Gas recognition with terahertz time-domain spectroscopy and spectral catalog: a preliminary study

H. Lin, W. Withayachumnankul, B. M. Fischer, S. P. Mickan, and D. Abbott

Centre for Biomedical Engineering and School of Electrical & Electronic Engineering,
The University of Adelaide Adelaide, SA 5005, Australia

ABSTRACT

The ability to detect and recognise a variety of gases under certain environmental conditions has significant potential impact in many areas: from hazardous gas detection in the industrial domain to physical sciences in academia. Gas sensing has long received attention with microwave and infrared spectroscopy. With many molecular resonances occurring in the THz (T-ray) range leading to simple, unique spectral features, THz time-domain spectroscopy (TDS) shows promise as a tool for gas detection and recognition. This paper presents a preliminary study on real-time gas recognition with THz-TDS. In particular, a simple method is used that involves extracting line positions from gas species without a reference pulse and classifying them by means of the minimum Euclidean distance using the Submillimeter, Millimeter, and Microwave Spectral Line Catalog.

Keywords: Gas recognition, THz-TDS, gas spectroscopy, JPL catalog, T-rays, ultrafast measurement

1. INTRODUCTION

Fundamentally, the interaction between molecules and an electromagnetic wave is determined by the incident energy, molecular species, substance condition, etc. A molecule can absorb and re-emit the wave at certain frequencies, specific to the energy transitions of either electronic, vibrational, or rotational modes. Each molecular species absorbs the wave in a unique spectral pattern and even though the substance and surrounding environment can, to a certain extent, affect the absorption pattern, the absorption features are generally unique to each molecule.

In the gas phase, the rotational transition modes occur in polar molecules, and these transitions broadly span from the microwave to infrared (IR) spectra. The pure rotational transitions result in an absorption spectrum that contains Lorentzian resonances at discrete frequencies. The spectrum, as mentioned above, is unique to the molecule. This uniqueness enables the classification and recognition of polar gases via either microwave, T-ray, or infrared spectroscopy of a gas sample.

The Terahertz (THz) or T-ray region of the electromagnetic spectrum is loosely defined in the literature as the frequency range from 0.1 to 10 THz*, situated between the microwave and IR regions. Historically, T-rays have been employed in astronomy for the spectral characterisation of the rotational and vibrational resonances, and thermal emission lines of simple molecules.[1] Prior to the 19890s, the T-ray region was described as the THz gap, due to poor hardware capability, high atmospheric absorption, and strong thermal background noise. However, significant progress has been made in the generation and detection instrumentation, thus making the THz gap evermore accessible.

Terahertz time-domain spectroscopy (THz-TDS) represents a breakthrough in T-ray generation and detection. At the generation end, it commonly employs photoconductive antenna (PCA)[2] or nonlinear electro-optic (EO) crystal[3] to convert an ultrashort optical burst into a coherent T-ray pulse. At the detection end, an optically gated antenna or EO crystal is employed, thus enabling the recording of high-SNR time-resolved T-ray waveforms.

Email addresses: hlin@eleceng.adelaide.edu.au (H. Lin); withawat@eleceng.adelaide.edu.au (W. Withayachumnankul); spmickan@eleceng.adelaude.edu.au (S. P. Mickan); bfischer@eleceng.adelaide.edu.au (B. M. Fischer); dabbott@eleceng.adelaide.edu.au (D. Abbott)

*Some texts define it as 0.3 to 30 THz

Terahertz Photonics, edited by Cunlin Zhang, Xi-Cheng Zhang, Proc. of SPIE
Vol. 6840, 68400X, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.760558

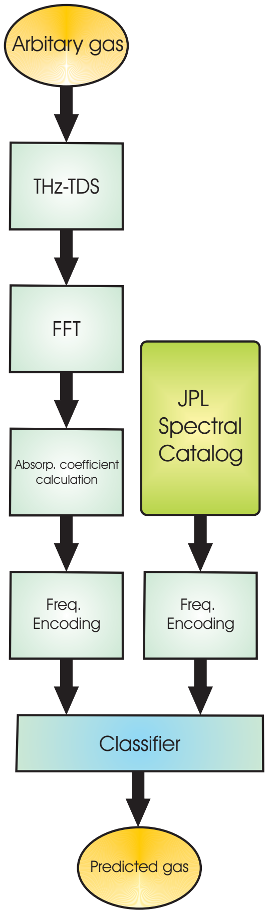


Figure 1. A flow chart for the proposed approach. THz-TDS is performed on an arbitrary gas sample with the resonant frequencies being obtained from the spectrum and the estimated reference pulse. These frequencies are then encoded as boolean values for classification purposes with a library of boolean encoded cataloged gases.

Earlier works on real-time gas recognition with THz-TDS have modeled the power spectrum of various gases and identified them with some success.[4, 5] In particular, linear predictive coding (LPC) has been demonstrated to be an efficient technique for capturing the sharp spectral features as opposed to Fourier analysis methods. However, determining the number of LPC coefficients is by no means a trivial task as too many lead to overfitting, while too few result in an inaccurate estimation. Furthermore, the technique is reported to fail for gas species with many absorption lines.[5] In this paper, we present a simple technique (Figure 1) that compares the spectral features from an existing spectral database with measured sample features, with an aid from recent work on reference pulse estimation.[6]

The paper is organized as follows: Section 2 presents the background of research by briefly reviewing existing gas sensing technologies and THz-TDS; Section 3 describes the spectral feature extraction process while Section 4 demonstrates a simple Euclidean distance classifier for this preliminary work; The paper ends with Section 5, discussion on future work.

2. GAS RECOGNITION WITH THZ-TDS

2.1. Gas recognition technologies

Gas sensing technologies have been widely reviewed in the literature[7–9] and can be broadly classified into four major groups; gas chromatography and mass spectrometry (GCMS), electromechanical sensors, mass sensors and optical based techniques. GCMS is extensively used in the medical, pharmacological, environmental and law enforcement fields to analyse chemical mixtures. Even though GCMS has the best selectivity, high sensitivity, precision and reproducibility, the technology is very expensive and training maybe needed. In addition, the retention times in GC are typically in the order of several minutes, therefore inducing delays in the identification process,[10] thus limiting real-time applications. In contrast, electromechanical sensors are small, low cost, low powered components that have short shelf-lives with performances such as sensitivity and selectivity dependence on the operating environment. Likewise, piezoelectric acoustic waves sensors are also small sized and low powered, and are sensitive to various chemicals.[7] However, they cannot discriminate across unknown mixtures and durability remains to be a concern.[11] Fourier Transform Infrared (FTIR) spectroscopy or infrared sensing is a well established optical method that uses fingerprint absorption signatures resulting from molecular vibrations for identification. But due to the nature of molecular vibrations, only certain gas species with a permanent dipole moment maybe fingerprinted and hence identified. In addition, the technology is expensive, complex, and can be affected by water vapor and humidity. Other optical sensing methods like microwave spectroscopy (300 MHz to 300 GHz), commonly known today as Fourier transform microwave spectroscopy (FT-MW),[12] is also a highly selective method for many polar gases that exhibit unique absorption signatures of high spectral resolution. With the availability of highly stable microwave sources, FT-MW also has the added benefit of overall system stability and durability. In general, optical techniques can be highly selective and robust and require less calibration than other sensing technologies,[11] thus making them a perfect trade-off between cost, robustness and performance.

2.2. Fundamental considerations

On a fundamental level, microwaves and mid-infrared radiation has its own respective advantages and disadvantages for spectroscopy. Principally, heavy gas molecules tend to have rotational resonances at microwave frequencies, whereas light gas molecules have rotational resonances in the mid-infrared. The absorption resonances recorded in the mid-infrared region are generally congested and complicated by the combination of rotational and ro-vibrational modes. This therefore hinders the feature extraction process. In contrast, the rotational transition modes in the microwave region are relatively low, resulting in few available features. T-rays therefore present an ideal trade-off. In addition, the interactions between EM wave and polar molecules in the T-ray region are three to six orders of magnitude stronger than that in the microwave region.[13]

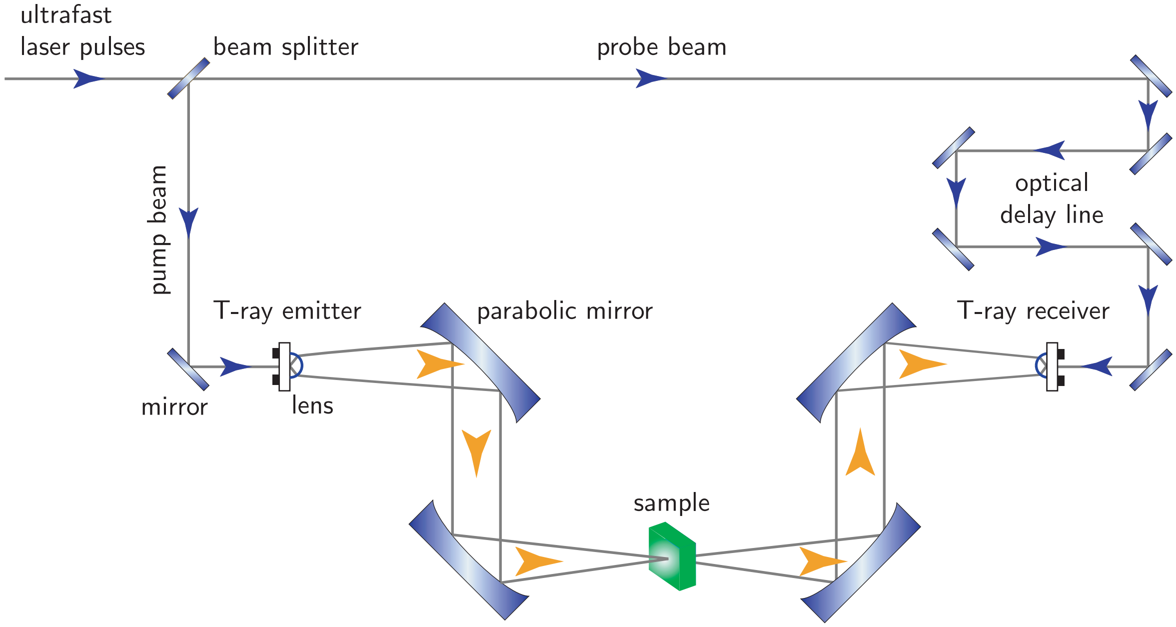


Figure 2. THz-TDS system configured in transmission mode with PCA generation and detection. The system consists of an ultrafast optical laser, T-ray emitter/receiver, an optical delay line, a set of mirrors, and a material sample. The optical beam paths are indicated by small blue arrowheads, and the T-ray beam paths by the large orange arrowheads.

2.3. THz-TDS system

THz-TDS in general has a better resolution compared to FTIR[13] and is able to collect the data almost instantaneously leading to real-time detection. In particular, amplitude and phase information can be obtained giving rise to more features for classification purposes. Even though FTIR spectroscopy has sensitivity spanning over many frequencies (from as low as 1 THz up to the visible spectrum), T-ray spectrometers are more sensitive below 3 THz[14] where the rotational modes lie. With many gases exhibiting absorption spectra due to molecular rotations in this region, THz-TDS promises to complement the relatively mature FTIR and expand the number of detectable gas species. The THz-TDS system shown in Figure 2 is mainly composed of an ultrafast optical laser, T-ray emitter/receiver, an optical delay line, a set of mirrors, and a material sample. The ultrafast optical pulse is divided into two paths, a probe beam and a pump beam, by a beam splitter. At the emitter, the optical pump beam stimulates T-ray pulsed radiation via either charge transport[2] or optical rectification effect,[3] depending on the emitter type. The diverging T-ray beam is collimated and focuses onto the sample by a lenses and a pair of parabolic mirrors. After passing through the sample, the T-ray beam is re-collimated and focused onto the receiver by an identical set of lenses and mirrors. At the receiver, the initially divided probe beam optically gates the T-ray receiver with a short time duration compared with the arriving T-ray pulse duration. Synchronizing between the optical gating pulse and the T-ray pulse allows the coherent detection of the T-ray signal at a time instance. A complete temporal scan of the T-ray signal is enabled by the discrete micro-motion of a mechanical stage controlling the optical delay line.

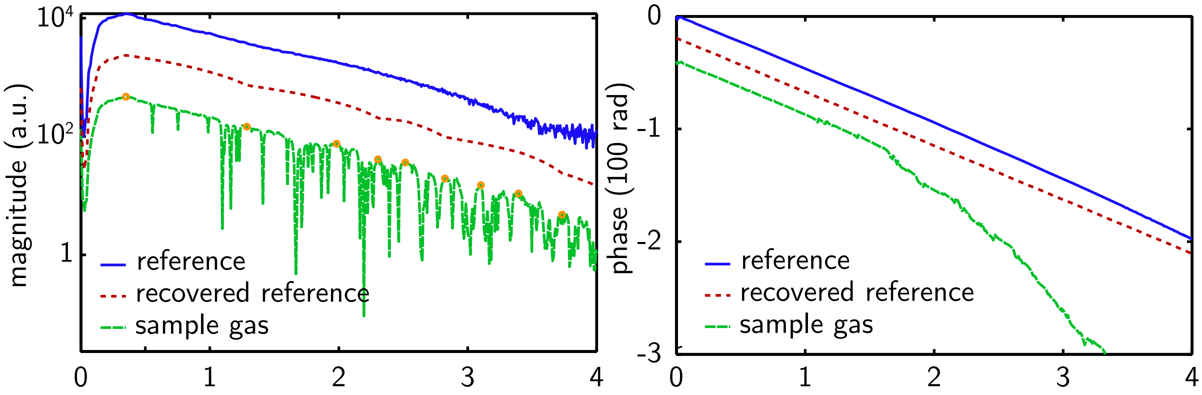
3. FEATURE EXTRACTION

The goal of feature extraction is to select features such that the samples from each class are in forms of clusters well separated from one another. In a T-ray spectrum, polar gases with a sufficient concentration, exhibit an ensemble of strong and sharp resonances at discrete frequencies due to quantised rotational energy transitions.[5] As the molecules differ in structure, the resonances' position and relative strength are unique to each gas species. This uniqueness is hence used as a classification feature for gas recognition. Specifically, the resonances' position or frequency of the peaks is considered because it is rarely affected by the gas pressure, gas temperature, nor system's dynamic range as is the absorption strength. The absorption strength is useful for the selection of well-defined resonances and the elimination of resonance-like spikes caused by random noise in the spectrum. A system modeling approach with LPC has been successfully employed to capture the resonant positions and strengths efficiently.[5] In particular, the coefficients for particular gases are extracted from the time-domain T-ray waveform to estimate the sample spectra for comparison and identification. The technique, however, does not take account of the phase information leading to inability in identifying gases with many absorption lines such as CH_3CN.[5] Although the number of coefficients is fixed for each waveform thereby making subsequent comparisons simple, choosing the optimal number is difficult as it is dependent on the gas species. Fourier series coefficients have also been attempted, but it was found that more coefficients were required to model the sharp resonant lines but lead only to minor performance improvement.[4] Also exploiting the simplicity and uniqueness of the T-ray spectral signatures, this feature extraction method aims to utilise the the spectral features present from an existing spectroscopic catalog for subsequent comparison with the sample features extracted without a reference pulse.

3.1. Spectroscopic catalog features

Computer-accessible spectroscopic information like resonances position, strength and width, of many common gas species are available in many databases or catalogs, for example:

- Molecular Spectroscopy Team, at Jet Propulsion Laboratory (JPL) reports the spectral line intensity, strength, and other related parameters of 331 interstellar gas species in the submillimeter, millimeter, and microwave frequency ranges.[15]
- Cologne Database for Molecular Spectroscopy (CDMS) added 280 more species to its own database, with the identical data format used at JPL.[16]



(a) T-ray magnitude spectra

(b) T-ray phase spectra

Figure 3. Magnitude and phase spectra for reference, recovered reference, and water vapor. (a) The water vapor or sample spectrum possesses a set of clear resonances. To determine the reference magnitude, the sample spectrum is spatially averaged and the local maxima are determined, as indicated by the superimposed dotted. Simply connecting the local maxima by linear functions recovers the reference magnitude. (b) The phase spectrum of the reference also needs recovery. By fitting a linear function to the sample phase spectrum, the reference phase can be estimated.

- HIgh-resolution TRANsmission molecular absorption database (HITRAN) provided a compilation of spectroscopic parameters of 39 molecules and its isotopologues, spanning the microwave and ultraviolet bands.[17] Apart from the line intensity and position, the list consists of the Einstein *A*-coefficients, statistical weight for upper and lower levels of the transitions, etc.

- Gestion et Etude des Informations Spectroscopiques Atmosphériques (GEISA) database contains the spectroscopic parameters of 42 molecules (96 isotopes) in the frequency range between 0 and 1076 THz.[18]

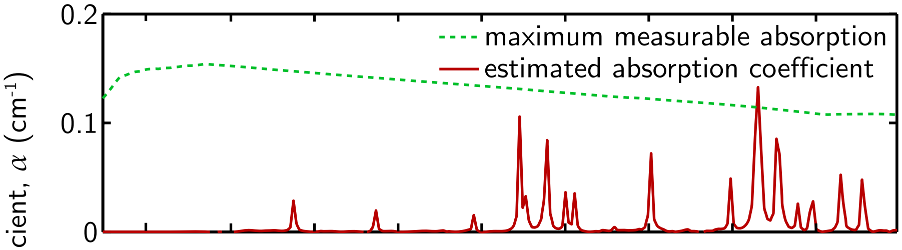
The Submillimeter, Millimeter, and Microwave Spectral Line catalog consists of many gas species with spectral lines in the frequency range between 0 and 10000 GHz, which overlaps with the T-ray range of 0.1 to 2 THz considered in this paper. For the sake of resonant frequency extraction, only the frequency and intensity columns are used.[15] In this region, the absorption lines are due to rotational and vibrational modes. Since the rotational lines are typically stronger than the vibrational lines in this range, in small molecules like water, a threshold is set to filter out the rotational modes and extract their frequencies.

3.2. Sample spectral features

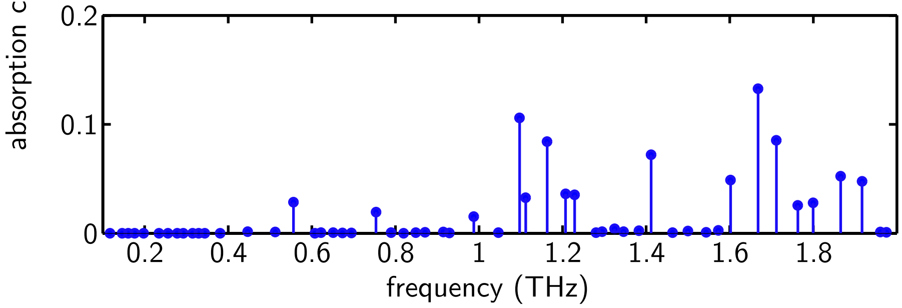
An issue arises as the gas recognition process proposed here eliminates the retrieval of a reference pulse to facilitate the measurement. This means that only the gas spectrum is available without the baseline to determine the absolute value of resonance strength. An idea to estimate the reference data from the sample data is therefore introduced. Once the reference is recovered, the full optical constants, including the absorption coefficient and the refractive index, can be determined. Extracting the resonances' position and strength from the absorption coefficient is more robust.

The recovery process could possibly apply with the sample data either in the time or frequency domain. In the frequency domain, the rotational resonances modify the reference only at narrow frequency bands, and the spectral shape of the reference is still preserved in general. In the time domain, the resonances cause significant distortion in the main pulse and also strong oscillations in the tail after the main pulse, leaving only small resemblance to the reference pulse. Therefore, recovery of the reference in the frequency domain is more reasonable.

Both magnitude and phase of the sample spectrum must be modified in order to recover the reference. The magnitude and shape of the reference are very dependent on the system. On interaction with a polar gas, the



(a) absorption coefficient



(b) extracted features

Figure 4. Absorption coefficient and extracted features. (a) The absorption coefficient of water vapor is extracted from the measured sample spectrum and the recovered reference. (b) The coefficient in Figure (a) is determined for its local maxima, which represent the position and strength resonance features ready for gas recognition.

magnitude exhibits a number of sharp Lorentzian resonances or abysses. Filling up the resonance abysses simply regains an approximation of the reference magnitude. More specifically, the magnitude spectrum is spatially averaged to eliminate small spikes. Subsequently, the local maxima are determined. These local maxima are connected linearly, resulting in the estimated reference magnitude. The process is depicted in Figure 3 (a).

The phase of the reference is nearly linear over the whole frequency range. The existence of a polar gas in the propagation path causes small phase perturbations at resonance frequencies. Applying the least squares fitting to the sample phase could smooth out these perturbations. A comparison among reference, sample, and modified phases are shown in Figure 3 (b).

After both the magnitude and phase of the reference are recovered, they can be used in conjunction with the sample spectrum to estimate the frequency-dependent optical constants. The absorption coefficient obtained from the measured sample and the recovered reference is plotted in Figure 4 (a). From the estimated absorption coefficient, the resonances' position and strength can be easily determined, as shown in Figure 4 (b). These features can be recognized by comparing against the gas spectral parameters in a spectroscopic catalog.

4. CLASSIFICATION

4.1. Data Encoding

In order to simplify comparisons between different gas species that differ in number of absorption peaks, a number of equal sized frequency bins are constructed in the frequency range of 0.1 to 2.0 THz, with the frequency resolution based on the measurement data. A typical THz-TDS system provides a frequency resolution of 37.5 GHz so this give rise to 50 or so frequency coefficients. An absorption peak at a particular frequency marks the respective frequency bin with a Boolean one, otherwise the default value is a Boolean zero indicating no peak. This simple data encoding technique is applied to both the catalog and sample features determined in Section 3. Figure 5 shows a data encoding example for water spectral features recorded in the catalog and the spectral features for an arbitrary sample.

Catalog (H_2O)

- - -	0	0	1	0	0	- - -	
- - -	0.488	0.525	0.563	0.600	0.638	- - -	THz

Sample (arbitrary)

- - -	1	0	0	1	1	- - -	
- - -	0.488	0.525	0.563	0.600	0.638	- - -	THz

Figure 5. Data encoding for the catalog and sample spectral features. The frequency resolution illustrated here is 37.5 GHz. Water has an absorption peak at 0.5569 THz and this is encoded with a boolean one in the 0.563 THz bin and zero elsewhere.

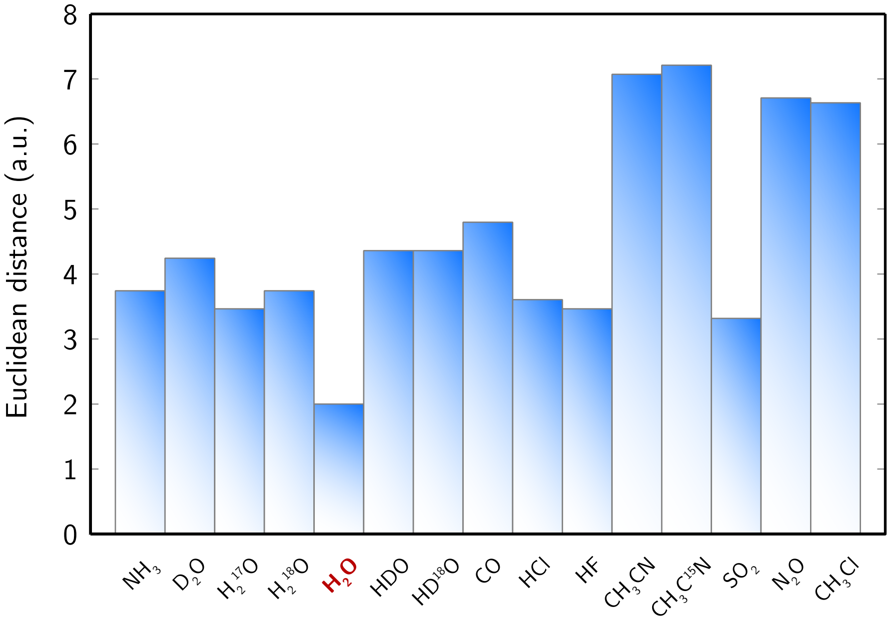


Figure 6. A column plot of the Euclidean distance measures between an unknown sample (water vapor) and various catalog gas species including water isotopes. Clearly, water can be recognised (highlighted in red) with the minimum Euclidean distance.

4.2. Euclidean Distance Classifier

Once the spectrum is encoded, the sample gas is classified by means of minimum Euclidean distance to the gas species in the catalog. The Euclidean distance measure is a simple, fast metric that has a straightforward geometric interpretation. In general, even though the metric suffers from the drawbacks of incomparable scales of different variables and an inability to identify correlation between the variables, it nevertheless is easy to implement and computationally efficient and is hence sufficient for the initial purposes. By determining the Euclidean distance between an unknown gas sample and catalog gas species, the sample is considered to be a particular gas that has the minimum Euclidean distance. Specifically, the distance between each respective frequency bin is computed, combined and compared. Figure 6 shows the column plot of various Euclidean distance measures between an unknown gas sample (water vapor) and various catalog gas species that include many water isotopes (HDO, D_2O) and common gases. Evidently, the unknown sample can be recognised automatically with the minimum Euclidean distance. In an effort to further minimise the Euclidean distance, the idea of zero padding is investigated, but was found to have little impact on the distance indicating the frequency resolution of the sample is sufficient. Interestingly, there is no direct correlation between the distance measure between water and its isotopes, which is due to the fact that the absorption peaks are essentially random. Classification accuracy is largely affected by frequency resolution and the effect of noise. Frequency resolution determines the resolving power of the resonances and this is governed by the THz-TDS hardware. A typical resolution of 37.5 GHz appears to be sufficient for this work. Noise is inevitable in the THz-TDS setup and may propagate to the measured frequency. Further information on the propagation of noise is provided in Withayachumnankul et al. (2007).[19] To combat the effect of noise, more measurements on the unknown sample would be necessary for subsequent averaging and comparisons.

5. CONCLUSION & DISCUSSION

The ability to detect and measure a variety of gases under specific environmental conditions has a significant impact in many areas. With many molecular resonances occurring in the THz range leading to unique spectral features, THz-TDS is a promising technique for real-time gas sensing. This paper has presented a preliminary study of real-time gas recognition that involves extracting resonances' positions from gas species without the reference pulse and classifying them by means of the minimum Euclidean distance with the Submillimeter, Millimeter, and Microwave Spectral Line catalog. Immediate work will focus on improving the robustness of this technique. In particular, we will expand on the number of catalog gas species and experiment with more gas samples. Future work will concentrate on gas recognition in gas mixtures.

REFERENCES

1. P. Siegel, "Terahertz technology in biology and medicine," *IEEE Transactions on Microwave Theory and Techniques* **52**(10), pp. 2438–47, 2004.
2. P. Smith, D. Auston, and M. Nuss, "Subpicosecond photoconducting dipole antennas," *IEEE Journal of Quantum Electronics* **24**(2), pp. 255 – 60, 1988.
3. L. Xu, X.-C. Zhang, and D. Auston, "Terahertz beam generation by femtosecond optical pulses in electro-optic materials," *Applied Physics Letters* **61**(15), pp. 1784 – 6, 1992.
4. R. Jacobsen, D. Mittleman, and M. Nuss, "Chemical recognition of gases and gas mixtures with terahertz waves," *Optics Letters* **21**(24), pp. 2011–2013, 1996.
5. D. Mittleman, R. Jacobsen, R. Neelamani, R. Baraniuk, and M. Nuss, "Gas sensing using terahertz time-domain spectroscopy," *Applied Physics B (Lasers and Optics)* **B67**, pp. 379 – 90, Sept. 1998.
6. W. Withayachumnankul, H. Lin, B. M. Fischer, S. P. Mickan, and D. Abbott, "Blind parameter estimation of gases from THz-TDS," *in preparation* , 2007.
7. B. Drafts, "Acoustic wave technology sensors," *IEEE Transactions on Microwave Theory and Techniques* **49**(4), pp. 795 – 802, 2001.
8. D. Wilson, S. Hoyt, J. Janata, K. Booksh, and L. Obando, "Chemical sensors for portable, handheld field instruments," *IEEE Sensors Journal* **1**(4), pp. 256 – 74, 2001.

9. P. A. Lieberzeit and F. L. Dickert, "Sensor technology and its application in environmental analysis," *Analytical and Bioanalytical Chemistry* **387**(1), pp. 237 – 247, 2007.
10. M. D. Brickhouse, W. R. Creasy, B. R. Williams, K. M. Morrissey, R. J. O'Connor, and H. D. Durst, "Multiple-technique analytical characterization of a mixture containing chemical-weapons simulant from a munition," *Journal of Chromatography A* **883**(1), pp. 185 – 198, 2000.
11. C. K. Ho, M. T. Itamura, M. Kelley, and R. C. Hughes, "Review of chemical sensors for in-situ monitoring of volatile contaminants," tech. rep., Scandia National Laboratories, 2001.
12. R. Dicke and R. Romer, "Pulse techniques in microwave spectroscopy," *Review of Scientific Instruments* **26**(10), pp. 915 – 928, 1955.
13. F. C. De Lucia, *Sensing with Terahertz Radiation*, vol. 85 of *Springer Series in Optical Sciences*, Springer, 2003.
14. P. Han, M. Tani, M. Usami, S. Kono, R. Kersting, and X.-C. Zhang, "A direct comparison between terahertz time-domain spectroscopy and far-infrared fourier transform spectroscopy," *Journal of Applied Physics* **89**(4), pp. 2357 – 9, 2001.
15. H. Pickett, R. Poynter, E. Cohen, M. Delitsky, J. Pearson, and H. Muller, "Submillimeter, millimeter, and microwave spectral line catalog," *Journal of Quantitative Spectroscopy and Radiative Transfer* **60**(5), pp. 883 – 890, 1998.
16. H. S. P. Müller, F. Schlöder, J. Stutzki, and G. Winnewisser, "The Cologne Database for Molecular Spectroscopy, CDMS: A useful tool for astronomers and spectroscopists," *Journal of Molecular Structure* **742**, pp. 215–227, 2005.
17. L. S. Rothman, D. Jacquemart, A. Barbe, D. C. Benner, M. Birk, L. R. Brown, M. R. Carleer, C. Chackerian, K. Chance, L. H. Coudert, V. Dana, V. M. Devi, J. M. Flaud, R. R. Gamache, A. Goldman, J. M. Hartmann, K. W. Jucks, A. G. Maki, J. Y. Mandin, S. T. Massie, J. Orphal, A. Perrin, C. P. Rinsland, M. A. H. Smith, J. Tennyson, R. N. Tolchenov, R. A. Toth, J. Vander Auwera, P. Varanasi, and G. Wagner, "The *HITRAN* 2004 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer* **96**, pp. 139–204, 2005.
18. N. Jacquinet-Husson, E. Arié, J. Ballard, A. Barbe, G. Bjoraker, B. Bonnet, L. R. Brown, C. Camy-Peyret, J. P. Champion, A. Chédin, A. Chursin, C. Clerbaux, G. Duxbury, J.-M. Flaud, N. Fourrié, A. Fayt, G. Graner, R. Gamache, A. Goldman, V. Golovko, G. Guelachvili, J. M. Hartmann, J. C. Hilico, J. Hillman, G. Lefèvre, E. Lellouch, S. N. Mikhaïlenko, O. V. Naumenko, V. Nemtchinov, D. A. Newnham, A. Nikitin, J. Orphal, A. Perrin, D. C. Reuter, C. P. Rinsland, L. Rosenmann, L. S. Rothman, N. A. Scott, J. Selby, L. N. Sinitsa, J. M. Sirota, A. M. Smith, K. M. Smith, V. G. Tyuterev, R. H. Tipping, S. Urban, P. Varanasi, and M. Weber, "The 1997 spectroscopic GEISA databank," *Journal of Quantitative Spectroscopy and Radiative Transfer* **62**(2), pp. 205–254, 1999.
19. W. Withayachumnankul, H. Lin, S. P. Mickan, B. M. Fischer, and D. Abbott, "Analysis of measurement uncertainty in THz-TDS," *Photonic Materials, Devices, and Applications II* **6593**(1), p. 659326, SPIE, 2007.