Review in terahertz spectral analysis
Josette El Haddad, Bruno Bousquet, Lionel Canioni, Patrick Mounaix

To cite this version:

HAL Id: hal-00788749
https://hal.archives-ouvertes.fr/hal-00788749
Submitted on 15 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Review in terahertz spectral analysis
Josette El Haddad, Bruno Bousquet, Lionel Canioni, Patrick Mounaix

Abstract: Nowadays terahertz spectroscopy is a well-established technique and recent progresses in technology demonstrated that this new technique is useful for both fundamental research and industrial applications. Varieties of applications such as imaging, non destructive testing, quality control are about to be transferred to industry supported by permanent improvements from basic research. Since chemometrics is today routinely applied to IR spectroscopy, we discuss in this paper the advantages of using chemometrics in the framework of terahertz spectroscopy. Different analytical procedures are illustrates. We conclude that advanced data processing is the key point to validate routine terahertz spectroscopy as a new reliable analytical technique.

Keywords: Terahertz spectroscopy; far infrared spectroscopy; vibrational spectroscopy; analytical chemistry; chemometrics; multivariate calibration,

Josette El Haddad
LOMA, Bordeaux University, CNRS UMR 5798, 351 cours de la Libération, 33405 Talence

Bruno Bousquet
LOMA, Bordeaux University, CNRS UMR 5798, 351 cours de la Libération, 33405 Talence

Lionel Canioni
LOMA, Bordeaux University, CNRS UMR 5798, 351 cours de la Libération, 33405 Talence

P.Mounaix *
LOMA, Bordeaux University, CNRS UMR 5798, 351 cours de la Libération, 33405 Talence

*Corresponding Author
Tel +33 5 4000 2602
Email p.mounaix@loma.u-bordeaux1.fr

1 .Introduction.

Terahertz waves (1 THz = 10¹² Hz) are electromagnetic waves ranging from 0.3 to 10 THz. The related low energy interactions can thus be studied by these waves in a large number of materials and consequently new data can potentially be retrieved to complete the knowledge of material behavior when compared with FIR and Raman spectroscopy. One of the most attractive properties of THz waves is their ability to penetrate inside a wide variety of materials. As a consequence, the terahertz waves can be used to analyze the content through many types of packaging materials, including paper, plastics, leather and wood [1, 2]. Moreover, the interaction with THz waves is non destructive, which allows quality control applications for industry [2-6] and probing biological materials for identification [7] and for medical diagnostics[8].

Ranging between far infrared and microwaves (cf. Fig. 1), THz spectroscopy is a powerful tool for characterizing vibrational modes, such as rotational, torsional, phonon, intramolecular and intermolecular modes [2, 6, 9]. THz spectroscopy is generally considered as being different from conventional far-infrared spectroscopy because terahertz response is linked to collective behavior of molecules in their environment and can distinguish polymorphism and chirality. Moreover, THz spectroscopy doesn’t need the use of the Kramers-Krönig equations, since this technique gives direct access to the electric field amplitude E(t). Consequently, both the amplitude and phase of the electric field in the frequency domain are directly obtained by Fourier transform. THz spectroscopy uses coherent sources and ultrafast sensitive detectors [2, 6, 10-12], which permits improving both the signal-to-noise ratio and the dynamic range.

In this paper, we briefly review the state of the art of the THz spectroscopy techniques with special emphasis on the data processing. We describe the typical setups for THz emission and detection including the major difference between static and transient THz spectroscopies. Then, we discuss the physical interactions between matter and the terahertz field. And finally, we show the potential advantages of using chemometrics in the framework of terahertz spectroscopy.
2. Generation and detection for THz TDS and TRTS

Here is a brief overview of two typical laser-based experiments of THz spectroscopy. The first one is THz time-domain spectroscopy (THz-TDS)[13], and the second one is the time-resolved THz spectroscopy (TRTS)[14, 15]. On the one hand, THz-TDS enables the determination of the complex permittivity of a sample over a frequency range of 0.2-4 THz. Both transmission or reflection setups are built and provide information about the static properties of the sample. On the other hand, TRTS gives access to the dynamic of the photocarriers. Here, the samples are photoexcited with an ultrashort laser pulse (pump) tunable from the UV to the mid-IR and even to the THz range. Then, the time-dependent THz spectrum is extracted with the THz pulse (probe) thanks to a regular pump-probe experiment. One important application of TRTS is the photophysics, and photodynamics of semiconductors [14, 15]. To this purpose, THz generation and detection have been developed from the optical or microwave sides. The emergence of this new technology has carried along a lot of books and publications in which different ways to generate, detect and analyze interaction between terahertz fields and matter are described. Most of the technological difficulties regarding the THz generation and detection have been overcome using ultrafast optical techniques and ultrafast semiconductors since the late 1980’s [6]. There are two major mechanisms for the generation and detection of ultrafast THz pulses. Both involve a nonlinear interaction of the pumping optical electric field with a material, namely THz generation through photon-induced conductivity changes (photoconductivity) in an ultrafast semiconductor. This resonant interaction creates mobile electron-hole pairs which are the origin of the terahertz generation. As the photocarriers are accelerated by a static field applied to the active device, the resulting transient current and charge separation induces a transient polarization in the semiconductor. The second type of mechanism, non-resonant, is optical rectification [11, 12] in crystals (or even air) presenting high electro-optic coefficients. Since the generated polarization does not rely on carrier transport, extremely high bandwidth pulses can be generated, limited only by the duration and bandwidth of the generating optical pulse. TRTS is very close to THz-TDS but requires a third laser beam. Generally, TRTS has been developed with amplified laser sources operating at 1 kHz to take advantage of higher laser pulse energy. It gives access to the dynamical properties of the sample thanks to a variable time delay between an optical pump pulse and the THz probe pulse. In this way, the THz frequency dependent, complex-valued conductivity can be determined as a function of pump-probe delay time. Two types of experiments are then possible. “Pump” scans provide an average transformation of THz absorption as a function of delay time. “Probe” scans give the frequency-dependent permittivity at a given pump-probe delay time. The most complete information is generated by performing a 2D scan of both the pump and probe pulses [14, 15].

3. Physical origin of THz interactions.

3.1 Gas spectroscopy

THz-TDS has been used for identification of molecules and compounds, with great interest for basic science as well as for commercial applications such as stand-off detection of explosives to fight against terrorism. The first demonstration of a THz-TDS system, in 1989, reported the absorption spectrum of water vapor [16]. The main purpose is to analyze the observed absorption profile which is primarily due to the molecular rotational transitions in the vibrational ground or excited state. From the detailed study of individual rotational transitions, it is possible to extract various properties including parameters such as molecular structure and rotational constants. Regarding the performance of the technique, a limit of detection as low as of part-per-trillion (ppt) was recently reported in the context of quantitative analysis of combustion products of a cigarette[17]. It’s worth pointing out that THz spectroscopy is orders of magnitude more sensitive than microwave spectroscopy and presents also a better resolution than infrared spectroscopy because of lower Doppler effects. In the far infrared, the rotational modes are dominant and differ from the ro-vibrational and ro-vibronic modes for which the rotational energy is affected by the vibrational and/or electronic energies. Moreover, quantum mechanics and selection rules apply to the rotational modes revealed by THz spectroscopy and may differ according to the technique used to observe the transition. An important database for the millimetric and sub-millimetric ranges can be found on the website of the Jet Propulsion Laboratory of the NASA (http://spec.jpl.nasa.gov).

3.2 Liquid spectroscopy

From a general point of view, the dielectric properties of liquids in the THz range mainly depend on the interaction and relaxation of dipoles inside the liquid. These dipoles can be either permanent dipoles in polar liquids or induced dipoles in non-polar liquids[18]. Hydrogen bonding can also have a significant effect on the THz response.
Reorientation of dipoles in liquids on a picosecond time scale is essential for both chemical reactions and biological functions [19]. As expected, water and aqueous solutions are the most widely studied liquids mainly for biological applications. Reflection spectroscopy and attenuated total reflection spectroscopy [20] are also exploited. More generally, the dielectric responses of liquids have been extensively studied in the THz range [6] and modeled using two Debye relaxation processes a slow one (few ps) and a fast one (less than fs) depending on the temperature. Moreover, the mixing of water with another liquid (e.g. ethanol) has revealed the existence of a third relaxation process [21] and THz-TDS spectroscopy was successfully applied to the analysis of ionic liquids and ionic liquid-water mixtures [22].

3.3 Solid spectroscopy

THz-TDS has been widely applied to many different molecular solids. THz absorption fingerprints have been recorded for a variety of explosives, poisons, and illicit drugs [23]. THz radiation being also sensitive to the chirality, it potentially offers a way to detect different polymorphs of drug substances, which is of major interest for the pharmaceutical industry. Absorption spectra in the THz range for many common materials can be found in databases such as the Terahertz Database (http://thzdb.org), operated by RIKEN and NICT in Japan. Solid samples are commonly prepared by mixing in a non-absorbing powder such as polyethylene, grinding into a fine powder, and then the resulting composite is finally prepared as pressed pellet. With this method, a fingerprint of the molecules is available even if both the sample uniformity and the scattering effects must be carefully considered. An elegant approach is to deposit the pure sample on the surface of a parallel plate waveguide (PPWG) [24]. It provides longer interaction length of the THz radiation with the sample, which improves the sensitivity of the measure. Assigning resonances in experimental absorption spectra to specific modes typically requires significant simulation and modeling efforts.

Applying THz spectroscopy to solid state physics, a considerable effort was devoted to semiconductor materials [25] because the related time scales perfectly match the characteristics of the TDS-THz. In addition, these materials are of major importance for the generation and the detection of THz fields especially through the understanding and the improvement of the dynamics of carriers. Moreover, optical pump and terahertz probe, with variable temperature experiments allowed new insight of involved mechanisms such as trapping, scattering and conduction in ordered or disordered materials as well as the dependence of carrier life time and phonon scattering [13, 25]. In addition, polymers [26], nanomaterials [27] and metamaterials [28] have also been investigated by THz spectroscopy.

To conclude this section, different types of interaction induced by the THz radiation for a series of typical materials are given in Table 1.

<table>
<thead>
<tr>
<th>Material</th>
<th>Excitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semiconductor</td>
<td>Free carrier, phonon</td>
</tr>
<tr>
<td>Ferroelectrics</td>
<td>Soft mode</td>
</tr>
<tr>
<td>Superconductor</td>
<td>Energy gap, quasiparticle excitation, Josephson plasma, 2D Plasmon polariton</td>
</tr>
<tr>
<td>Photonic crystal</td>
<td>Photonic band gap</td>
</tr>
<tr>
<td>Liquid</td>
<td>Relaxation mode</td>
</tr>
<tr>
<td>Gas</td>
<td>Rotational level</td>
</tr>
<tr>
<td>Polymer</td>
<td>Exciton, polaron bipolaron, hoping</td>
</tr>
<tr>
<td>Biomolecules</td>
<td>Collective excitation , biological function</td>
</tr>
</tbody>
</table>

Table 1. Physical effects of the THz radiation on different materials.

4. Data processing of THz spectra

The purpose of this section is to review the different approaches allowing the analysis of THz spectra. In spectroscopy, analysis means identification and quantification. For identification purpose, we propose here to discuss how THz spectroscopy could take advantage of multivariate methods currently used in other fields of spectroscopy like infrared spectroscopy. And for quantitative analysis, two approaches are presented. The first one is based on modeling and can be applied to a single spectrum, without any calibration. At the opposite, the second approach consists in a preliminary calibration based on a series of reference samples in order to finally extract quantitative information from unknown samples. Both univariate and multivariate calibrations are discussed.

4.1 Modeling the THz spectra

Numerical simulations are usually carried out for modeling THz spectra using commercial software and various packages for density functional theory. The underlying strategy is to understand the origin of the absorption lines of organic molecules for single crystals, microcrystalline and powder samples. Here, samples like pesticides, explosives, and drugs are extensively studied as well as amino acids and other biomolecules.

The lattice vibrations giving rise to the terahertz absorption spectra of molecular crystals are a product of the intermolecular interactions specific to the unique three-dimensional arrangement of these ordered molecules. However, since there is no group-characteristic frequency in the THz spectra of crystalline molecular solids, understanding the basic vibrational motions at the origin of the absorption bands is very challenging. Consequently, the interpretation of THz absorption features must be assisted by computational approaches well designed to adequately model the appropriate intermolecular forces at the origin of the potential energies of the related vibrations (cf. Fig 2). Solid-state density functional theory (DFT) [29] was successfully applied to modeling of THz spectra despite its incapacity to correctly describe long-range dispersive interactions. However, the ability of DFT for identifying crystal structures and polymorphism based on observed THz spectra is
now well established [30]. The most common predictive software for the evaluation of the absorption band in the THz spectra of solid materials are: Gaussian, Crystal, CPMD, Castep, CHARMM, VASP and D®mol [6, 29, 31].

The possibility of a simultaneous fitting of both absorption coefficient and index of refraction allows an accurate diagnosis even with relatively weak experimental spectral features. Nonetheless, it is clear that, in the case of low-frequency vibrations in molecular solids, one must carefully decide which molecular modeling approach to apply. And while the numerical methods discussed here rely on the knowledge of the sample composition, additional basic analytical methods must be applied for predicting the response of unknown material.

4.2 Identification of samples

THz spectroscopy can be applied to sample identification, recognition and sorting. Quantitative analysis is not required in this case and a simple comparison between selected spectral features can be performed for different samples to study their similarities. Let’s point out that THz spectroscopy should take advantage of chemometrics, which is already routinely applied in other fields of spectroscopy. Up to now, only sparse papers report on the use of chemometrics for terahertz spectroscopy and as a consequence, we propose here to highlight few selected results.

Principal component analysis (PCA) is probably the most common chemometric method applied for samples sorting. For example, Watanabe et al. [33] obtained a map with the spatial distribution of each component in a pellet. In addition, J.Axel Zeitler et al. [34] used PCA to investigate the variable temperature and variable hydration experiments in the absorbance of pharmaceutical materials. A review of the recent pharmaceutical applications of terahertz pulsed spectroscopy and imaging is proposed by Shen [35]. Here, a series of spectra are simultaneously compared through the measure of the covariance (see for example fig 3). This multivariate approach also allows revealing the correlations between different spectral features and consequently is routinely used for data reduction [36]. Indeed PCA provides new data, which are linear combinations of the initial data. The related mathematical transformation is a change of basis and when the original dataset is displayed on this new basis, it is usually very easy to separate the samples into classes. But this approach is restricted to identification and doesn’t allow quantitative results.

4.3 Quantitative analysis

The simplest method in use for quantitative analysis is univariate calibration, which refers to the construction of a calibration curve, i.e. the experimental THz signal versus the varying parameter, for instance, the analyte concentration. This calibration curve is generally supposed to be linear and the quality of the calibration is evaluated through a linear least squares regression. Nguyen et al. [37] measured the THz spectra of a series of samples composed of cocrystal of an equimolar mixture of phenazine and mesaconic acid. Each sample composition corresponded to a given level of conversion to the cocrystal. In this case the signal for calibration was the area under the peak at 1.2 THz and the good correlation between this signal and the amount of the cocrystal was characterized by R²=0.915. A linear relationship between absorbance intensities at characteristic peak (THz) and concentration values of pharmaceutical component was found to be better for crystalline than for amorphous structure where no strong peak was observed. [38]. Since a single spectral feature is sufficient to obtain a good calibration, univariate analysis is preferred. However, when the spectrum contains many spectral features correlated to the varying parameter, it is necessary to turn to multivariate analysis.

The absorbance of terahertz spectrum being additive according to the Lambert–Beer law, the traditional multi-linear regression was used [37]. After estimating the molar coefficient values ε (ν) for specific mixtures, and extracting the absorbance spectrum, they succeed in the quantification of Amino acids components thanks to a multi- linear regression with a relative error of 12%. Theoretically the traditional multi-linear regression method is known to be efficient in the absence of correlation between variables. But King et al. [39, 40] evidenced the existence of such correlation and explained it was the result of simultaneous absorption by two or more polymorphs in the mixture. In this case an alternative method called Partial Least Square (PLS) regression could be used. PLS algorithm is similar to the one of PCA calculating the maximum covariance but in the case of PLS, the variations taken into account have to be correlated with the analyte concentration. Let’s emphasize that the calculation is based on linear algebra and consequently, this is by definition a multi-linear method. But, since the absorbance of terahertz radiation satisfies the Lambert–Beer’s law, a multi-linear regression appears to be perfectly convenient to calibrate THz data [37, 41, 42]. To go deeper in this approach of chemometrics, interested people are encouraged to read a review of the multi-linear regression methods applied to mid-IR spectroscopy[43]. In 2007, the advantage of using multivariate approach was demonstrated for process analytical technology (PAT)[42]. Multivariate analyses and in particular PLS were successfully applied to THz spectra of pharmaceutical samples in 2005 [41]. In this pioneering work, the authors correlated THz spectra with the analyte concentrations for pharmaceutical products (acetylsalicylic acid (aspirin), lactose, and 4-acetaminophenol) by THz pulsed spectroscopy. In 2007, the advantage of using multivariate approach was reported for process analytical technology (PAT)[42].
In order to assess the performance of a calibration method, it is necessary to split the samples in two sets, namely the calibration set dedicated to the calibration itself and the validation set exploited to evaluate the ability of the calibration model to predict results for unknown samples. This method of cross-validation is general for any calibration method. Typical figures of merit are usually calculated in order to evaluate the calibration, namely the correlation coefficients $R_c$ and the root mean square error of calibration (RMSEC) for the calibration set and the correlation coefficients $R_p$ and the root mean square error of prediction (RMSEP) for the validation set [44, 45]. The optimum calibration model obviously corresponds to the higher value of correlation and to the lower values of errors. Moreover, if the error of the calibration set is much lower than those of the validation set, there is over-fitting [46]. Let’s also emphasize that a small number of PLS components is required to get the best performance [41, 44]. Consequently, the THz spectral window to be analyzed should be reduced in order to remove most of the noise while keeping relevant data correlated with the concentration of the analyte. More precisely, one should remove the high frequency range of the THz spectra prior to any data processing [47] because of poor experimental repeatability in the high frequency experimentally.

4.4. Limitation of experimental procedure

The samples under investigation are often in powder form and then, pellets made with a transparent matrix material are fabricated for spectroscopic analysis. As a consequence, spectral distortions are commonly observed in the terahertz spectra in the case of granulated materials. These false spectral features are caused by scattering due to the refractive index mismatch between the particles and their surrounding medium. Scattering is a major problem in precise measurement of quasi-optical parameters of material samples; some papers propose mitigating techniques and recommend a novel method that allows calculating true absorption spectra for samples with some limitations [48].

The quality of calibration also depends on the pre-processing methods like base line correction, smoothing, first and second derivative, mean centered filtering, normalization, multiplicative scatter or (signal) correction (MSC), standard normal variate (SNV). Wu et al. [38] reported that the analyses of Theophylline and Lactose amounts after direct compression tablets gave the best performance without any pre-processing. At the opposite, for wet granulation tablets, pre-processing was required. More precisely, MSC and 1st derivative gave the best prediction for wet granulation tablets with respectively absolute average error of prediction 5.94%-5.92% with MSC for Theophylline and 6.05%-5.95% with 1st derivative for Lactose amounts.

Ostuka et al.[45] studied the best parameters and preprocessing methods in the case of transmittance THz spectra of the forms I and II of the polymorph Mefenamic acid. The spectra were really hard to interpret in quantitative analysis, even after smoothing and normalizing, but after a convenient pre-treatment, i.e. the selection of a frequency range (0.46-6.05 THz) with smoothing and second derivative, they obtained a sufficient model to quantify the form I into mixtures containing the two forms. This pre-processing provided the highest value of correlation coefficient (R= 0.9692). Until now there is not a unique method of data preprocessing, and one should evaluate the performance of different pretreatments in order to select the most suitable.

To go further, when the relationship between the signal and the analyte concentration is nonlinear, it is necessary to turn to more advanced treatment. Artificial neural network (ANN) is a mathematical method inspired by biological neural networks which is perfectly well-suited to deal with nonlinearities in the calibration. ANN consists of an interconnected set of artificial neurons. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Many different types of ANN have been developed. In the case of spectroscopy, input values are provided by the spectra and output values are the analyte concentrations. Since the performance of ANN relies on the choice of the input data, it is very important to carefully select only relevant data and to remove noise and redundancy information. The best way to select the most relevant data and to reduce their number could be to use PCA. Indeed, few papers implemented PCA for imaging or spectral identification of explosives by THz [49, 50]. To our knowledge, ANN which is routinely used for quantification in NIR, IR, and LIBS spectroscopies should be efficiently applied to quantitative analysis in terahertz spectroscopy.

5. Conclusion

Recent progress on terahertz technology pushes away the limits for the intensive use of this spectral range. Nowadays terahertz spectroscopy is a well established technique with dozens of companies around the world that build and sell turnkey systems. Varieties of applications such as imaging, non-destructive testing, and quality control have been transferred to industry supported by basic research. In this paper, the most relevant experimental setups for THz spectroscopy were presented and quantitative analysis was discussed from the simplest univariate approach to the most complex methods referenced as chemometrics. With the advantage of a direct access to the phase compared to IR spectroscopy, we strongly believe that terahertz spectroscopy should take advantage of advanced data processing and that should become the key point for further development of this method as a new reliable analytical spectroscopic technique. In the future, we strongly promote systematic works for the detailed assessment of THz spectroscopy with chemometrics, including pre-processing, data selection and calibration methods.
References:


