

Exploring interaction effects in two-component gas mixtures using orthogonal signal correction of ultrasound pulses

Johan E. Carlson^{a)} and Pär-Erik Martinsson

EISLAB, Department of Computer Science and Electrical Engineering, Luleå University of Technology, SE-971 87 Luleå, Sweden

(Received 1 October 2004; revised 24 February 2005; accepted 28 February 2005)

Within Sweden and the EU, an increased use of biogas gas and natural gas is encouraged to decrease emission of carbon dioxide. To support more effective manufacturing, distribution, and consumption of energy gases, new methods for the measurement of the calorimetric value or the gas composition are needed. This paper presents a method to extract and visualize variations in ultrasound pulse shape, caused by interaction effects between the constituents of a two-component gas mixture. The method is based on a combination of *principal component analysis* and *orthogonal signal correction*. Pulse-echo ultrasound experiments on mixtures of oxygen and ethane in the concentration range from 20% to 80% ethane show that the extracted information could be correlated with the molar fraction of ethane in the mixture. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1893565]

PACS numbers: 43.60.-c, 43.20.Ye, 43.35.Yb [YHB]

Pages: 2961–2968

I. INTRODUCTION

Natural gas and biogas contain mixtures of several gases, with the major component being methane. Other components are ethane, hydrogen, and other higher order hydrocarbons (such as propane, butane, etc.). Sometimes the gas also contains small, but highly undesired, fractions of oxygen or water vapor. Water and oxygen are both corrosive and can cause severe damage to the pipeline systems. Therefore, knowledge and methods for monitoring the amount of undesired impurities in the gas are important.

Furthermore, with the use of different sources of gas (i.e., different gas fields) the energy content of the gas delivered to customers may vary considerably. Variations of as much as 20% can occur.¹ It is of interest to both provider and customer to know the composition of such gas mixtures, since this determines the energy content (calorimetric value) and, thus, the combustion properties and the monetary value of the gas.

Today, the energy content of gases is measured using either gas chromatography or calorimetry. Both methods are accurate, but require samples of the gas to be removed and analyzed separately, which makes them relatively slow. They are also rather expensive. Because of this, the existing techniques are not suitable for on-line measurement at the customer side of the distribution line.

In conjunction with flow meters the temperature and pressure are measured in order to relate the operating flow conditions to a reference condition. Given the calorimetric value at the reference condition, the energy content can be measured on-line, during operation. However, this approach assumes that the composition of the flow is constant. New methods for on-line measurement of the calorimetric value have been investigated by, for example, Jaeschke *et al.*^{2,3} Jaeschke modified the technique by adding measurement of

relative permittivity, speed of sound, and CO₂ molar fraction as input to a correlation model in order to improve the accuracy of the energy content estimate.

In this study, we investigate what information that can be extracted from the shape of ultrasound pulses that propagated through the gas. Today, ultrasonic flow meters are used to measure the volume flow, and a method that does not require many additional sensors would be attractive. With the approach presented here, the data available in a conventional flow meter are processed to extract information about the gas composition.

Typically, both the speed of sound and the attenuation of sound within a gas vary with temperature, frequency, pressure, etc. Hence, both these properties can help to monitor changes in experimental conditions. This has recently been studied in both theory and experiments by Dain and Lueptow,^{4,5} Martinsson,⁶ and Townsend and Meador.⁷

The frequency dependence of the attenuation is fairly easy to measure, but the speed of sound is much more difficult.⁸ Both of these affect the shape of the received pulse.

Figure 1 shows three pulses obtained using the experimental setup described in Sec. IV A. The first pulse was measured in pure oxygen, the second in pure ethane, and the third in a mixture of the two, containing molar fraction of 40% of ethane. As the figure shows, there is a small change in pulse shape between pure oxygen and ethane. It is, however, more difficult to notice how this changes when gases are mixed. The pulses in Fig. 1 have been normalized and aligned in time for the purpose of showing differences in shape. The remaining differences in pulse shape are caused by frequency-dependent attenuation and dispersion.

In this paper, we develop a subspace-based filter that can be used to suppress variations in pulse shape originating from the pure gases, hence, leaving only variations originating from interaction effects between the constituent gases. The filter is based on a principle known as *orthogonal signal*

^{a)}Electronic mail: johan.carlson@csee.ltu.se

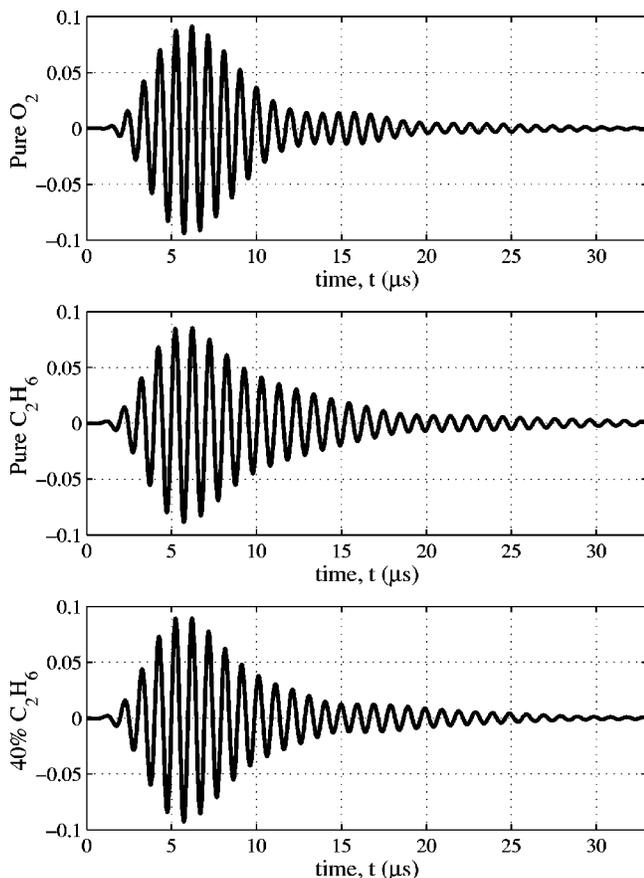


FIG. 1. Example of pulses measured in pure oxygen, pure ethane, and a mixture with a molar fraction of 40% ethane, all measured at a static pressure of 5.0 bar. The temperature was 20 °C. The pulses have been normalized to unit energy and aligned in time.

correction (OSC).⁹ Pulses measured in a mixture will be affected by the pure gases and by the inter gas interaction effects. These interaction effects are much smaller than the individual effects of the constituent gases, but highly interesting when the goal is to study the composition of a mixture. The OSC filter is implemented as a projection matrix that projects the measured pulses in gas mixtures onto the orthogonal complement of a basis spanning the experimental variation caused by the pure gases. The remaining experimental variation is then analyzed using *principal component analysis* (PCA).¹⁰ The work presented herein is an extension of the results presented at a recent conference.¹¹

The method is evaluated with experiments on pure oxygen (O_2), pure ethane (C_2H_6), and mixtures of the two for molar fractions of ethane in the range of 20%–80%. The

results show that the remaining variation can be explained by one principal component, which correlates well with the molar fraction of ethane.

II. NOTATION

We will use bold capital letters (e.g., \mathbf{X}) to denote matrices, and small bold letters to denote column vectors (e.g., \mathbf{x}_i , for column i of the matrix \mathbf{X}). Scalars and matrix row and column indices are denoted with small nonbold (italic) characters (e.g., i, k, n). For example, the first column of the matrix \mathbf{X} is denoted \mathbf{x}_1 . Matrix transpose is denoted with T . The notation $x[n]$ denotes the n th element of a vector.

III. THEORY

Any experimentally observed data, x , is composed of two parts: ξ , which is a systematic part, and ε , which is a random noise part, such that $x = \xi + \varepsilon$.

Principal component analysis (PCA)¹⁰ is a well-established tool for analyzing and modeling multivariate data. The central idea is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, but at the same time preserving as much as possible of the systematic experimental variation, ξ , in the data. Figure 2 illustrates the principle for a simple case where a three-dimensional data set is projected onto a two-dimensional subspace.

In order to analyze the shape of ultrasound pulses by PCA, we need to represent them in matrix form. Let $x_i[n]$ ($n=1, \dots, N$) be a sampled version of an ultrasound pulse and let \mathbf{x}_i be the column vector representation of the same pulse. Now each measured pulse can be seen as a point in an N -dimensional space. For each of the different experimental settings, a new vector is obtained, and in order to analyze the whole set by PCA, they are stored as columns of a matrix, \mathbf{X} .

In the experiments (see Sec. IV), we have two data sets, the first containing pulse-echo measurements of pure oxygen and pure ethane, and the second containing measurements on mixtures of the two. If the effects of oxygen and ethane were to add linearly when mixing the two, the dimensionality of the data set would not increase. That is, the same set of principal components would describe the experimental variation in both data sets. If, however, there are any interaction effects between the gases in the mixture, we would need some additional components to describe these (i.e., the dimensionality of the data set is increased). Now, if the interaction effects are small, which they most likely will be when

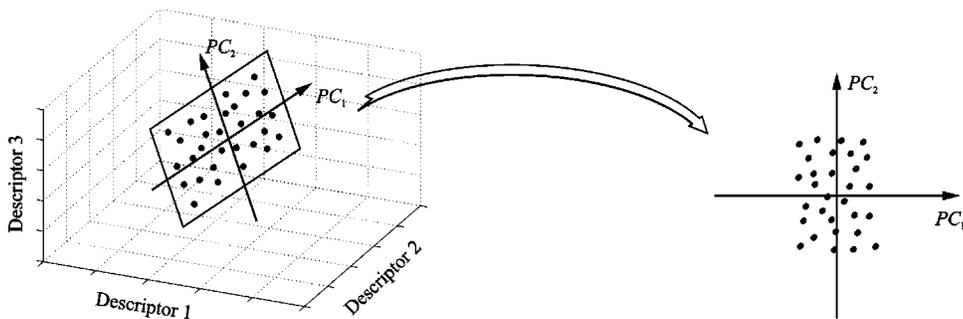


FIG. 2. Schematic description of how PCA reduces the dimensionality of a data set.

the molar fraction is small, they will be practically drowned by the effects of the pure gases. If the goal is to extract the effects of mixing gases, we need some method to remove the effects originating from the pure gases. The way we approach this problem is by *orthogonal signal correction* (OSC).⁹

First, a PCA is performed on measurements on pure gases. This results in a set of components that explain the variations in pulse shape caused by the pure gases, for the different experimental conditions. The pulses measured in gas mixtures will vary due to the individual constituents, but also due to interaction effects between the gases. The OSC process can be seen as a filter that removes the effects of the pure gases from the mixture data, leaving a data set that is uncorrelated (orthogonal) to the measurements in pure gases. The remaining variations in pulse shape originate from interaction effects between the constituent gases.

For this to work, the data set has to be preprocessed to remove the effect of propagation delay and scalar attenuation. This information is, of course, still accessible, but by normalizing the pulses, we do not *a priori* assess any greater statistical significance to any of the measured pulses. Pulses measured in different gas compositions will have a different propagation delay, because of variations in sound velocity. If the delays are not compensated for, this will result in an apparent increase in dimensionality of the data set (i.e., more principal components are needed). This is accomplished by aligning all pulses with respect to one of the measurements (e.g., the first column in the matrix containing all pulses).

The next subsection describes the preprocessing of the ultrasound pulses, and Sec. III B then describes the OSC and the PCA.

A. Preprocessing

The pre-processing consists of two steps:

- (1) Normalizing the pulses to unit energy, thus removing the effect of a scalar attenuation due to the propagation distance.
- (2) Aligning the pulses in time and thereby removing the effect of changes in propagation delay through the medium, and the effect of any sampling jitter caused by the digitizing hardware.

1. Normalizing

The normalized pulse, \mathbf{x}_i , is calculated as

$$\mathbf{x}_i = \frac{\bar{\mathbf{x}}_i}{\sqrt{\mathbf{x}_i^T \bar{\mathbf{x}}_i}}, \quad (1)$$

where $\bar{\mathbf{x}}_i$ denotes the vector containing a sampled version of the pulse, before normalizing.

2. Aligning pulses

Aligning the pulses consists of two steps:

- (1) Estimate time-of-flight differences between pulses, with an accuracy of fractions of the sampling time.
- (2) Align pulses according to the estimated time-of-flight difference.

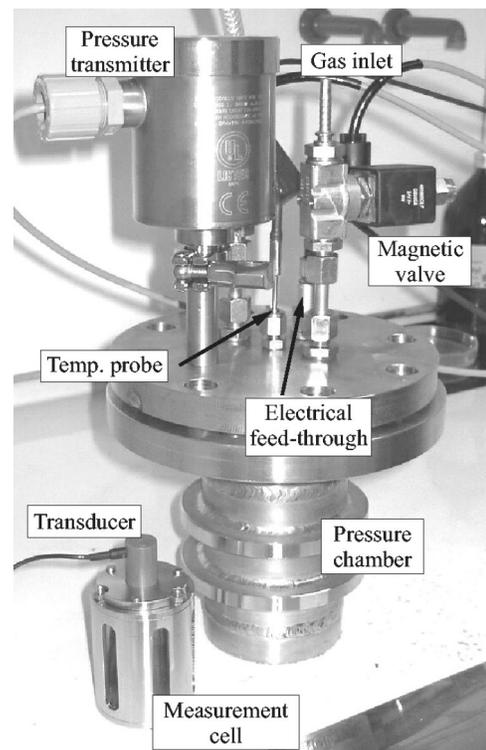


FIG. 3. Description of the measurement cell and the pulse-echo scheme. In this paper the first echo, as indicated in the figure, is analyzed.

In this paper we consider only changes in pulse shape of an ultrasound echo that propagated back and forth through the volume of gas (see Fig. 3). Because we only use the first echo (as indicated in Fig. 3), we do not measure speed of sound and attenuation explicitly, but rather changes in pulse shape caused by frequency-dependent attenuation and speed of sound. However, when the experimental conditions change, the time-of-flight through the gas will also vary. The set of echoes recorded for the different measurement configurations are therefore aligned, with respect to one of the echoes (e.g., the first measurement).

The reason for aligning and normalizing is to better reveal any changes in pulse shape. As a consequence of these changes, it becomes difficult to determine time-of-flight differences, since techniques for this are based on the assumption that the *only* differences between the pulses are a time-of-flight difference and possibly a scalar attenuation.

As a best-effort attempt, we use the analytic cross-correlation technique proposed by Marple¹² to estimate the time-of-flight differences in whole samples (i.e., an integer multiple of the sampling time of the system). Marple's method uses the maximum of the envelope of the cross-correlation function, instead of the cross-correlation function itself. The result is an estimate that is less sensitive to small changes in pulse shape than the standard cross-correlation method, since the envelope is much less sensitive to phase shifts. The details of this method are left out since they are very well described in the original paper by Marple.

A cross-correlation technique will result in an estimate of the time-of-flight difference as a multiple of the sampling time. Variations in propagation speed or jitter in the digitizing hardware can, however, result in delays on a subsample

level (i.e., fractions of the sampling time). To further improve the performance of the PCA, the time-of-flight difference estimates are therefore refined using the estimator proposed by Grennberg and Sandell.¹³ This estimator has been shown to work especially well for narrow-band pulses, and for small delays.

Aligning the pulses according to an integer delay (i.e., a multiple of the sampling time) is straightforward, but to shift pulses a fraction of the sampling time, we need to use some interpolation scheme. In this work we have chosen the method of Lagrange interpolation,¹⁴ since this can be implemented as a simple linear filter. Let the original pulse be $\bar{x}_i[n]$. Then, the shifted pulse, $x_i[n]$, is given by

$$x_i[n] = \bar{x}_i[n] * h_l[n, \theta], \quad (2)$$

where θ is the time-of-flight difference, $*$ denotes the convolution, and $h_l[n, \theta]$ is the impulse response of the interpolation filter, according to Lagrange's interpolation formula. For a three-point interpolation, the impulse response is given by

$$h_l[n, \theta] = \begin{cases} \frac{1}{2} \theta(\theta - 1), & n = 0, \\ (1 - \theta^2), & n = 1, \\ \frac{1}{2} \theta(\theta + 1), & n = 2. \end{cases} \quad (3)$$

B. Orthogonal signal correction

After the preprocessing, we are left with a set of measurements that essentially vary only in pulse shape. This section describes the remaining steps of the analysis:

- (1) Find a basis for the experimental variation caused by the pure gases (PCA), for different static pressures.
- (2) Project the measurements of mixtures onto the orthogonal complement of the basis determined in step 1 (OSC).
- (3) Find a new basis for the remaining experimental variation (PCA).

Let \mathbf{X}_0 be a matrix where the columns are pulses measured in pure oxygen (O_2) and pure ethane (C_2H_6), for different pressures. Let \mathbf{X}_1 be the matrix with columns corresponding to pulses measured in mixtures of the gases, each representing different molar fractions of ethane, also for different static pressures.

Finding a basis for the experimental variation spanned by the columns of \mathbf{X}_0 means determining the principal components (PCs) of \mathbf{X}_0 . In this paper, the PCA is implemented using *singular value decomposition* (SVD).¹⁵ With the SVD, any rank r matrix \mathbf{X}_0 can be factored as

$$\mathbf{X}_0 = \mathbf{U}_0 \mathbf{S}_0 \mathbf{V}_0^T, \quad (4)$$

where the columns of \mathbf{U}_0 and \mathbf{V}_0 are the unit-norm eigenvectors of $\mathbf{X}_0 \mathbf{X}_0^T$ and $\mathbf{X}_0^T \mathbf{X}_0$, respectively. The nonzero diagonal elements, σ_i of \mathbf{S}_0 are the square-roots of the eigenvalues of $\mathbf{X}_0 \mathbf{X}_0^T$ and $\mathbf{X}_0^T \mathbf{X}_0$, sorted so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$. This can be seen as a change of basis, where the new basis for the columns of \mathbf{X}_0 is given by the columns of \mathbf{U}_0 (also called *principal components* or *loading vectors*), and the weights for the new basis (*scores*) are given by the columns of $\mathbf{S}_0 \mathbf{V}_0^T$.

The change of basis means that all pulses (columns of \mathbf{X}_0) can now be represented as a linear combination of the columns in \mathbf{U}_0 , using the weights in $\mathbf{S}_0 \mathbf{V}_0^T$, i.e.,

$$\mathbf{x}_i = \mathbf{u}_1 \sigma_1 v_{i1} + \mathbf{u}_2 \sigma_2 v_{i2} + \dots + \mathbf{u}_r \sigma_r v_{ir}, \quad (5)$$

where r is the rank of \mathbf{X}_0 . If fewer components than r are used, this yields an optimal low-rank approximation of \mathbf{X}_0 (in a least-squares sense). This property is what is used when approximating a data set with a small number of principal components.

To examine how much of the total experimental variation (in %) each principal component explains, we study the scaled singular values, since the singular value σ_i is proportional to the experimental variation in the direction of the i th principal component.¹⁰ Let $\bar{\sigma}_i$ be scaled versions of the r nonzero singular values σ_i :

$$\bar{\sigma}_i^2 = 100 \frac{\sigma_i^2}{\sum_{k=1}^r \sigma_k^2}. \quad (6)$$

Now, if the effects of the constituent gases were to add linearly, the same set of principal components (columns of \mathbf{U}_0) would also span the variations in pulse shape caused by the gas mixtures. That is, the columns of \mathbf{X}_1 could all be written as linear combinations of the columns of \mathbf{U}_0 . If this is not the case, the interaction effects can be extracted from \mathbf{X}_1 by projecting onto the orthogonal complement of \mathbf{X}_0 as

$$\tilde{\mathbf{X}}_1 = \mathbf{\Pi}_{\mathbf{X}_0}^\perp \mathbf{X}_1, \quad (7)$$

where the projection matrix $\mathbf{\Pi}_{\mathbf{X}_0}^\perp$ is given by¹⁵

$$\mathbf{\Pi}_{\mathbf{X}_0}^\perp = \mathbf{I} - \hat{\mathbf{U}}_0 (\hat{\mathbf{U}}_0^T \hat{\mathbf{U}}_0)^{-1} \hat{\mathbf{U}}_0^T, \quad (8)$$

where $\hat{\mathbf{U}}_0$ is the matrix consisting of the n ($n < r$) most significant components of \mathbf{U}_0 . This is determined by looking at the cumulative sum of the scaled singular values from Eq. (6). Since the columns of $\hat{\mathbf{U}}_0$ are by construction orthonormal,¹⁵ Eq. (8) simplifies to

$$\mathbf{\Pi}_{\mathbf{X}_0}^\perp = \mathbf{I} - \hat{\mathbf{U}}_0 \hat{\mathbf{U}}_0^T. \quad (9)$$

A basis for the interaction effects is then given by the principal components of $\tilde{\mathbf{X}}_1$, given by the columns of $\tilde{\mathbf{U}}_1$, in

$$\tilde{\mathbf{X}}_1 = \tilde{\mathbf{U}}_1 \tilde{\mathbf{S}}_1 \tilde{\mathbf{V}}_1^T. \quad (10)$$

In other words, we can say that the columns of $\tilde{\mathbf{U}}_1$ form a basis for the experimental variation that remains after removing the contribution of the pure gases. The matrix $\mathbf{\Pi}_{\mathbf{X}_0}^\perp$ in Eq. (9) projects the data in \mathbf{X}_1 onto a smaller subspace, orthogonal to the subspace spanned by $\hat{\mathbf{U}}_0$. This is why this process is called *orthogonal signal correction*.⁹

C. Summary of the algorithm

The analysis principle described in the previous sections can be summarized as follows:

- (1) Store all pulses as columns of the matrix $\bar{\mathbf{X}}$.
- (2) Normalize and align the pulses in $\bar{\mathbf{X}}$ as described in Sec. III A in order to obtain \mathbf{X}_0 .

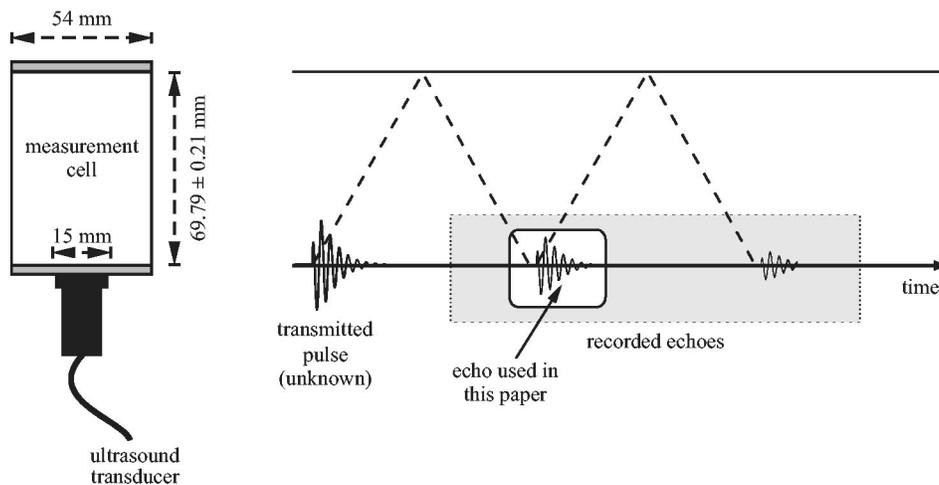


FIG. 4. The measurement equipment. All measurements were performed in a pressure chamber. The pressure was varied between 1.54 and 7.4 bar. The air transducer has a center frequency of 1 MHz.

- (3) Calculate principal components (\mathbf{U}_0) of the pure gases using Eq. (4).
- (4) Remove the effect of the pure gases from the gas mixture using Eq. (7).
- (5) Determine the principal components, $\tilde{\mathbf{U}}_1$, of the remaining variation ($\tilde{\mathbf{X}}_1$) using Eq. (10).

IV. EXPERIMENTS

A. Setup

A pulse echo scheme was used to acquire ultrasound pulses that propagated through the gas mixtures. The acoustic properties of interest vary with frequency, f , and pressure, P . In fact, it is the frequency to pressure ratio, f/P , which is the appropriate scale to use.⁴ The temperature and frequency dependence of sound velocity and acoustic attenuation have previously been investigated by, for example, Martinsson,⁶ Lueptow,^{4,5} and Bhatia.¹⁶ There are two ways to vary the f/P ratio in a pulse echo system. Transducers of different center frequency can be used for a fixed pressure, or one transducer can be used while the pressure is changed. The latter of the two principles was chosen for the work in this paper. A 1-MHz air transducer was used. The effective diameter of the transducer was 15 mm (cf. Fig. 3). Diffraction losses were assumed to be negligible, since the pulse-echo measurements are in the near-field region of the transducer.¹⁷

A custom-built pressure chamber (see Fig. 4) was used to achieve different static pressures. The pressure was varied between 1.54 to 7.4 bar in 12 steps for each gas. Since the attenuation in ethane is extremely high at low pressures and high frequencies, we were limited to make measurements at higher pressure for that particular gas (above 1.86 bar).

The pressure in the chamber was measured with an *ANDERSON TPP Pressure Transmitter* with a range of up to 13.6 bar above atmospheric zero. The transmitter has an accuracy of approximately 30 mbar. This includes the combined effects of linearity, hysteresis, and repeatability.

The transducer was mounted on a stainless steel measurement cell, as seen in the lower left corner of Fig. 4. The measurement cell was then immersed into the pressure chamber. Figure 3 shows the details of the measurement cell. For all calculations in this paper, only the first of the recorded echoes was used, as indicated in the figure. The whole setup

was then placed in a temperature controlled chamber (*Heræus Vötsch HT4010*), where the temperature was kept constant at the desired temperature.

To excite and receive acoustic pulses from the transducer, a *Panametrics Pulser/Receiver Model 5072* was used. For the transmitting mode, the pulser/receiver was set to deliver maximum energy to the transducer, which corresponds to a short voltage impulse with 360-V amplitude, corresponding to an excitation energy of 104 μJ . In receive mode, the signals are amplified 20 dB.

All pulses were sampled at 100 MHz with an 8-bit *Tektronix TDS 724*, 1-GHz digitizing oscilloscope. For each experimental setting, 50 signals were recorded and transferred to a PC, where they were averaged and further processed (see Sec. III A).

For each measurement, the temperature was recorded using an encapsulated PT100 sensor mounted through the wall of the pressure chamber.

The transducer used was originally designed for operation in air which has an acoustic impedance $Z_{\text{air}} = 4154 \text{ Pa}\cdot\text{s/m}$ (at $T = 20^\circ\text{C}$ and $P = 1 \text{ bar}$). Since the acoustic impedance of the gases used are different from air, the transducer will not operate at its optimal performance, and consequently not transmit as much power as desirable.

B. Results

The first set of experiments was with pure ethane and pure oxygen, at 20°C , for pressures of 1.54–7.4 bar. This resulted in 12 measurements of ethane and 12 of oxygen. After aligning and normalizing, as described in Sec. III A, all 24 were stored as columns of the matrix \mathbf{X}_0 . The second set of experiments was with mixtures of ethane and oxygen, for molar fractions of 20%, 40%, 60%, and 80%, for the same pressure range, also at 20°C . In total $4 \times 12 = 48$ experiments were made on the mixtures. These pulses were aligned and normalized and stored as columns of the matrix \mathbf{X}_1 .

Looking at the cumulative sum of the squares of the scaled singular values (cf. Fig. 5), $\bar{\sigma}_i^2$, of \mathbf{X}_0 we see that the first three components are enough to describe approximately 99% of the total experimental variation. The rest of the components are assumed to represent mainly noise. Figure 6 shows the corresponding *principal components*, \mathbf{u}_i (i

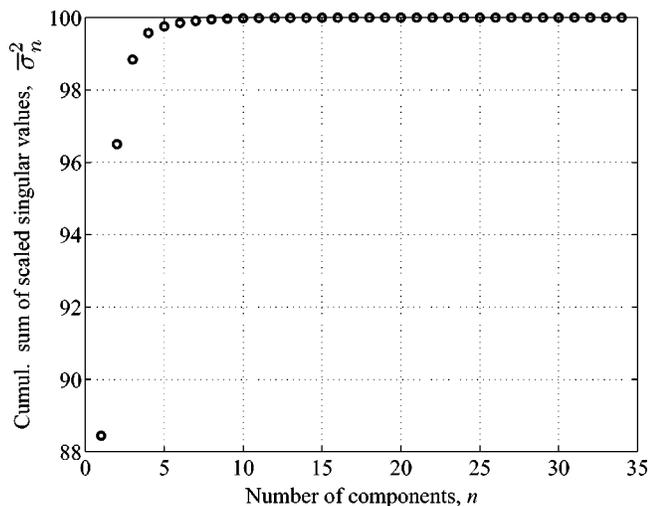


FIG. 5. Cumulative sum of the singular values, $\bar{\sigma}_n^2$, of \mathbf{X}_0 (i.e., eigenvalues of $\mathbf{X}_0^T \mathbf{X}_0$), describing the amount of experimental variation explained as components are added to the model [see Eq. (6)]. Note that the first three components are enough to describe approximately 99% of the total variation.

$= 1, 2, 3$). Since these three components account for approximately 99% of the experimental variation, this means all measured pulses can be accurately approximated as a linear combination of these, as given by Eq. (5).

Following the steps described in Sec. III C, we obtain the *score matrix*, $\tilde{\mathbf{S}}_1 \tilde{\mathbf{V}}_1^T$, and the *loading matrix*, $\tilde{\mathbf{U}}_1$.

The data set for the gas mixtures was then filtered using the OSC filter given by Eq. (7). In order to study how much systematic variation that remains after the OSC filter was applied, a PCA was performed on $\tilde{\mathbf{X}}_1$. From Fig. 7 we see that the remaining variation is significantly smaller. Determining the significance of these components is more diffi-

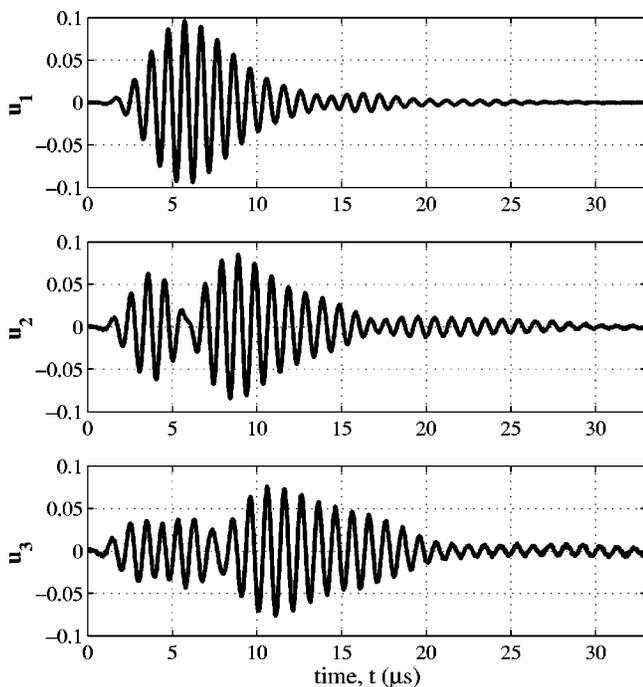


FIG. 6. First three loading vectors, \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , representing approximately 99% of the total experimental variation.

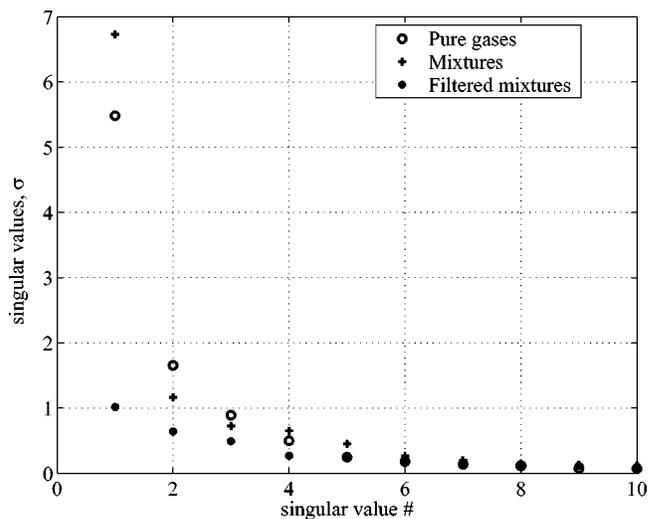


FIG. 7. Singular values of the pure gases (\mathbf{X}_0), the mixtures \mathbf{X}_1 , and the mixtures after projection $\tilde{\mathbf{X}}_1$. It is clear that the pure gases contribute to most of the variation. The filtering reveals the relatively small interaction effects.

cult, since the singular values do not vary as much as for the unfiltered mixtures. There are two possible reasons for this:

- (1) The components contain mainly noise.
- (2) All components contain important variation.

The second is not likely, since we know already that approximately 98% of the original variation has been removed. To determine whether the remaining components represent any systematic variation, we need to study the scores associated with each of the components. In Fig. 8 these are shown for the first component. These show a clear interdependence with the physical variables of interest (i.e., f/P ratio and molar fraction). Studying the scores of the other components did not show the same systematic relationship to the molar fraction, and therefore no figures of this have been included in the paper.

Determining the number of necessary principal components can be done in several ways. In this paper we deter-

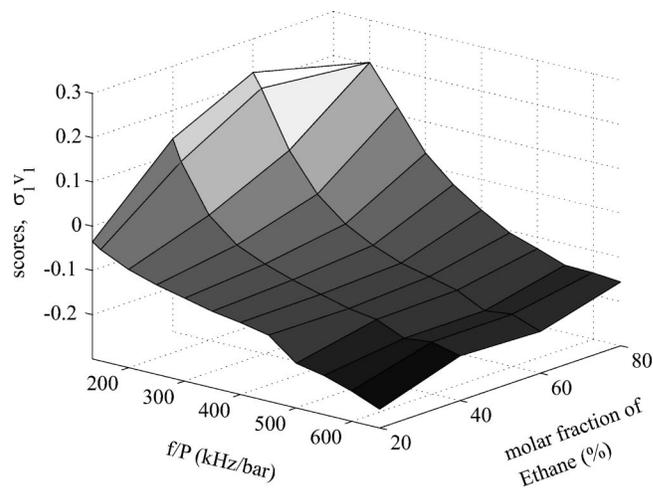


FIG. 8. Scores (coefficients), $\tilde{\mathbf{v}}_1 \sigma_1$, of the first principal component of the filtered data set as a function of frequency to pressure ratio and molar fraction of ethane.

mined this by looking at how many components are needed to capture >98% of the experimental variation, using Eq. (6). Depending on the purpose of the analysis (e.g., regression or classification) other methods such as cross validation¹⁸ or the Akaike information criterion¹⁹ could be used.

In Fig. 7 it appears that for the mixtures some of the singular values are smaller than the corresponding values for the pure gases, and vice versa. Some questions were raised whether this contradicts the conclusion that the pure gases contribute to most of the systematic variation. The pure gases and the mixtures are two different data sets, and the intrinsic structures of these sets may vary, and thus the relative size of the singular values can differ. Looking at the singular values of the filtered mixtures, it is, however, clear that most of the variation has been removed. The remaining set is uncorrelated with the pure gases and the conclusion, and thus the remaining variation is due to noise and intergas interaction effects.

Figure 8 shows the *scores* (or coefficients) $\tilde{v}_1\sigma_1$ of the first component for the filtered mixture data as function of frequency to pressure ratio (f/P) and molar fraction of ethane. For high f/P values (i.e., low static pressures), the acoustic attenuation is very high, and the signal-to-noise-ratio (SNR) drops. For these f/P values, the intergas interaction effects are almost drowned by the experimental noise, and thus become difficult to extract. This can be seen in Fig. 8 as the surface levels out, i.e., the scores describe mainly noise. For higher SNR, there is visible interdependence between the molar fraction of ethane and the scores of the first principal component. To estimate molar fractions, the surface in Fig. 8 could be modeled as a function of the frequency to pressure ratio and molar fraction.

V. DISCUSSION

A first step towards a physical model of a complex system is to identify variables that affect the experimental variation. The methodology described in this paper helps revealing small, but highly interesting properties of gas mixtures, related to intergas interaction effects. The results in Fig. 8 show that these properties vary with the physical properties of interest. It is reasonable to believe that the changes in pulse shape are due to dispersion and frequency-dependent attenuation. For some individual gas components, these effects are predictable using physical models. It is, however, much more difficult to describe the physics of the interaction effects. The focus of this paper is to show that these effects can be observed from measured data, but we have not yet attempted to describe the underlying physical processes. Although the surface in Fig. 8 could be modeled empirically, a good physical model would be more attractive. Current research focuses on studying these phenomena, but to date it is not possible to draw any conclusions regarding this. The physical principles of wave propagation in gases are described in detail elsewhere, and since we do not use the physical models here, these details have been left out. The interested reader is referred to the work by Dain and Lueptow^{4,5} or the classic by Bhatia.¹⁶ A challenging problem

for future research is to find means of extracting systematic variation from observed data that can be used to estimate parameters in physical models.

In this paper we have only considered a two-component gas mixture. The long-term goal is to develop method for on-line measurement of the energy content of energy gases. In practice, these are always mixtures of more than two gases. Although molar fractions would enable us to calculate the energy content, it might be easier to quantify the energy content directly, using the same methodology as described in this paper.

In Sec. III A, we mention the importance of aligning the pulses in time before processing. If this is not done, the time delays will cause the apparent increase in dimensionality of the data set. Compensating for the time delays is, however, rather difficult. The change in pulse shape that we are interested in extracting is also a source of error when we try to estimate the time delays. The reason for this is that the time delay between pulses with different shape is not well defined. This is a well-understood problem that future research will have to address (see Ref. 20 and references therein).

It should also be mentioned that, except for the methodology described here, it is also possible to run a PCA directly on a matrix containing measurements on both pure gases and gas mixtures. The principal components obtained by this procedure also contain information about the gas composition.²¹ The OSC principle does, however, better isolate the specific interaction effects caused by mixing the gases.

VI. CONCLUSIONS

In this paper we have described how OSC and PCA can be used to extract information from ultrasound pulses that vary with the molar fractions of ethane in an ethane/oxygen mixture. OSC enables us to describe variations in pulse shape caused by interaction effects between the gases that cannot be described by the pure gases alone. The OSC can be interpreted as a subspace-based filtering process that removes experimental variation described by the pure gases, thus revealing the intergas interaction effects. These effects are significantly smaller than the effects of the pure gases and, as indicated by Fig. 1, very difficult to estimate from the measured pulses directly. The variation remaining after OSC filtering was analyzed with PCA, and the results show that one principal component is sufficient to describe most of the interaction effects, and that this shows a visible interdependence with the molar fraction of ethane in the two-component gas mixture.

ACKNOWLEDGMENTS

The authors wish to express their sincerest gratitude toward Professor Anders Grennberg and to Professor Rolf Carlson, for their valuable input, and to Professor Jerker Delsing, for supporting this work. Generous grants from the *Kempe Foundation* and *Swedish Energy Agency* are also gratefully acknowledged. Finally, the authors like to thank the reviewers for their valuable comments leading to significant improvements of the manuscript.

- ¹J. Delsing and I. Blom, "On-Line Measurements of Energy Content of Bio Gas and Natural Gas Mixtures," Technical Report No. ISSN: 0282-3772, Lund Institute of Technology (unpublished).
- ²M. Jaeschke, "Thermodynamic research improves energy measurement of natural gas," *Thermochim. Acta* **382**, 37–45 (2002).
- ³M. Jaeschke, P. Schley, and R. J. van Rossmalen, "Thermodynamic research improves energy measurement in natural gas," *Int. J. Thermophys.* **23**(4), 1013–1031 (2002).
- ⁴Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—theory," *J. Acoust. Soc. Am.* **109**, 1955–1964 (2001).
- ⁵Y. Dain and R. M. Lueptow, "Acoustic attenuation in three-component gas mixtures—Results," *J. Acoust. Soc. Am.* **110**, 2974–2979 (2001).
- ⁶P.-E. Martinsson and J. Delsing, "Ultrasonic Measurements of Molecular Relaxation in Ethane and Carbon Monoxide," in *Proc. IEEE Int. Ultrason. Symp.* (IEEE, Munich, Germany, 2002), pp. 494–499.
- ⁷L. W. Townsend and W. E. Meador, "Vibrational relaxation and sound absorption and dispersion in binary mixtures of gases," *J. Acoust. Soc. Am.* **99**, 920–925 (1995).
- ⁸J. E. Carlson, F. Sjöberg, and P.-E. Martinsson, "A Noise-Tolerant Group Delay Estimator Applied to Dispersion Measurement in Gases," in *Proc. IEEE Int. Ultrason. Symp.* (IEEE, Honolulu, HI, 2003), pp. 254–257.
- ⁹S. Wold, H. Antti, F. Lindgren, and J. Öman, "Orthogonal Signal Correction of Near-Infrared Spectra," *Chemom. Intell. Lab. Syst.* **44**, 175–185 (1998).
- ¹⁰I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer Verlag, New York, 2002).
- ¹¹J. E. Carlson and P.-E. Martinsson, "Ultrasonic Measurement of Molar Fractions in Gas Mixtures by Orthogonal Signal Correction," in *Proc. IEEE Int. Ultrason. Symp.* (IEEE, Montréal, Canada, 2004), pp. 821–825.
- ¹²S. L. Marple, Jr., "Estimating Group Delay and Phase Delay via Discrete-Time "Analytic" Cross-Correlation," *IEEE Trans. Signal Process.* **47**(9), 2604–2607 (1999).
- ¹³A. Grennberg and M. Sandell, "Estimation of Subsample Time Delay Differences in Narrowband Ultrasonic Echoes Using the Hilbert Transform Correlation," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **41**(5), 588–595 (1994).
- ¹⁴*Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, Dover Books on Advanced Mathematics*, 8th ed., edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).
- ¹⁵G. Strang, *Linear Algebra and its Applications*, 3rd ed. (Harcourt Brace Jovanovich, San Diego, 1986).
- ¹⁶A. B. Bhatia, *Ultrasonic Absorption—An Introduction to the Theory of Sound Absorption and Dispersion in Gases, Liquids and Solids* (Dover, New York, 1985).
- ¹⁷G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
- ¹⁸J. Shao, "Linear model order selection by cross-validation," *J. Am. Stat. Assoc.* **88**(422), 486–494 (1993).
- ¹⁹H. Akaike, "A new look at statistical model identification," *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
- ²⁰R. L. Trousil, K. R. Waters, and J. G. Miller, "Experimental validation of the use of Kramers-Kronig relations to eliminate the phase sheet ambiguity in broadband phase spectroscopy," *J. Acoust. Soc. Am.* **109**(5), 2236–2244 (2001).
- ²¹P.-E. Martinsson and J. E. Carlson, "Investigating the Feasibility of Using Principal Component Analysis for Ultrasonic Classification of Gas Mixtures," in *Proc. IEEE Int. Ultrason. Symp.* (IEEE, Honolulu, HI, 2003), pp. 1396–1399.