**Target audience of paper: scientists wanting to reproduce this work, professionals interested in gas analysis**

**Include only:**

* **Things I can defend!**
* **Everything necessary to reproduce and improve this work** 
  + **Describe own work extensively**
  + **Describe others’ work as necessary**
* **Value of this work and experiments as a whole: place of this work in the scope of the entire project, competition, value proposition, customers (all blood-diagnosed disease specialists, other markets?), key partners and contributions (QCL OEM, MATLAB, any company whose products are used)**

**Intro to gas analysis and chemometrics**

~~Why and how gas analysis and why use QCL spectroscopy~~

**SHOW STEP BY STEP HOW THE SIGNAL IS PROCESSED AND WHAT VARIABLES INFLUENCE THE SIGNAL.** Reference to Adonis/Zhe for their elimination of noise by each of these variables. This gives a clear overview of all noise sources and shows which still need work. Show clearly why every calibration step is taken, and what the result is.

**Experimental setup**

~~Describe system on the basis of figure~~

~~Use Adonis’ paper, Zhe’s thesis, questionspaper~~

~~Mention resolution~~

**From intensity to absorbance and Data calibration**

~~Wavenumber calibration – (compare with Adonis’) sound system calibration~~

Start with diagram showing overview of process. (Mention 10 measurements for each signal)

Compare st.dev. of 20 monitors before and after, and the absorbance (and st.dev.) before and after calibration

~~H2O & CO2 calibration~~

**Concentration and compound determination**

~~Finding p-values + Explanation of p-values~~

~~Determine compounds with high intensity in regions of low p value~~

~~Determine compounds without need for p-values (and explain pros/cons)~~

~~Determine concentration from compounds~~

**~~Remove CO2 in order to better estimate underlying ethyl-alcohol~~** ~~(check Adonis’ Analytical Chemistry paper for theory/understanding. Did this, Adonis’ paper uses the removal, but doesn’t state if or why it should improve concentration detection.)~~

**~~Compound\_filter.m~~** ~~Is put in subsection of molecule determination.~~

**Clustering and classification**

Got 80% accuracy on sorting healthy vs. asthmatic using SVM

**Appendix**

List of molecules: Standard, chosen by p-region, total list of all molecules processed.

Info on the samples processed: Measurements were taken by [WHOM?] of children in the [SOPHIA KINDERZIEKENHUIS] in ROTTERDAM. 70 breath samples of healthy children and 70 breath samples of asthmatic children were used. Most children had two samples taken which are treated as separate measurements, so the measurements originate from about 70 children total.

Data as found in between algorithms such as molecule lists, wavenumber vector range and spacing, and regions of low p-values. Check separate paragraphs for what input/output they use or reference to.

**Discussion**

\section{Wavenumber calibration}

Even after the wavenumber calibration the resulting absorbance spectrum doesn’t match with \(CO\_2\) and \(H\_2O\) peaks from literature. This is caused by the high variation of the twenty monitor measurements of which the mean is the basis of the wavenumber calibration. Regardless, wavenumber calibration is still useful since it is known that the different measurements do not match even though they should. The measurements should match since a cause of the variation between measurements is identified to be the piëzo element and its hysteresis, and this is expected to be the only major contributor.\\

\begin{comment}

The goal of the wavenumber calibration is to decrease the uncertainty of the absorbance, which in turn decreases the uncertainty of the determined concentrations. Discuss why st.dev. of absorbance decreases the amount it does. \\

\end{comment}

\section{Determination of molecules}

A general disadvantage of choosing molecules based on their molar absorptivity is that some molecules might have a large impact on the eventual absorbance due to a high concentration, despite having a low absorptivity. They would therefore be falsely eliminated from the list of molecules to check for.\\

There is also the drawback of excluding molecules which might be an important part of the absorbance, but aren’t selected since they do not have adequate molar absorptivity in the p-regions. This makes the subsequent determination of concentrations less accurate. A trade-off can be made by allowing for more p-regions, either by decreasing the minimum amount of consecutive points needed or heightening the maximum p-values of which these regions consist. This way more molecules will have some absorptivity in one of the regions and therefore won’t be scrapped. \\

Both problems to do with selecting molecules based on intensity can be resolved in advance by putting molecules that play an important role in the eventual absorbance in the standard list of molecules. This requires a-priori knowledge of the important molecules which is obtained from literature\cite{a review of volatiles} and simply by performing the least squares method to get the concentration on a large list of molecules. The concentrations won’t be as accurate for a large list, but all molecules that contribute a lot to the spectrum can be picked out for later runs with strict molecule lists.

\section{Determination of concentrations}

From the shapes of $CO\_2$ and $H\_2O$ in \autoref{fig:4.1} and autoref{fig:4.2} it can be seen that the concentrations calculated using the \textit{lsqnonlin} function are definitely in the right order of magnitude. The signal-to-noise for the rest of the measured absorbance is however still quite large, and the calculations for the molecules with lower concentrations are therefore not very accurate.

[Placeholder barplot of concentraions with st. dev.]

Currently the concentration determination is not yet accurate enough to use for distinguishing between healthy and asthmatic children, as can be seen in \autoref{barplot of concentrations with st. dev.}. The uncertainty of the measurements themselves can be improved by finding a moving element of which the displacement over voltage can be better characterized than the current piëzo. Alternatively an attempt can be made to calibrate the signals as they come straight from the measurement apparatus, before even the pre-calibration.

\section{Categorizing health status by classification}  
The accuracy of the classification can possibly be improved by looking more closely into the used algorithms, and improving them using a-priori knowledge of the samples such as regions of relatively high signal-to-noise ratio. Wavenumber regions can be given weights based on the signal-to-noise ratios so the algorithm depends more heavily on these accurate regions.

The error rate could also tell a lot more if the amount of false positives and false negatives is given. This would give separate error rates based on whether the algorithm judges the sample as healthy or asthmatic. It is possible that one of these error rates is much lower than the other, thereby increasing the certainty of the result above the total error rate.

\section{noise}?

Cause of noise (in rawest signal): Thermal fluctuations expand laser box(, causing different wavelengths to be stable and emitted, seen in the signal as intensity fluctuations? [source: Adonis]). Piëzo hysteresis causes wavenumber (horizontal) shift and stretch. Mode-hopping is caused by piëzo + grating allowing short periods were multiple wavelengths are supported [source: Adonis]. Mode-hopping is also caused by random gas fluctuations?

In order to achieve higher accuracy with the QCL setup the intensity fluctuations should be minimized. The main cause X could be replaced by …. The wavenumber shifting and stretching caused by the hysteresis of the piëzo can be fixed using a similar method as for the wavenumber calibration.

Optional/what else was done during this project

Get NEC as in Florian’s paper

Separate CO2H2O concentration determination

Helped Adonis filter molecules from organic molecule paper: Compound\_filter.m was used to find which of 199 molecules of paper found by adonis has absorption in the 1020-1100 cm-1 wavenumber region. These were 69 molecules. Can be found in email correspondence of 26/11/2014 “I'm working from home.” What was done with these molecules eventually? Used as standard molecules, or added to large collection of 456 molecules?

Learned R: reading and editing of Erasmus people their exhaled.r script.

Learned Python: used for machine learning data ordering and CO2 removal script.

Wrote remove\_CO2\_get\_concentration.py script in python.

Took Optical fabrication course