

Chemical recognition of gases and gas mixtures with terahertz waves

R. H. Jacobsen,* D. M. Mittleman,[†] and M. C. Nuss

Bell Labs, Lucent Technologies, 101 Crawfords Corner Road, Holmdel, New Jersey 07733

Received August 9, 1996

A time-domain chemical-recognition system for classifying gases and analyzing gas mixtures is presented. We analyze the free induction decay exhibited by gases excited by far-infrared (terahertz) pulses in the time domain, using digital signal-processing techniques. A simple geometric picture is used for the classification of the waveforms measured for unknown gas species. We demonstrate how the recognition system can be used to determine the partial pressures of an ammonia-water gas mixture. © 1996 Optical Society of America

The analysis of chemical compositions is a fundamental issue in molecular spectroscopy. In the far-infrared region of the electromagnetic spectrum, time-domain spectroscopy has been facilitated by a novel development of pulsed broadband sources. This spectroscopic technique, commonly known as terahertz (THz) time-domain spectroscopy, is based on optoelectronic techniques^{1,2} that use photoconductive antennas excited by femtosecond laser pulses to generate and detect subpicosecond electromagnetic pulses. The THz waveforms consist of a single-cycle oscillation of the electromagnetic field and have spectra covering frequencies from 100 GHz to more than 2 THz, and signal-to-noise ratios as high as $10^6 : 1$ have been obtained.³

The impulse character of the THz waveforms has motivated studies of the coherent transient response of resonantly excited gases^{4,5} and gas mixtures.⁶ The coherent transient phenomena show up in the time-domain spectra as free induction decays⁷ (FID's) characterized by fast oscillations⁴ and commensurate echoes⁵ that occur after the impulse excitation. The FID is a unique time-domain representation of the absorption and dispersion characteristics of the sample and serves as a fingerprint of its chemical composition. However, despite its time-resolved nature, THz spectroscopy has so far been based on conventional Fourier-transform techniques.

In this Letter we demonstrate a time-domain chemical-recognition system for the analysis of the chemical compositions of infrared active gas systems. We model the action of a resonantly excited gas or gas mixture by a linear digital filter that reshapes an incident THz waveform to produce the corresponding FID. These circumstances strongly resemble the basic problems of spectral estimation for phonetic recognition,⁸ and theory and algorithms from speech recognition systems have been adopted. In particular, we apply a correlation type of analysis known as linear predictive coding^{9,10} (LPC) to extract and parameterize the spectral features of a waveform. A simple geometric picture permits the use of these parameters for classification of individual gas species as well as mixture analysis.

In linear prediction, subsequent values in a time series of a digital signal, $s(n)$, can be generated from a weighted sum of the past M values:

$$s(n) = \sum_{k=1}^M a_k s(n-k) + e(n),$$

where $e(n)$ is the residual, and a_k are the parameters of the linear prediction. To illustrate the basic physics involved we introduce the z transform of Eq. (1):

$$S(z) = \sum_{k=1}^M a_k z^{-k} S(z) + E(z), \quad (2)$$

where $z = \exp(i2\pi f\Delta)$ and Δ^{-1} is the sampling frequency. We find that the signal $S(z)$ is constructed from the residual $E(z)$ by an all-pole filtering:

$$S(z) = H(z)E(z), \quad (3)$$

where

$$H(z) = \frac{1}{1 - \sum_{k=1}^M a_k z^{-k}}. \quad (4)$$

The LPC coefficients a_k contain the information on the spectral content of the waveform. The advantages of using LPC for chemical recognition of gases are twofold: (1) Poles provide an accurate representation for an underlying power spectrum that has sharp spectral lines. This is in contrast to Fourier analysis, in which a signal waveform is expanded into a Fourier series. Such series can have only zeros, not poles, and must attempt to fit sharp spectral features with a polynomial that requires a large number of coefficients. (2) By using the LPC we can always rely on the same number of coefficients M , regardless of the number of resonances excited. This enables us to use simple geometric pictures instead of complicated template-matching techniques for the classification of waveforms.

The optimum values for the LPC coefficients are calculated from least-squares principles, i.e., we wish to minimize the total squared residual: $\mathcal{E} = \sum e(n)^2$. The analysis yields a system of normal equations¹⁰ that relates the LPC coefficients to values of autocorrelation function $\sum_k s(k)s(k-n)$ of the time series representing the waveform.

The parameters from a LPC analysis are treated as a vector, $\mathbf{a} = (a_1, a_2, \dots, a_M)$, in an M -dimensional vector space. Vectors representing known gas species are stored in a codebook, and the codebook is built by

successive measurements and subsequent LPC analysis for a variety of different gases. The final codebook contains a set of linearly independent vectors $\{\mathbf{a}_i\}$, $i = 1, 2, \dots, p$, $p < M$, that span a subspace of the M -dimensional vector space. A single-species recognition procedure consists of a full search of the codebook to find the best match between the trial vector, i.e., the LPC coefficients of an unknown species, and the vectors in the codebook. We accomplish this by calculating the Euclidean distances between the vectors and use the minimum distance as the classification criterion.

A gas-mixture analysis uses the fact that the LPC vectors from different gases add in the M -dimensional vector space. Hence, a coordinate representation x_i of the trial vector in the $\{\mathbf{a}_i\}$ basis becomes a measure of the mole fractions of the species present in the mixture. We illustrate these principles in Fig. 1, using a binary mixture as an example. The LPC vector representing the mixture \mathbf{a} is the vector sum of coded vectors for the individual species weighted by their mole fractions. In practice, an orthogonal basis $\{\mathbf{b}_i\}$ is constructed and used to calculate the orthogonal projections y_i . A linear transformation establishes the relation between the two bases, $\{\mathbf{a}_i\}$ and $\{\mathbf{b}_i\}$.

For experimental realization of our proposed chemical-recognition system we use a standard setup for THz time-domain spectroscopy.^{2,4,5} In the experiment we pass a collimated THz beam through a 30.5-cm-long gas cell.

We have studied the gases HCl, NH₃, H₂O, and CH₃CN in a pressure range from 0.3 to 13 kPa. Figure 2 shows an example of a THz waveform resulting from propagation in a HCl atmosphere at a pressure of 13 kPa. The measured FID shows oscillations that are due to the frequency beating among resonances at 0.626, 1.251, and 1.876 THz.¹¹ We obtain the waveforms by sampling $N = 1024$ points with a temporal resolution of $\Delta = 93$ fs. That is, the sampling of a particular waveform is usually interrupted before the FID has completely decayed. This illustrates that frequency resolution $(N\Delta)^{-1}$ is not an issue in our chemical-recognition system. In fact the LPC is able to predict sharp resonances by using a very small number of data points ($N \approx M$), where Fourier analysis fails to resolve the resonances. This is due to the predictive properties of the LPC analysis, and hence the LPC offers better quantitative analysis for a small number of data points.

Before the LPC analysis, the waveforms are preprocessed by finite impulse response filtering and time-domain windowing. We do this to stabilize the LPC by reducing the risk of predicting spurious frequency peaks. Because we are interested in only the coherent irradiance from the gas, we begin the LPC analysis for values appearing after the impulse excitation.

Figure 3 shows a comparison between the maximum-entropy-method power spectrum,¹⁰ calculated from the LPC coefficients, and the Fourier-transform power spectrum for a waveform that has been propagated through H₂O vapor. The maximum-entropy method spectrum is estimated from the power spectrum of the all-pole filter. The comparison be-

tween the two spectra serves as a confidence measure for the LPC but is not used for recognition purposes. The frequency resolution obtained from the Fourier analysis is better than the resolution obtained from the LPC. However, the LPC analysis correctly predicts the power, i.e., the integral of the power spectral density, in an absorption line.

To demonstrate the use of the chemical-recognition system for mixture analysis, we have chosen to study the binary mixture of NH₃ and H₂O. We find this mixture interesting from a recognition point of view because resonances at 0.572 THz (NH₃) and 0.557 THz (H₂O), and again at 1.168 THz (NH₃) and 1.163 THz (Ref. 11, H₂O) coincide substantially within the frequency resolution of our experiment. The result is shown in Fig. 4. We prepared the mixtures by starting out with NH₃ vapor at 12.9 kPa and successively exchanging it with H₂O vapor in portions of few percent per volume. For each mixture a waveform was recorded. From the LPC analysis and the geometric interpretation of the LPC vector we calculate the mole fractions of the two species. We

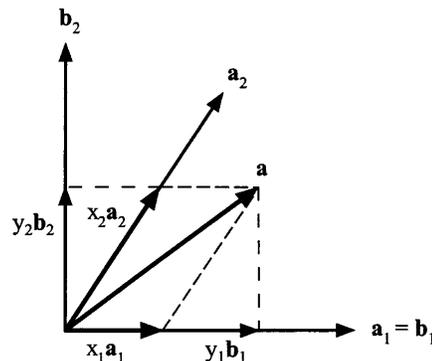


Fig. 1. Two-dimensional representation of a codebook. The vectors \mathbf{a}_1 and \mathbf{a}_2 represent two different gas species. From these vectors two orthogonal vectors, \mathbf{b}_1 and \mathbf{b}_2 , are constructed. Vector \mathbf{a} corresponds to a mixture that contains the fraction x_1 of species 1 and the fraction x_2 of species 2.

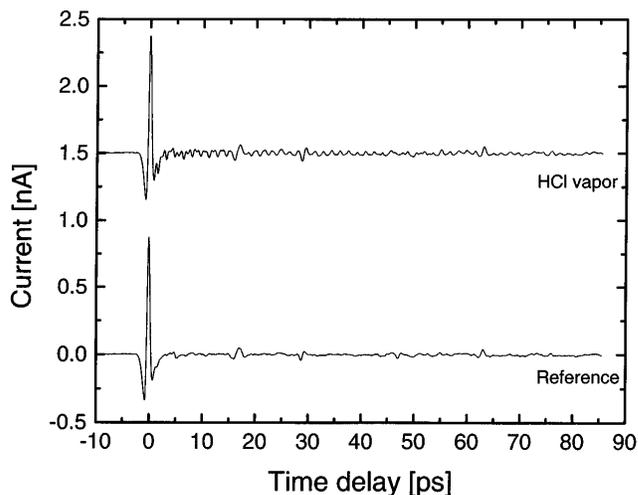


Fig. 2. THz waveform for HCl vapor at 13 kPa and a reference waveform measured for an evacuated gas cell. The HCl waveform has been shifted by 1.5 nA for clarity. The occurrences at 17, 29, and 47 ps are due to reflections in the setup.

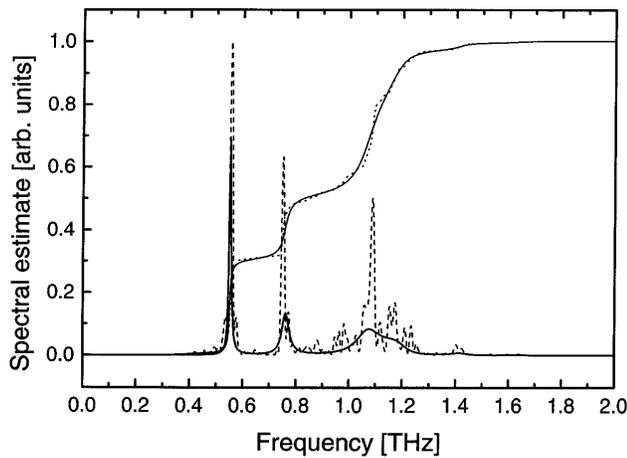


Fig. 3. Power spectrum estimated from a LPC analysis of order $M = 50$ (thick solid curve) and from a Fourier analysis (dashed curve) for H_2O vapor at 2.8 kPa. The thin solid curve and the dotted curve represent the integrated power spectral density for the LPC and the Fourier analysis, respectively.

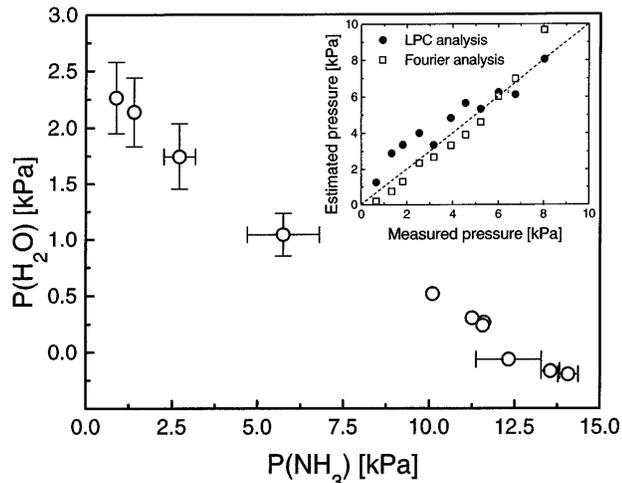


Fig. 4. Result of chemical recognition of mixtures of NH_3 and H_2O . For each mixture we plot a single point representing the estimated partial pressure of the two gas species. The inset shows the estimated versus the measured pressure from a single-species recognition of HCl.

convert the mole fractions into partial pressures by scaling to the pressure of the corresponding coded vectors. Nonvanishing projections onto species in the codebook that were not present in the mixture (HCl and CH_3CN) are attributed to experimental uncertainties and are used to estimate error bars. Error bars are displayed only when they become larger than the actual data points. The inset of Fig. 4 shows the result of a single-species recognition of HCl at various pressures. We used both LPC analysis and Fourier analysis to estimate the actual pressure of HCl. The dashed line shows the ideal case. In terms of getting quantitative pressure information

both techniques are useful but neither is significantly better.

From Fig. 4 we find that the chemical-recognition system is able to trace the transition from NH_3 -dominated mixtures (lower right) to H_2O -dominated mixtures (upper left). When a mixture is strongly dominated by NH_3 the prediction yields a slightly negative pressure for H_2O and a pressure for NH_3 that is larger than the starting pressure. Obviously, this situation does not have any physical meaning and results from insufficient statistics for the coded vectors. By use of clustering techniques⁸ of vectors obtained from repeated measurements on a particular gas species, we should be able to eliminate this inaccuracy.

In conclusion, we have demonstrated the use of THz waveforms for chemical recognition of gases and gas mixtures. The recognition system is based on a correlation type of analysis and is carried out in the time domain. Using a simple geometric picture, we were able to estimate the partial pressures of components in a gas mixture. We foresee potential real-world applications such as trace-gas sensing and industrial process control for our proposed chemical-recognition system.

We thank Y. LeCun for helpful discussions.

*Permanent address, Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark.

†Present address, Department of Electrical and Computer Engineering, Rice University, P.O. Box 1892, Houston, Texas 77251-1892.

References

1. P. R. Smith, D. H. Auston, and M. C. Nuss, *IEEE J. Quantum Electron.* **24**, 255 (1988).
2. M. van Exter and D. Grischkowsky, *IEEE Trans. Microwave Theory Technol.* **38**, 1684 (1990).
3. I. Brener, D. Dykaar, A. Frommer, L. N. Pfeiffer, J. Lopata, J. Wynn, K. West, and M. C. Nuss, *Opt. Lett.* **21**, 1924 (1996).
4. M. van Exter, Ch. Fattinger, and D. Grischkowsky, *Opt. Lett.* **14**, 1128 (1989).
5. H. Harde, S. Keiding, and D. Grischkowsky, *Phys. Rev. Lett.* **66**, 1834 (1991).
6. R. A. Cheville and D. Grischkowsky, *Opt. Lett.* **20**, 1646 (1995).
7. R. G. Brewer and R. L. Shoemaker, *Phys. Rev. A* **6**, 2001 (1972).
8. L. Rabiner and B.-H. Juang, in *Fundamentals of Speech Recognition*, A. V. Oppenheim, ed. (Prentice-Hall, Englewood Cliffs, N.J., 1993), p. 97.
9. J. Makhoul, *Proc. IEEE* **63**, 561 (1975).
10. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications* (Prentice-Hall, Englewood Cliffs, N.J., 1996).
11. H. M. Pickett, R. L. Poynter, and E. A. Cohen, *Submillimeter, Millimeter, and Microwave Spectral Line Catalog*, accessed from the Jet Propulsion Laboratory, Pasadena, California, via the World Wide Web (<http://spec.jpl.nasa.gov>).