The myth of data acquisition rate

Attila Felinger [a,b,*], Anikó Kilár [a], Borbála Boros [b]

[a] MTA–PTE Molecular Interactions in Separation Science Research Group, Ifjúság útja 6, H-7624 Pécs, Hungary
[b] Department of Analytical and Environmental Chemistry and Szentágothai Research Center, University of Pécs, Ifjúság útja 6, H-7624 Pécs, Hungary

HIGHLIGHTS

- Data acquisition rate has no influence on band broadening and hence resolution.
- Software packages often contain undocumented features.
- Undisclosed digital filtering is often coupled with data acquisition.

GRAPHICAL ABSTRACT
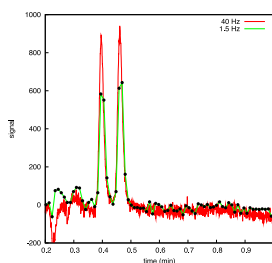
ABSTRACT

With the need for high-frequency data acquisition, the influence of the data acquisition rate on the quality of the digitized signal is often discussed and also misinterpreted. In this study we show that undersampling of the signal, i.e. low data acquisition rate will not cause band broadening. Users of modern instrumentation and authors are frequently misled by hidden features of the data handling software they use. Very often users are unaware of the noise filtering algorithms that run parallel with data acquisition and that lack of information misleads them. We also demonstrate that undersampled signals can be restored by a proper trigonometric interpolation.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The quality of digitized signals has been widely studied. Most often the effect of data acquisition rate is studied in terms of minimum data points per peak required for the accurate quantitation.

Different studies based on the accuracy of the numerical integration of peak area state that at least 10 data per σ are required when a Gaussian peak is digitized [1–3]. Some authors claim that 6–7 points per peak, or one data per σ is sufficient for signal analysis [4], which latter density is, of course, in accordance with the Nyquist frequency, but such a low sampling rate is obviously improper when dealing with real signals where baseline noise is omnipresent [5,6].

The influence of data acquisition rate is often revisited as faster detectors become available. The need for high-frequency data acquisition is sometimes illustrated using fast detectors intentionally set to a low-frequency acquisition rate. In those instances the authors of scientific papers [7] or instrumentation brochures [8] claim that peak shapes become distorted and broad when the signal is sampled with a less-than-optimum sampling rate.

The aim of this study is to shed light on this misconception that accompanies data acquisition.

* Corresponding author at: Department of Analytical and Environmental Chemistry and Szentágothai Research Center, University of Pécs, Ifjúság útja 6, H-7624 Pécs, Hungary. Tel.: +36 72 501500x24582; fax: +36 72 501518.
E-mail address: felinger@ttk.pte.hu (A. Felinger).

2. Materials and methods

2.1. Sampling theorem

When merely the theory of digital signals is considered, the Nyquist or Shannon sampling theorem suggests that at least *two samples* for the *highest frequency* present in the continuous signal must be collected [5]:

$$\Delta t \leq \frac{1}{2f_{max}} \tag{1}$$

where Δt is the sampling time[1] and f_{max} is the highest frequency to be preserved after digitalization. Obviously, in order to accurately determine the location of peak maxima or the peak area, a higher sampling frequency is required than it is established by Eq. (1).

When the signal is recorded over a time period T, altogether $n_p = T/\Delta t$ data points are acquired. The frequency resolution of the digitized signal will be $\Delta f = 1/T$.

2.2. Trigonometric interpolation

The sampling rate and the total acquisition time of a digitized signal determine both the maximum observable frequency and the resolution of the Fourier spectrum. The maximum frequency f_{max} is inversely proportional to the sampling time Δt (see Eq. (1)).

Low sampling rate will set a constraint the maximum observable frequency. The undersampled signal can be Fourier transformed, and the length of the Fourier transformed signal is increased by adding zeros after the last point, formally zero-intensity contribution of frequencies higher than f_{max} is added. When $(k-1)n_p$ zeros are added after f_{max} to extend the n_p-point Fourier spectrum, the maximum frequency in the spectrum becomes:

$$f'_{max} = \frac{kn_p}{2T} = \frac{k}{2\Delta t} = kf_{max} \tag{2}$$

Therefore, when the zero-filled Fourier spectrum is inverse transformed, the data density will be k-fold compared to the original sampled signal. Thus, the above procedure is an interpolation of sampled signals. Interpolation by zero filling can be used to improve the resolution of insufficiently sampled signals [9–11]. It must be emphasized, however, that interpolation may help to isolate maxima or minima, but it is not a real resolution enhancement. When resolution is insufficient due to undersampling, although the maximum observable frequency f_{max} is increased to f'_{max} by adding zeros, but all data within $[f_{max}, \ldots, f'_{max}]$ are zeros. Therefore, that frequency region contains no additional new information at all.

2.3. Chemicals and reagents

Toluene (99.5%) was purchased from Sigma–Aldrich (St. Louis, MO, USA). Ethylbenzene (>99%) was from Fluka (Buchs, Switzerland). Methanol (Chromasolv for gradient elution) and water (Chromasolv Plus for HPLC) were obtained from Sigma–Aldrich (St. Louis, MO, USA).

2.4. Instrumentation and conditions

The analyses were performed on an Agilent 1100/1200 Series HPLC and a Shimadzu Prominence UFLC XR systems.

The Agilent HPLC system consisted of a micro vacuum degasser (Agilent 1100, Agilent, Palo Alto, CA, USA), binary pumps (Agilent 1100), thermostated column compartment (Agilent 1200),

Diode Array detector (Agilent 1100) at 261 nm, and an autosampler (Agilent 1100). The Agilent ChemStation (B.01.03(204) version) software was applied on the HPLC system.

The Shimadzu UFLC XR system consisted of a liquid chromatograph (Prominence Liquid Chromatograph LC-20 ADXR, Shimadzu), a DGU-20 A3 micro vacuum degasser, an SIL-20 ACXR auto sampler, an SPD-M20A diode array detector, a CTO-20 AC column oven, and a CBM-20 controller. LabSolutions (Shimadzu) software was used to control the UFLC system and for data processing.

The same Ascentis Express C$_{18}$ column (50×2.1 mm, 2.7 μm, Supelco, Bellefonte USA) was used with both instruments. For the isocratic separations, the mobile phase was methanol:water = 80:20 (v/v %). Sample concentration was 0.13 (v/v%) of toluene and ethylbenzene in the solvent of the mobile phase.

Operating conditions were as follows: flow rate 0.5 mL min^{-1}, column temperature was 25 °C and injection volume of the standards solvent was 1 μL (three replicate injections). The data acquisition frequency of the diode array detector was changed from 0.3 to 80 Hz on the Agilent instrument and the optical (cell) slit widths were 1 and 4 nm. The maximum data acquisition frequency in the Shimadzu instrument was 40 Hz.

3. Results and discussion

3.1. Misleading experimental data

In order to reproduce the effect of sampling frequency on peak width, first we carried out the separation of a toluene–ethylbenzene mixture on an Ascentis Express C$_{18}$ (50×2.1 mm, 2.7 μm) column using an Agilent 1100 instrument. The mobile phase was MeOH:H$_2$O = 80:20 (v/v%). The chromatograms recorded at 80 Hz and 1.25 Hz sampling rates are plotted in Fig. 1.

When the sampling rate is 80 Hz, the number of theoretical plates is $N = 1078$ for toluene and it is $N = 1344$ for ethylbenzene. The resolution factor for this separation is $R_s = 1.60$. Note that we did not attempt to optimize the separation efficiency and the extra-column band broadening caused by the instrumentation is significant. Albeit this fast separation is a good demonstration
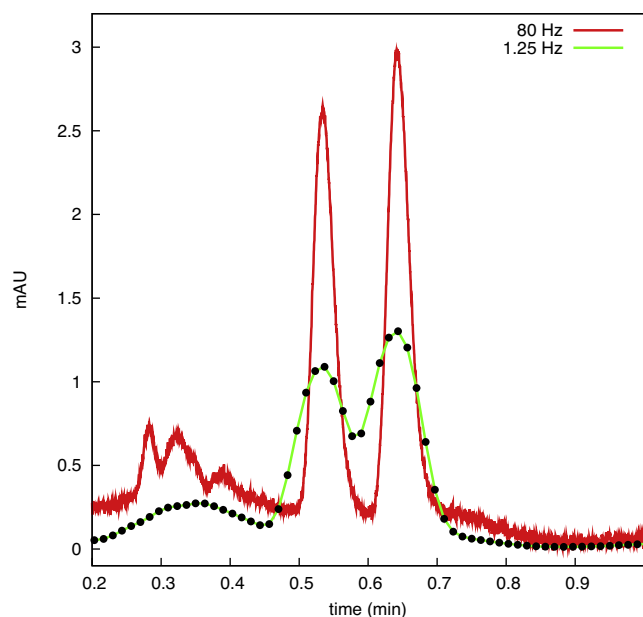

Fig. 1. Chromatograms recorded on the Agilent 1100 instrument at 80 Hz and at 1.25 Hz sampling rate.

---

[1] The data acquisition or sampling rate (frequency) is defined as $1/\Delta t$.