Contrast improvement of terahertz images of thin histopathologic sections

Florian Formanek,* Marc-Aurèle Brun, and Akio Yasuda

Sony Corporation, Life Science Laboratory, Advanced Materials Laboratories, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan
*florian.formanek@jp.sony.com

Abstract: We present terahertz images of 10 μm thick histopathologic sections obtained in reflection geometry with a time-domain spectrometer, and demonstrate improved contrast for sections measured in paraffin with water. Automated segmentation is applied to the complex refractive index data to generate clustered terahertz images distinguishing cancer from healthy tissues. The degree of classification of pixels is then evaluated using registered visible microscope images. Principal component analysis and propagation simulations are employed to investigate the origin and the gain of image contrast.

OCIS codes: (170.3880) Medical and biological imaging; (170-6795) Terahertz imaging; (300.6495) Spectroscopy, terahertz.

References and links

1. R. M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," Phys. Med. Biol. **47**(21), 3853–3863 (2002).
2. R. M. Woodward, V. P. Wallace, R. J. Pye, B. E. Cole, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging of ex vivo basal cell carcinoma," J. Invest. Dermatol. **120**(1), 72–78 (2003).
3. V. P. Wallace, A. J. Fitzgerald, S. Shankar, N. Flanagan, R. Pye, J. Cluff, and D. D. Arnone, "Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo," Br. J. Dermatol. **151**(2), 424–432 (2004).
4. A. J. Fitzgerald, V. P. Wallace, M. Jimenez-Linan, L. Bobrow, R. J. Pye, A. D. Purushotham, and D. D. Arnone, "Terahertz pulsed imaging of human breast tumors," Radiology **239**(2), 533–540 (2006).
5. T. Löffler, T. Bauer, K. J. Siebert, H. G. Roskos, A. Fitzgerald, and S. Czasch, "Terahertz dark-field imaging of biomedical tissue," Opt. Express **9**(12), 616–621 (2001).
6. P. Knobloch, C. Schildknecht, T. Kleine-Ostmann, M. Koch, S. Hoffmann, M. Hofmann, E. Rehberg, M. Sperling, K. Donhuijsen, G. Hein, and K. Pierz, "Medical THz imaging: an investigation of histo-pathological samples," Phys. Med. Biol. **47**(21), 3875–3884 (2002).
7. S. Nakajima, H. Hoshina, M. Yamashita, C. Otani, and N. Miyoshi, "Terahertz imaging diagnostics of cancer tissues with a chemometrics technique," Appl. Phys. Lett. **90**(4), 041102 (2007).
8. E. Berry, J. W. Handley, A. J. Fitzgerald, W. J. Merchant, R. D. Boyle, N. N. Zinov'ev, R. E. Miles, J. M. Chamberlain, and M. A. Smith, "Multispectral classification techniques for terahertz pulsed imaging: an example in histopathology," Med. Eng. Phys. **26**(5), 423–430 (2004).
9. T. Löffler, K. Siebert, S. Czasch, T. Bauer, and H. G. Roskos, "Visualization and classification in biomedical terahertz pulsed imaging," Phys. Med. Biol. **47**(21), 3847–3852 (2002).
10. J. Nishizawa, T. Sasaki, and T. Ken Suto, T. Yamada, T. Tanabe, T. Tanno, T. Sawai, and Y. Miura, "THz imaging of nucleobases and cancerous tissue using a GaP THz-wave generator," Opt. Commun. **244**(1-6), 469–474 (2005).
11. S. M. Kim, F. Hatami, J. S. Harris, A. W. Kurian, J. Ford, D. King, G. Scalari, M. Giovannini, N. Hoyler, J. Faist, and G. Harris, "Biomedical terahertz imaging with a quantum cascade laser," Appl. Phys. Lett. **88**(15), 153903 (2006).
12. G. M. Png, J. W. Choi, B. W.-H. Ng, S. P. Mickan, D. Abbott, and X.-C. Zhang, "The impact of hydration changes in fresh bio-tissue on THz spectroscopic measurements," Phys. Med. Biol. **53**(13), 3501–3517 (2008).
13. H. Hoshina, A. Hayashi, N. Miyoshi, F. Miyamaru, and C. Otani, "Terahertz pulsed imaging of frozen biological tissues," Appl. Phys. Lett. **94**(12), 123901 (2009).
14. M.-A. Brun, F. Formanek, A. Yasuda, M. Sekine, N. Ando, and Y. Eishii, "Terahertz imaging applied to cancer diagnosis," Phys. Med. Biol. **55**(16), 4615–4623 (2010).
15. P. U. Jepsen, U. Møller, and H. Merbold, "Investigation of aqueous alcohol and sugar solutions with reflection terahertz time-domain spectroscopy," Opt. Express **15**(22), 14717–14737 (2007).

16. S. W. Smith, "Linear image processing," in *Digital Signal Processing: a Practical Guide for Engineers and Scientists*, (Newnes, 2003), pp. 397–422.
17. B. Balasko, J. Abonyi, and B. Feil, Fuzzy clustering and data analysis toolbox (2005). *http://www.fmt.vein.hu/softcomp/fclusttoolbox/*
18. J. R. Fletcher, G. P. Swift, D. C. Dai, J. M. Chamberlain, and P. C. Upadhya, "Pulsed terahertz signal reconstruction," J. Appl. Phys. **102**(11), 113105 (2007).

1. Introduction

In recent years several groups have demonstrated the application of terahertz (THz) imaging for medical studies, especially for diagnosing cancers. Due to the strong absorption of terahertz radiation by water present in biological tissues, these studies were performed either in reflection geometry [1–4] using bulk raw samples or in transmission geometry using paraffin-embedded [5–9], dried [10,11], lyophilized [12] or frozen [13] millimeter thick slices. THz imaging has potential as a tool to aid in the diagnosis of cancers if used in combination with histopathologic examination. However, as they contain information in far more depth than visible photomicrographs, THz images of thick samples are not representative of the structures visible in histology sections. Moreover, the reproducibility of results is often hindered by sample preparation, including control of moisture levels and contact conditions with the imaging window.

In this paper, we show THz spectroscopic images of 10 µm thick sections cut from, formalin fixed, paraffin-embedded liver tissue blocks prepared according to standard histopathologic procedures. Such samples have the advantage of being available from the archive of any pathology department, requiring a minimum quantity of tissue and allowing direct registration with visible microscope images for diagnosis [14]. However, despite the high sensitivity of THz time-domain spectroscopy (TDS), important noise background associated with the measurements of very thin samples (in this case $\lambda/30$ at 1 THz) often leads to a poor signal-to-noise ratio (SNR). Besides the obvious system development aspect, advanced visualization and classification techniques [6–8] can be used to improve image rendering and extract information of interest. Here, we demonstrate that THz images of very thin sections measured in paraffin with water are improved compared to THz images obtained with raw (dewaxed) sections in water or with paraffin sections in air. The origin and the gain of image contrast are investigated using principal component analysis and Fresnel-based simulations.

2. Methods and experimental results

The experiments are performed with a homemade time-domain spectroscopy system working in reflection geometry (see Fig. 1). The system uses a Ti:Sapphire laser emitting femtosecond pulses at a 80 MHz repetition rate to gate photoconductive antennas made of low temperature grown gallium arsenide. The 10 µm thick histopathologic sections are mounted on 3 mm thick Z-cut quartz windows with $\lambda/4$ surface accuracy at 633 nm and parallel to less than 5 arcsec. Good surface flatness and parallelism for the imaging window is critical to achieve accurate determination of the refractive index for such very thin samples. The THz beam is focused onto the top surface of the window with an angle of incidence of 20° using off-axis parabolic mirrors. A lock-in amplifier records the current generated at the detector while mechanical stages raster scan the sample holder assembly. The system is enclosed in a purge box filled with dry nitrogen in order to reduce absorption of the THz beam by water vapor in the air. A CCD camera is used to acquire visible images for precise registration of microphotographs with the THz images. The sections are imaged at room temperature, directly in paraffin or dewaxed using xylene and ethanol solutions. Measurements in water are performed using a 500 µm thick spacer and a top window to prevent evaporation. Because of possible denaturation, raw tissues are not measured in air.
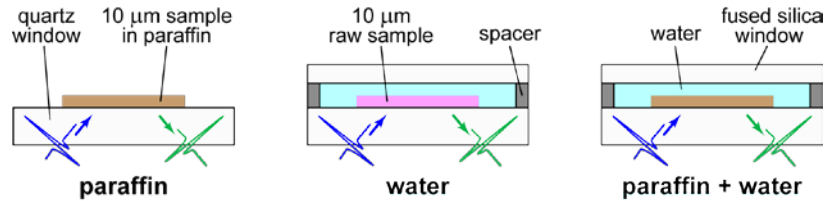
Fig. 1. Schematic diagram of the three measurement geometries.

Time-domain spectra consisting of 1024 points recorded over a 40 ps delay range with a 300 µs lock-in time-constant are Fourier transformed to reveal the frequency domain information. Images are formed by measuring a time-domain spectrum at each sample position. The total acquisition time is about 6 hours for an image consisting of 256 by 256 pixels. The frequency-dependent complex refractive index at each pixel is extracted using a method similar to the one described by Jepsen [15]. The calculation considers the influence of the incidence angle and the polarization of the THz beam. The error induced by the beam shift between reference and sample data is corrected using a calibration factor. Multiple reflections within the sample are also considered. A fixed sample thickness of 10 µm is assumed for the calculation. The averaged refractive indices of the tissues, calculated for each geometry, are shown in Fig. 2. From these graphs, the following values are extracted and used in the simulation presented in the next section: $n_{paraffin} = 1.545$, $n_{paraffin + water} = 1.75 - 0.175i$, $n_{water} = n \times 0.8$, where n is the theoretical index of water.
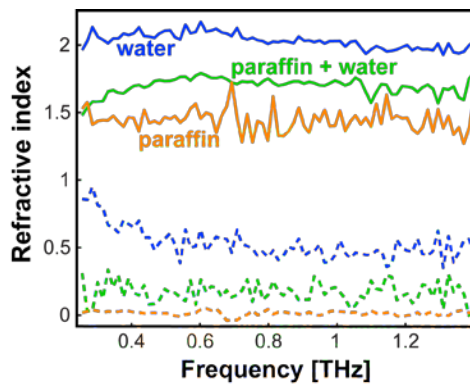

Fig. 2. Graph showing the refractive index (solid lines) and the extinction coefficient (dashed lines) of the healthy tissues calculated from the measurements performed on sample 1 in the three different geometries (water, paraffin + water, paraffin).

Each pixel of the image is represented by its coordinates in the multidimensional space defined by the real and imaginary parts of the refractive index. A circular averaging image filter (also called pillbox filter) with a radius equal to 1 is applied at each frequency image slice to reduce the noise [16]. Principal component analysis (PCA) is employed over the 0.8-1.2 THz range to reduce the dimensionality of the data into a small number of uncorrelated variables and provide an insight into the origin of contrast. For the all the images presented in this paper, the summed variance for the first three principal components accounts for more than 80% of the variance over pixels. The Gustafson-Kessel clustering algorithm [17], based on the Mahalanobis norm, is applied on the set of data consisting of the pixels (observations) and the first three principal values (coordinates). The result of the clustering is set according to thresholds, with memberships below 0.3 set to 0 and memberships above 0.7 set to 1. From the membership values, an artificially colored image is formed using a color look-up table.

Figure 3 shows the THz segmented images obtained with the different experimental geometries and the corresponding visible microphotographs used for comparison. Both

samples are part of an excised hepatocellular carcinoma of the liver from two distinct patients. Healthy parts appear in green while the tumors are shown in red. Yellow parts are intermediary and blue parts correspond to water. Unclassified pixels are displayed in black. The measurements reveal that the terahertz images obtained in paraffin do not correlate with the visible images and the patterns seem irrelevant. The observed contrast may come from variations in the sections' thickness or from a slope of the sample holder with respect to the incident THz beam. Measurements in water positively identify cancer and healthy tissues, though the images remain noisy. In the case of paraffin with water, the noise is clearly reduced and the contrast further improved. Especially, the margins of the tissues are more clearly defined and the different structures appear in more homogenous and brilliant colors, indicating an overall better pixel classification.
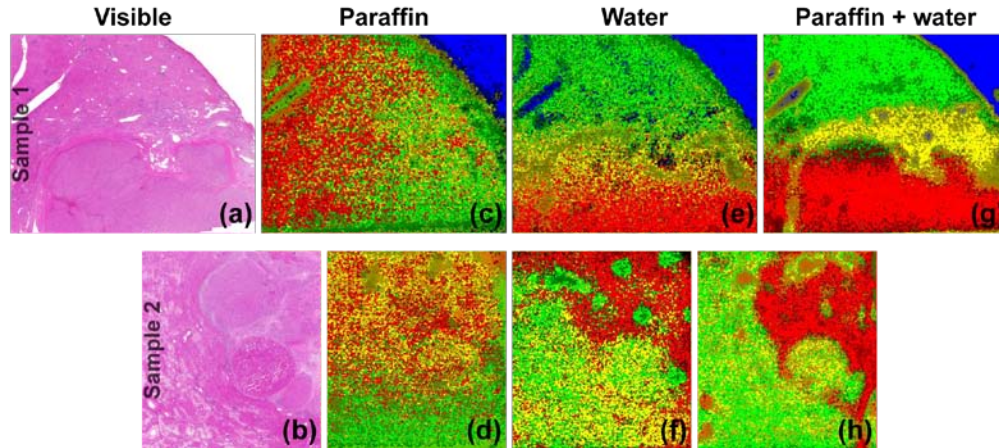

Fig. 3. (a), (b) Registered visible microphotographs of the histopathologic sections stained with hematoxylin and eosin (H&E). (c)-(h) Segmented THz images obtained in the different measurement geometries. Images are 21 x 17 mm^2 (288 x 232 pixels) and 17.4 x 19 mm^2 (228 x 248 pixels) for sample 1 (top row) and sample 2, respectively.

To evaluate each measurement geometry, the THz images are compared with the segmented images in Fig. 4(a) and 4(b) in which the clusters are defined to correspond to the different cellular structures observed at high magnification in the stained sections. Figure 4(c) shows the percentage of pixels correctly assigned to each clusters for each sample. The graph confirms that the best agreement is obtained for the samples measured in paraffin with water, with up to 75% for sample 1. The classification of pixels for the paraffin geometry is only slightly better than the classification that would be obtained with a random distribution of pixels in the different clusters. The reason for this result is that the correct classification obtained on some areas (mainly the upper-right blue zone corresponding to water) is counterbalanced by an a priori meaningless classification over the rest of sample.
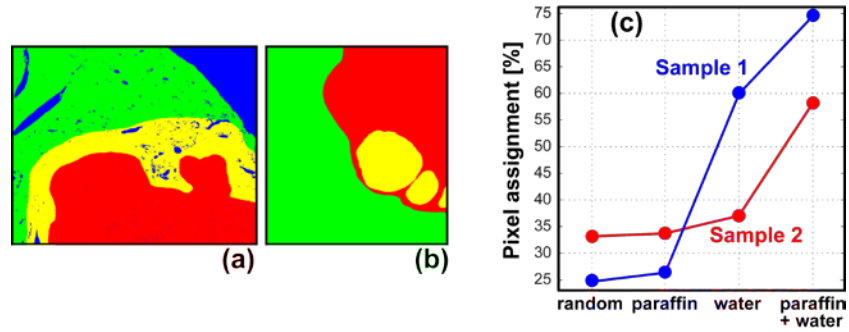
Fig. 4. (a), (b) Images segmented according to the stained sections visible in Fig. 3 (c) and used for comparison with the THz images. Percentage of pixels per THz image correctly assigned to each clusters depending on the measurement geometry.

Figure 5 shows the score plot of the first principal component versus the second principal component obtained from the clusters of Fig. 3(g). Since the blue dots correspond to water and the yellow dots represent tissues containing more holes, the fact that the clusters appear aligned along the second component indicates that the first component reflects the water content or the cell density of the tissues. This plot also suggests that identification of the yellow tissues using only the first component is possible. On the contrary, identification of cancer and healthy tissues cannot be achieved without using the other components.
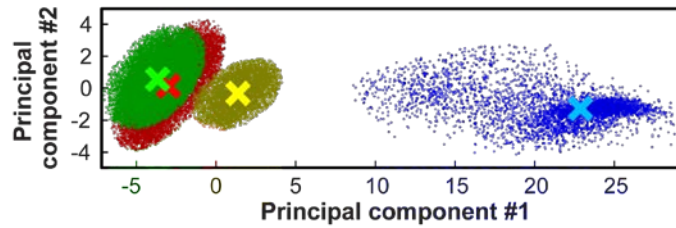

Fig. 5. Score plot of the first principal values versus the second principal values, obtained from the clusters of Fig. 3(g). The crosses indicate the median point within each cluster.

3. Discussion

Propagation simulations are carried out in MATLAB (MathWorks) to investigate the origin of the contrast improvement. The initial incident pulse is defined in the frequency-domain by the following equations [18]:

$$\tilde{F}(\omega) = e^{i\omega t_0} \left[\cos\left(\frac{\omega}{4 f_c}\right) \right]^{p} \left[i\sin\left(\frac{\omega}{2 f_c}\right) \right]^{q} \quad |\omega| \leq 2\pi f_c, \tag{1}$$

$$\tilde{F}(\omega) = 0 \quad |\omega| > 2\pi f_c; \tag{2}$$

where t_0 is the time of arrival of the main part of the pulse and f_c is the cut-off frequency. The integers p and q account for shape factors. The cut-off frequency and the shape factors are chosen so that the spectrum in amplitude of the simulated pulse matches the experimental one. The value of the time of arrival can be chosen arbitrarily. The following values are used for the different parameters: $t_0 = 20$, $f_c = 2.5$, $p = 15$ and $q = 1$. Propagation of the initial pulse is then simulated in the frequency-domain using the reflection, transmission and propagation coefficients corresponding to each geometry. Up to four internal reflections within the thin sample are considered. The refractive index absolute values are defined using the values calculated in the previous section to form the images:

$$n_{cancer}^{paraffin} = 1.545, \quad n_{healthy}^{paraffin} = 1.455,$$

$$n_{cancer}^{paraffin+water} = 1.8025 - 0.1802i, \quad n_{healthy}^{paraffin+water} = 1.6975 - 0.1697i,$$

$$n_{cancer}^{water} = n_{water} * 0.824, \quad n_{healthy}^{water} = n_{water} * 0.7760.$$

The relative difference of +/− 3% in refractive index between cancer and healthy tissues is estimated from the paraffin with water data which corresponds to the measurement giving the most accurate classification. Only the index of water is considered frequency-dependent, as other media have constant indices in the frequency range considered.

Figure 6 shows the initial incident pulse used in the simulation and the resulting reflected pulses on cancer and healthy area calculated for the three different geometries. In the case of paraffin, the pulses reflected over healthy or cancer tissues appear almost identical, especially in amplitude, with only a small time-domain shift in peak position. As we know that the delay stage jitter is the main source of noise for reflection measurements, giving incertitude on the time delay values, the results of the simulation provide an idea as to the origin of the poor signal-to-noise ratio on the paraffin images. On the contrary, the amplitude of the reflected pulses differs in the geometries with water.
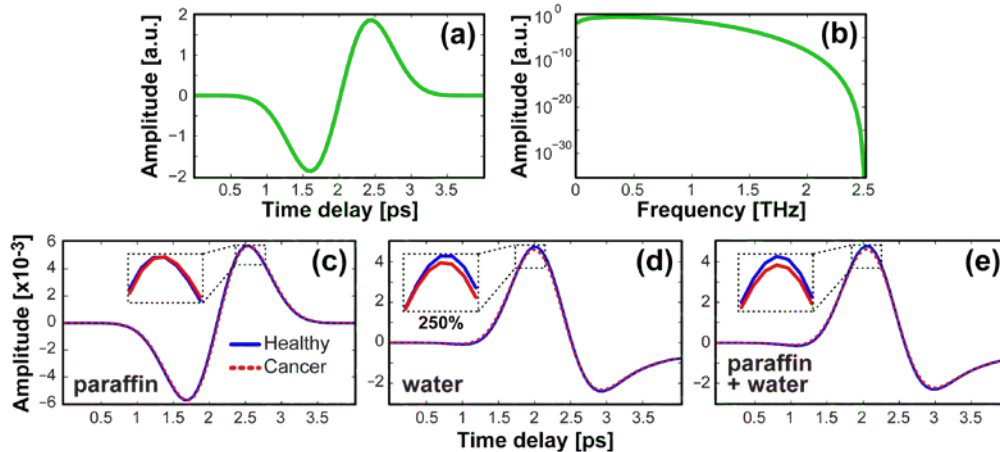

Fig. 6. (a), (b) Initial incident pulse used in the simulation and corresponding frequency-domain spectrum defined to reproduce the experimental usable bandwidth. (c)-(e) Calculated reflected pulses for the different geometries, with the initial time-delay values reset to zero. The insets show 250% magnification close-ups of the pulses' peak.

Figure 7 shows the difference in amplitude between a pulse reflected over healthy tissue and a pulse reflected over cancer tissue, considering different numbers of reflections within the sample. These graphs give insight into the importance of successive reflections on the observed signal. While for the water or the paraffin geometry the difference between reflected pulses is mostly accounted for within the first two reflections within the sample, the contribution of the third reflection cannot be neglected in the paraffin with water case. This means that the THz light effectively interacts with the sample over an additional round trip. Considering that the observed contrast is at least partially due to various scattering properties, this addition return trip may be the main reason to the contrast improvement between water and paraffin with water.
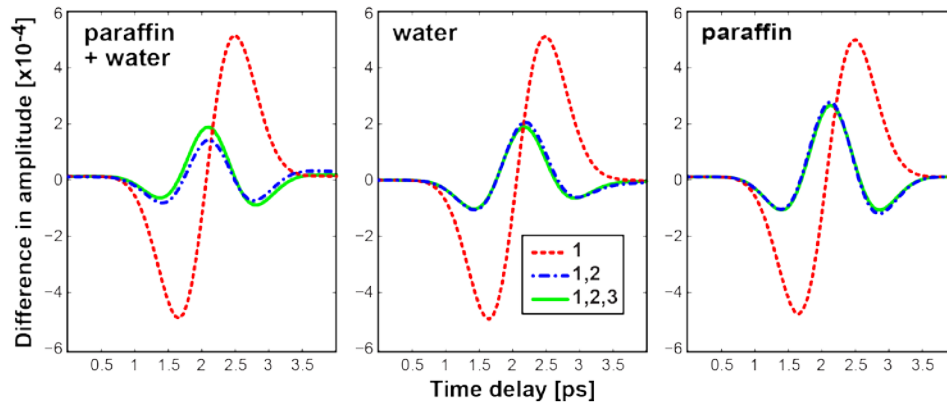
Fig. 7. Calculated spectra of the difference in amplitude between pulses reflected over healthy and cancer tissues, considering different numbers of reflections within a 10 µm thick section.

4. Conclusion

In conclusion, we have showed that segmented terahertz images of 10 μm thick histopathologic sections can be improved by performing the measurements in paraffin with water, rather than in air or with raw tissue in water. Especially, image noise is reduced and the percentage of pixels correctly assigned to each cluster is increased. Since the signal-to-noise ratio has a direct bearing on the speed at which images can be acquired with sufficient contrast, our technique allows reducing the total acquisition time, for example by lowering the lock-in time constant or increasing the scanning step size. With an improved signal-to-noise ratio, weaker denoising filters could also be used to preserve more details. Finally, we believe that even thinner sections can be successfully imaged with this technique.

Acknowledgments