. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, Cambridge, 1993.

[8] L. Rabiner, J. Allen, Short-time Fourier analysis techniques for FIR system identification and power spectrum estimation, *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-27 (2) (1979) 182–192.

[9] J. Schoukens, Y. Rolain, R. Pintelon, Improved frequency response function measurements for random noise excitations, *IEEE Transactions on Instrumentations and Measurements* 47 (1) (1998) 322–326.

[10] L. Rabiner, J. Allen, On the implementation of a short-time spectral analysis method for system identification, *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP*- 28 (1) (1980) 69–78.
[11] P. Welch, The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics* 15 (1967) 70–73.
[12] J. Mathews, D. Youn, Analysis of the short-time unbiased spectrum estimation algorithm, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 33 (1) (1985) 136–142.
[13] J. Schoukens, Y. Rolain, R. Pintelon, Analysis of windowing/leakage effects in frequency response function measurements, in: *16th IFAC World Congress*, Prague, Czech Republic, 4–8 July 2005.
[14] J. Schoukens, R. Pintelon, *Identification of Linear Systems*, Pergamon Press, Oxford, 1991.