Supplemental Information for "Low Resolution Raman: The impact of spectral resolution on limit of detection and imaging speed in hyperspectral imaging"

Xianli Wang,*[a], Chuanzhen Hu,*[a], Kaiqin Chu[a,b], Zachary J. Smith[a]

[a.] Key Laboratory of Precision Scientific Instrumentation of Anhui Higher Education Institutes, Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei, Anhui, China.

[b.] Hefei National Laboratory for physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui, China.

Table of contents

1. Theory of Low Resolution Raman

1.1 SPSF(f,y,λ) calculation

The SPSF(f,y,λ) is related to the focal length (f) of Lens 5 in Fig. 1A, wavelength (λ), and distribution along the slit (y). We measured spectra of a neon light source (4608, Edmund Optics) whose spectral peak width is below the resolution limit of all focal lengths considered. At each focal length, we firstly selected three peaks circled by the red dashed rectangle in Fig. S1, and then fitted them by a 2-D Gaussian distribution to obtain the SPSF(f,y,λ) corresponding to each peak (f = 80mm, 100mm, 150mm, 200mm, 250mm, 300mm, 400mm). We then used the average of these three values, SPSF(f,y,λ), to represent the spectral point spread function at each lens focal length, representing the spectral resolution. After calculation, the spectral resolutions of the system equipped with these lenses are 52.7cm^{-1}, 39.3cm^{-1}, 31.2cm^{-1}, 17.2cm^{-1}, 15.1cm^{-1}, 14.4cm^{-1}, and 11.5cm^{-1}, respectively.

1.2 Theoretical simulation

The flow chart of simulating spectral intensity at different resolutions is shown in Fig. S2A. As an example, we firstly consider a 2-D Raman peak W(y,λ) shown in Fig. S2B that in the wavenumber domain obeys a Lorentzian distribution with a 30cm^{-1} peak width, while in the spatial domain it obeys Gaussian distribution and peak width is narrow setting the standard deviation, σ, to 0.1 pixels and constant versus wavenumber. Secondly, we measured spectra of a neon light source, shown in Fig. S1, whose peak width is narrow compared to the spectral resolution at all focal lengths, and then we could obtain corresponding spectral point spread functions, SPSF(f,y,λ) via the method described in section 1.1. Fig. S2C shows the experimental SPSF(400,y,λ). Each peak on the detector should be the convolution of the ideal peak W(y,λ) and the SPSF(f,y,λ). Fig. S2D shows the result of the convolution of W(y,λ) and SPSF(400,y,λ). As we are interested in the impact of spectral resolution, we binned the 2-D blurred peak along the spatial direction (y) to get a 1-D peak distribution, $\hat{W}$(f,λ), as shown in Fig. S2E, as described in Eq. 1 in the main text. For a 30cm^{-1} width input peak, if we ignore the SPSF, then we would get the ideal peak shown by the red dashed

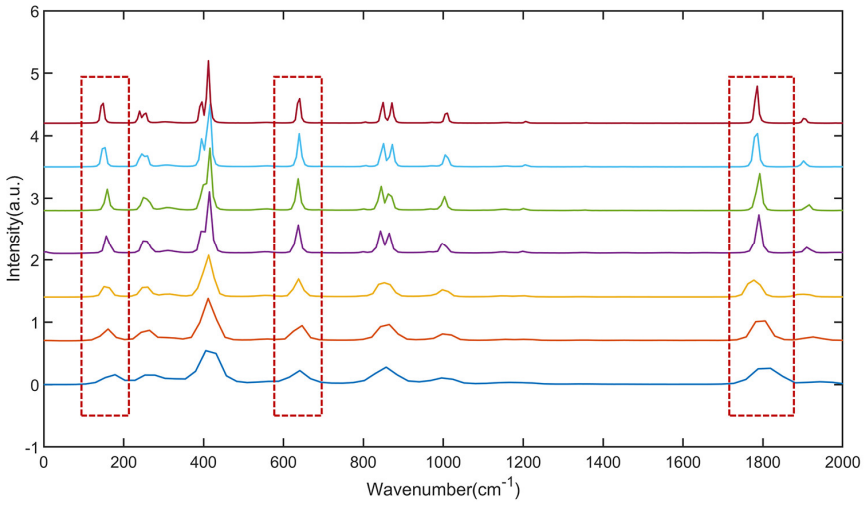

Fig. S1 The spectra of Neon light source with different spectral resolutions. The focal length corresponding to each spectrum from bottom to top is 80mm, 100mm, 150mm, 200mm, 250mm, 300mm, 400mm. Those peaks cirlced by red dashed rectangle are used to calculate spectral point spread function.

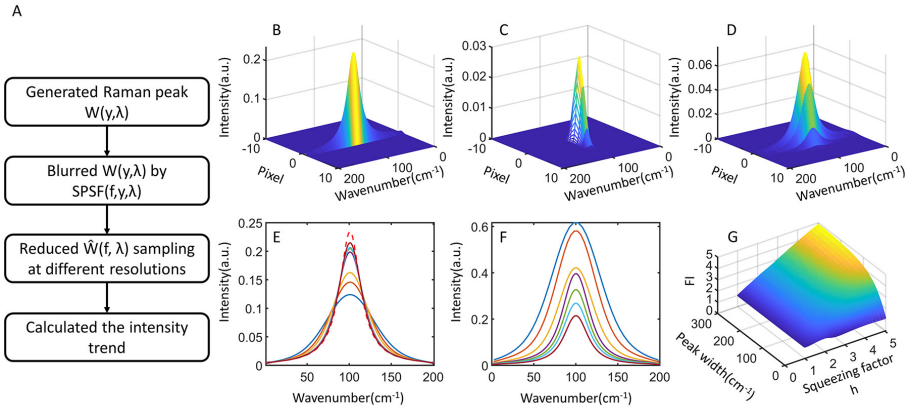

Fig. S2 Theoretical calcluation process. (A). Flow chart of Low Resolution Raman theory. (B). 2-D distribution of ideal Raman with peak width 30 cm^{-1} and narrow spatial width. (C) 2-D distribution of spectral point spread function. (D) 2-D distribution of blurred Raman peak, which is the convolution of ideal Raman peak and spectral point spread function. (E) 1-D intensity distributions along wavenumber direction acquired by summing peak along the spatial direction. The red dash line represents the unblurred spectrum. (F) 1-D intensity distributions after resampling those peaks in (E) according to the detector pixellation. (G) Theoretical intensity FI trend versus peak width and squeezing factor.

line on the dector. After blurring by the SPSF, we can clearly see all maximum intensity values are weaker than those of the ideal peak and all peak widths are broder than those of the ideal peak. However, while broadening the SPSF leads to a decrease in the peak value of a finely sampled peak, in the real spectrometer, broadening the SPSF also leads to decreased sampling of the spectrum, a decrease in the number of pixels spanned by those blurred peaks, and thus an enhancement of the peak intensity. To compare the peak intensity at different spectral resolutions, we took the peak intensity corresponding to the highest resolution, 11.5cm^{-1}, as reference peak intensity and defined a “squeezing factor”, h, as

described in the main text Eq. 2. All peaks are multiplied by the corresponding squeezing factor to represent the actual peaks with different spectral resolution as shown in Fig. S2F. Finally, we took the maximum value corresponding to the highest resolution as reference value and divided the maximum peak intensity corrseponding to each spectral resolution by the reference value to obtain the intensity increase trend. Performing this process for a variety of peaks with varying peak widths yields the surface shown in Fig. S2G. This theoretical model is validated in Section 3.1 of the main text.

2. LOD Calculation

Using the theory laid out in Section 1, we can compute the intensity of a given peak at different spectral resolutions. Firstly, we fitted the peak indicated by the black line in Fig. 3A by a Gaussian to obtain the peak width and peak height at the highest concentration (1.234M) and spectral resolution (11.5cm^{-1}) as a reference. The intensities at other resolutions can be theoretically calculated according to the FI calculated via Eq. 3 of the main text. The peak width and peak height obtained from the fitting result shown in Fig. S3A are 48cm^{-1} and 20.759, respectively. Fitting the theoretical data points with a line passing through the origin (assuming the ideal case of 0 signal at 0 concentration, fits shown as dashed lines in Fig. S3B), we obtain a slope and intercept that can be used in Eq. 10 of the main text. The resulting theoretical LOD values are shown in Table S1.

■ **Table S1** Theoretical LOD of each spectral resolution

Spectral resolution(cm^{-1})	FI	Peak intensity	LOD(M)
39.3	2.90	60.201	0.2278
17.2	1.86	38.612	0.3557
14.4	1.30	26.987	0.5074
11.5	1.00	20.759	0.6603

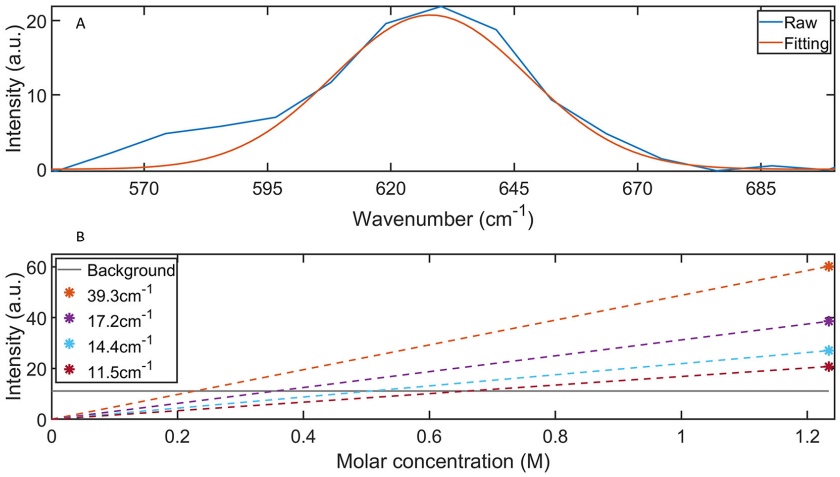

Fig. S3 Theoretical LOD calculation process. (A). Gaussian fitting result for the selected peak at highest spectral resolution, 11.5cm^{-1}, and highest concentration, 1.234M. (B) Theoretical peak intensity versus fructose concentration.

3. Endmembers of blind Raman image unmixing

Fig. S4A, B, and C represent the k-means endmember spectra from “fully-resolved”, “Narrowband low resolution”, and “Broadband low resolution” datasets from PS and PMMA beads, shown in Fig. 4 of the main text. Fig S4 A2, B2, and C2 represent corresponding pseudo-colored abundance maps of the PS (red) and PMMA (blue) endmembers. Note that despite very poor spectral resolution, clear differences between

PS and PMMA are observed in the low resolution datasets. Fig. S5 represents the k-means endmembers based on Fig. 5 of the main text. Colors correspond to cell wall (blue), nucleic acids (red), and cytoplasm/cell interior (yellow). The background is not shown for clarity. Finally, Fig. S6 shows blind

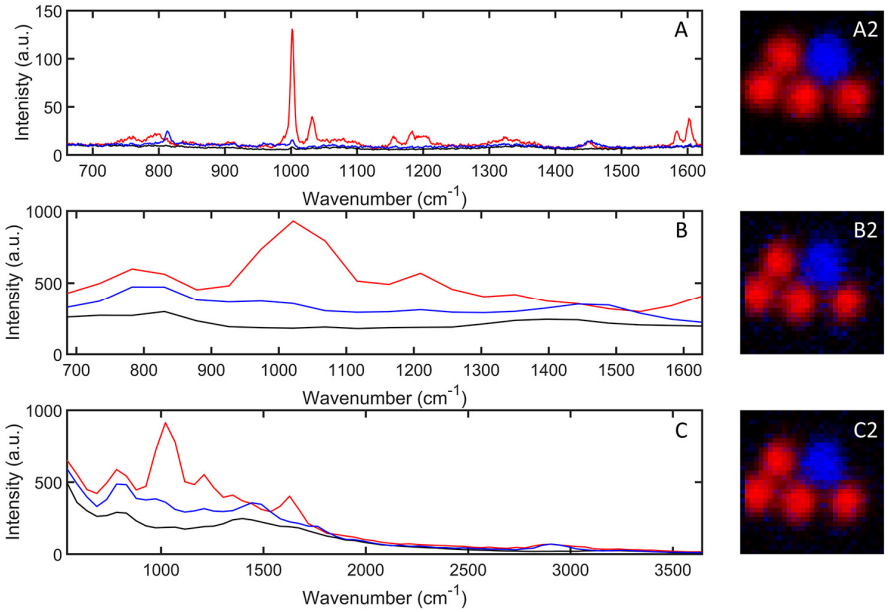

Fig. S4 Endmembers of PS/PMMA images. (A). Fully resolved spectra obtained by K-means. (A2). Pseudo color image corresponding to (A). (B). Narrowband low resolution spectra obtained by K-means. (B2). Pseudo color image corresponding to (B). (C). Broadband low resolution spectra obtained by K-means. (C2). Pseudo color image corresponding to (C)

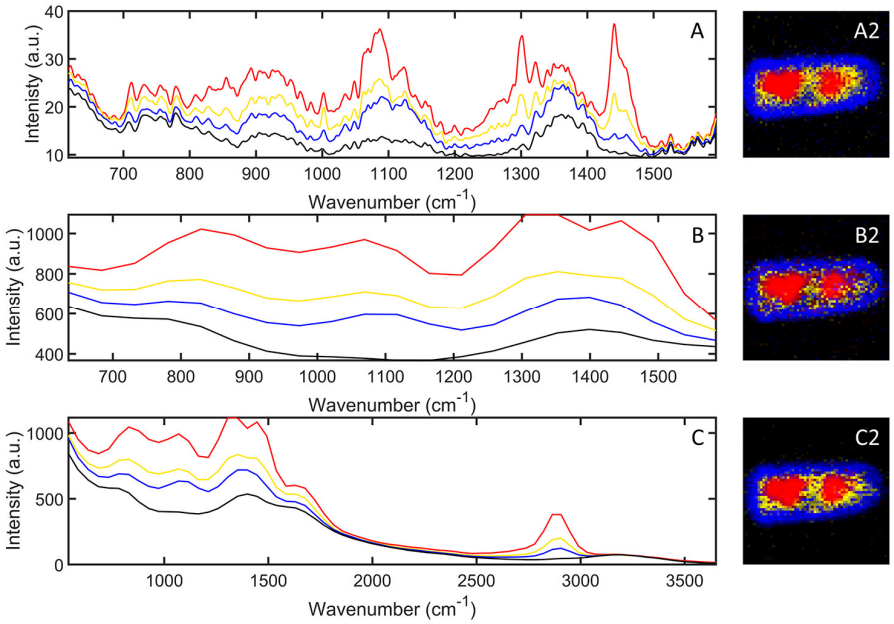

Fig. S5 Endmembers of yeast cell images. (A). Fully resolved spectra obtained by K-means. (A2). Pseudo color image corresponding to (A). (B). Narrowband low resolution spectra obtained by K-means. (B2). Pseudo color image corresponding to (B). (C). Broadband low resolution spectra obtained by K-means. (C2). Pseudo color image corresponding to (C)

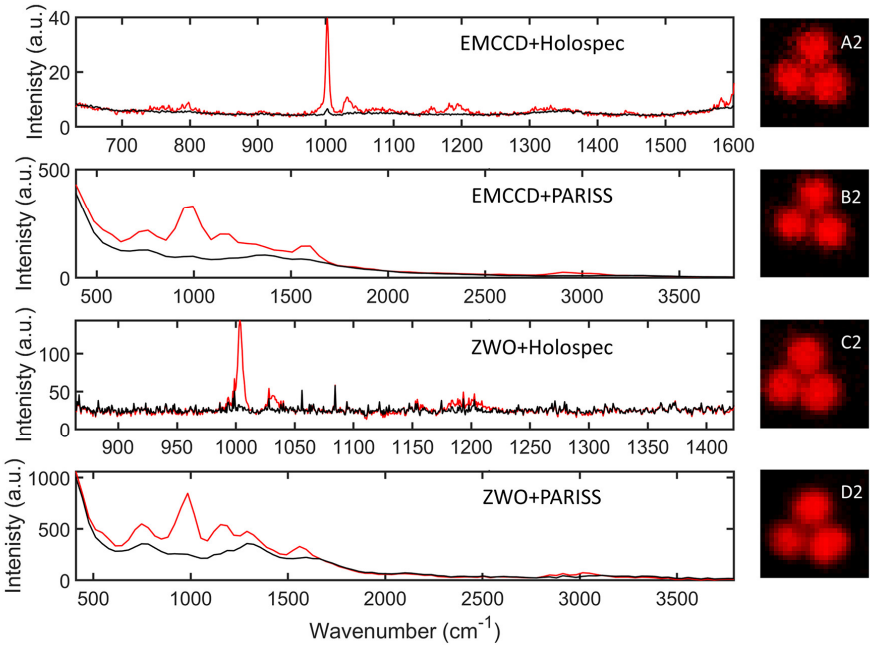

Fig. S6 Endmembers of PS images using different cameras and spectrometers. (A). Fully resolved spectra obtained by K-means using an EMCCD. (A2). Pseudo color image corresponding to (A). (B). Broadband low resolution spectra obtained by K-means using an EMCCD. (B2). Pseudo color image corresponding to (B). (C). Fully resolved spectra obtained by K-means using a low-cost CMOS camera. (C2). Pseudo color image corresponding to (C). (D). Broadband low resolution spectra obtained by K-means using a low-cost CMOS camera. (D2). Pseudo color image corresponding to (D).

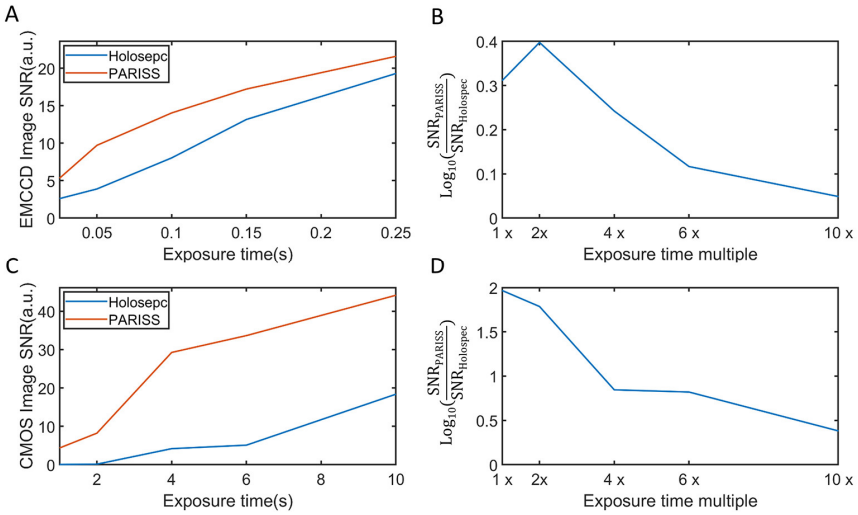

Fig. S7 Image SNR of polystyrene beads for different detectors, spectrometers, and exposure times. (A) Performance of the EMCCD with low resolution (PARISS) and fully resolved (Holospec) spectrometers. (B) The ratio of the SNRs in (A) versus exposure time. (C) Performance of the CMOS with low resolution and fully resolved spectrometers. (D) The ratio of SNRs in (C) versus exposure time. Note that at the shortest exposure times, both the EMCCD+PARISS and CMOS+PARISS have an image SNR of approximately 5.

unmixing results for a sample of PS beads shown in Fig. 6 of main text, with varying combinations of detector and spectrometer.

The raw image SNRs used to calculate the ratios in Fig. 6C are shown in Fig. S7A and S7C. From here it is clear that in the case of the EMCCD, the Holospec and PARISS have more similar SNRs, indicating that the performance improvement provided by low- resolution Raman is more modest. In the CMOS case, the SNR values are more dissimilar, indicating that the performance improvement is greater. The ratio of the SNRs, shown in Fig. S7B and S7D, quantify this performance improvement. As the SNR ratios can be quite different, Fig. 6C (and Fig. S7B and S7D) plots these curves using a log scale. Note additionally that the exposure times for the EMCCD and CMOS detectors were chosen such that their image SNR at the shortest exposure time when using the PARISS detector were similar (approximately 5), to give the two datasets a similar starting point.

4. Modeling of error vs. exposure time of multicomponent mixtures

4.1 Relationship between exposure time, spectral coefficient and prediction standard deviation

While Eq. 3 in the main text predicts how much the intensity of a given peak would be enhanced by reducing the spectral resolution, another question is how the accuracy in obtaining chemical concentrations using a linear model varies with spectral resolution. A typical linear model is given below:

$$\hat{\mathbf{X}} = z\mathbf{S}\mathbf{C} + \boldsymbol{\varepsilon} \tag{S4-1}$$

$\hat{\mathbf{X}}$ is an Nx1 vector representing the measured spectra. **C** is a Jx1 vector representing concentration of each pure component, **S** is an NxJ matrix representing spectra of pure components and **ε** represents Gaussian noise whose mean value is zero and standard deviation is denoted by σ. z is the vector of scaling factors between experimental measurements of $\hat{\mathbf{X}}$ and **S** (accounting for differences such as exposure time, measurement volume, etc.) The concentration vector **C** is typically of interest. In this section we derive how spectral resolution impacts the accuracy of concentration prediction. The least squares solution to Equation (S4-1) is given below in Equation (S4-2):

$$\hat{\mathbf{C}} = z\mathbf{S}^{\mathrm{T}}(\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}\hat{\mathbf{X}} \tag{S4-2}$$

The prediction error can be obtained by calculating the difference of the predicted concentration, $\hat{\mathbf{C}}$, and the real concentration, **C**. The mean value and standard deviation of prediction error can be expressed as Equations (S4-3) and (S4-4), respectively.

$$\mathrm{E}(\hat{\mathbf{C}} - \mathbf{C}) = \mathbf{0} \tag{S4-3}$$

$$\mathrm{std}(\hat{\mathbf{C}} - \mathbf{C}) = (z\sigma)\sqrt{(\mathbf{S}^{\mathbf{T}}\mathbf{S})^{-1}} \tag{S4-4}$$

In the absence of systematic bias, the mean value of the prediction error always converges to 0, but the standard deviation of the error, which determines the expected measurement accuracy, is related to the covariance of the pure component matrix (how similar are the two components), noise level σ and scaling factor z. Because all mixtures are binary in our experiements, we define the quadrature sum of the errors as Σ, which can be expressed as Equation (S4-5),

$$\Sigma = \sqrt{\mathrm{std}(\hat{\mathrm{C}}_1 - \mathrm{C}_1)^2 + \mathrm{std}(\hat{\mathrm{C}}_2 - \mathrm{C}_2)^2} = \frac{\sigma z}{\mathrm{SC}} \tag{S4-5}$$

where SC, “spectral coefficent”, is a parameter whose definition is shown in Equation (S4-6).

$$\mathrm{SC} = \frac{\sqrt{(\bar{\mathrm{S}}_1{}^{\mathrm{T}}\bar{\mathrm{S}}_1) * (\bar{\mathrm{S}}_2{}^{\mathrm{T}}\bar{\mathrm{S}}_2) - (\bar{\mathrm{S}}_1{}^{\mathrm{T}}\bar{\mathrm{S}}_2)^2}}{\sqrt{(\bar{\mathrm{S}}_2{}^{\mathrm{T}}\bar{\mathrm{S}}_2) + (\bar{\mathrm{S}}_1{}^{\mathrm{T}}\bar{\mathrm{S}}_1)}} \tag{S4-6}$$

SC depends on both the spectral intensity (which affects the vector norms), as well as the similarity between the two spectra (which affects the cross-term $S_1{}^{T}S_2$). So, as the spectral resolution decreases, the

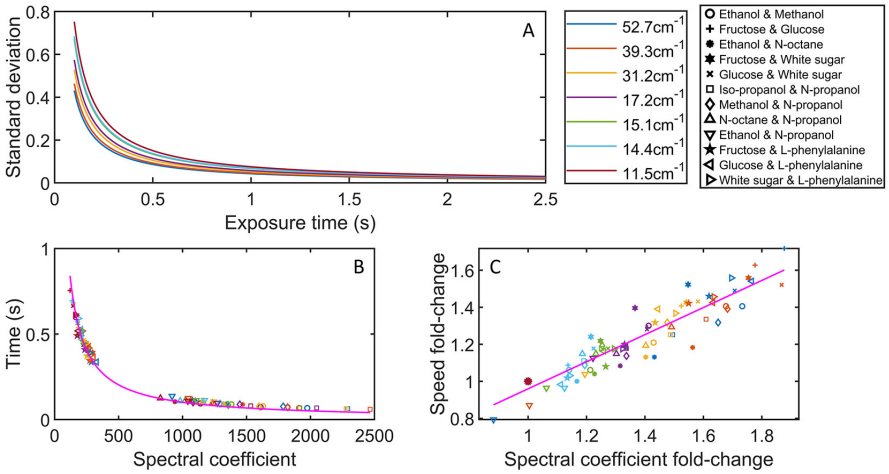

Fig. S8 (A) The fitting results between standard deviation of prediction error and exposure time for various spectral resolutions for a binary mixture of fructose and glucose. Resolution indicated by color as shown in the legend. (B) Spectral coefficient vs. time to reach a fixed standard deviation of prediction error for a several binary mixtures (indicated by shape, as shown in the legend) and spectral resolutions (indicated by color). (C) The relationship between spectral coefficient fold-change and speed fold-change is linear.

improvement in SNR due to increased intensity, and information loss due to increasing similarity between the two components, compete with each other. In the case of spectra with broad peaks and little similarity, reducing spectral resolution could be expected to favourably impact Σ. However, for spectra with sharp peaks and high overlap, Σ could conceivably degrade with reducing spectral resolution. Thus, the impact of spectral resolution on quantitative spectroscopy is highly case-dependent, but reducing spectral resolution yields generally mildly positive results, as shown below.

4.2 Experimental validation

In the last section, we derived that the dominant factors determining the analytical error are the spectral SNR, and the spectral overlap. This suggests two ways to improve standard deviation: either prolong the exposure time or reduce the resolution to improve the measurement SNR. The first solution is undesirable in many contexts due to finite limits on acceptable exposure times. Thus, in the following section we explore how much benefit reducing the system resolution can have on analytical spectroscopy. We used the system shown in Fig. 1A to acquire high SNR pure component spectra and spectra of mixtures of these pure components with different spectral resolutions and exposure times. For each mixture, we acquired 50 spectra at each spectral resolution and exposure time.

After predicting the concentration for each replicate of each spectral mixture at each exposure time and spectral resolution, we constructed the analysis presented in Fig. S8A, which shows the standard deviation of the errors versus exposure time and spectral resolution for the example case of a binary mixture of fructose and glucose. For clarity, Fig. S8A presents fits to the raw data. The fitting performance can be found in Fig. S9.

It's clear from Fig. S8A that to reach same given standard deviation, the required exposure time for low resolution Raman spectroscopy is shorter than that of high-resolution Raman spectroscopy. Further, the time required to reach a given standard deviation should have an inverse relationship with SC according to Equation (S4-5). If we set the given standard deviation to 0.1, then determine the required exposure times to reach that standard deviation from Fig. S8A, we can plot this time versus the SC for each measurement, as shown in Fig. S8B. From this figure it is obvious that large values of SC, indicating high SNR and low spectral similarity between the two binary components is required to reach a low error in a reasonable exposure time.

Taking all of these analyses together, we can conclude that the potential improvement in acquisition speed to reach a certain analytical error depends critically on the specific spectra of the mixture components. To analyse this, we normalize all measurements to the measurements made using the highest spectral resolution. Then, we can compute the fold-change in SC and the fold-change in time to

reach an error standard deviation of 0.1. The relationship between fold change in speed and fold change in SC is shown in Fig. S8C. In this plot spectral resolution is denoted by the symbol color (colors as in Fig. S8A), while the symbols denote the specific chemical mixture (see legend). Clearly, reducing the spectral resolution improves measurement speed for a given analytical error for the large majority of compounds. However, these gains are relatively modest. Furthermore, for some combinations, such as ethanol and N-octane (Inverted triangle), the speed improvement is smaller than 1 and its SC fold-increase is smaller than 1. This represents the fact that these chemicals have rather sharp peaks, and rather high degree of spectral similarity, suggesting that reducing spectral resolution is unlikely to provide a substantial boost to SNR, while giving a substantial penalty to the ability to discriminate the two compounds.

Thus, for a given system, if the SC fold-increase is less than 1, spectral information loss dominates, meaning that low resolution Raman spectroscopy is not suitable for this sample, on the contrary, when SC fold-increase is greater than 1, SNR improvement dominates, meaning that low resolution Raman spectroscopy can yield a gain in analytical accuracy at a given exposure time.

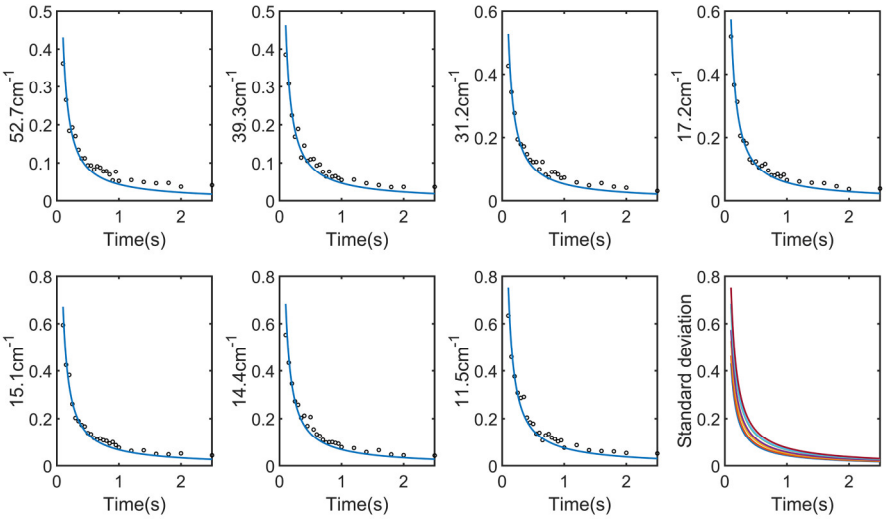

Fig. S9 Fitting performance for fructose and glucose mixture at different spectral resolutions. The sub-figure in the lower right corner shows fitting results of glucose and fructose at all spectral resolutions.