People Stink!: Towards identification of people from breath samples

Katri Salminen
School of Industrial Engineering
Tampere University of Applied Sciences
Tampere, Finland
katri.salminen@tuni.fi

Jussi Rantala
Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
jussi.rantala@tuni.fi

Philipp Müller
Faculty of Information Technology and Communication Sciences
Tampere University
Tampere, Finland
philipp.muller@tuni.fi

***Abstract*— The paper addresses the potential to use breath samples for identifying people. Participants were asked to exhale ten times for a length of five seconds to a tube attached to a commercial ion-mobility spectrometry device on three separate sessions. The data of each participant was divided into training (50% of the samples) and test data sets (50% of the samples) in random order. Classification decision tree (CDT), K nearest neighbor (KNN), naïve Bayes (NB), linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA) were used to analyze if the data could be classified correctly. Within a session, KNN (75.2%), NB (78.3%), and LDA (85.8%) were able to identify participants. Between sessions, the performance decreased.**

Keywords—* *Ion-mobile spectrometry, machine learning, multimodality, scent

I. Introduction (*Heading 1*)

The current paper investigates the use of the scent of breath to identify people. Classification of exhale breath samples is based on bacteria in the mouth and tongue that emit VOCs. While issues like diet or diseases can affect the smell of breath, there is also person-specific stability in the composition of breath that lasts over time [1, 2]. The exhale breath is most often analyzed by performing a chemical analysis which enables identification of a person with an accuracy of over 90% [3]. However, often chemical scent analysis is impractical. Taking samples and storing them, conducting the chemical analysis in a laboratory to decide the composition of the sample, and only then applying analysis is time-consuming and requires expertise in chemistry and laboratory measurement techniques [4]. To be able to proceed towards applications for human-computer interaction, the methodology to collect and analyze samples needs to be fast, simple and based on commercial devices.

The aim was to present a novel method to classify exhale breath samples to investigate the potential to identify people. The breath samples from one participant were collected on three different days so that it was possible to analyze the stability of the identification results over time. Exhale breath was chosen as an input method for three reasons. First, exhale breath measurement requires less preparation from the participant than body scent analysis [5, 6], where the person needs to shower and avoid odorants before measurements. Second, exhale breath can vary in volume, humidity, and temperature, making classification technically challenging in comparison to industrial cases with controlled scent presentation. Third, chemical analysis of exhale breath has been used to track down hormonal activity [7], lung cancer [8], the amount of acetone related to blood sugar level in diabetes [9], and lipid peroxidation [10] suggesting that there are several use cases beyond user identification and cryptography.

The procedure was following. A handheld ion-mobility spectrometry (IMS) device [11] was used to measure exhale breath from four participants in three separate sessions. The logging system saved the sensor readings before, during, and after an exhale to produce a "breath print". The breath prints were analyzed with five machine learning algorithms to compare their success in participant identification. Limitations of the study such as the effect of environmental factors like room air quality [2, 12] on classifier performance will be discussed.

II. Methods

A. Participants

A total of four fully-informed participants (2 females, 2 male) took part in three measurement sessions. The participants were non-smokers and had no medication or dental problems to ensure that the potential differences in breath prints were not due to, for example, the scent of tobacco. The participants were absent from consuming food or beverages other than water for an hour before the study.

B. Apparatus and Procedure

An IMS-based ChemPro 100i [11] by Environics Ltd. was used for the collection of the breath samples. The device uses an airflow of 1.3 l/min to suck in air with the help of a rotary vane pump. The molecules in the air are ionized using a radioactive source in the ionization chamber. The airflow is then pushed through an electric field where the ions come in contact with one of 14 electrodes. For each breath sample, the currents measured by all 14 electrodes were used as a sample (rate 1 Hz). The breath samples were collected in a ventilated laboratory. Indoor temperature and humidity were not controlled. Participants exhaled 10 times for five seconds near a small plastic tube attached to the ChemPro 100i. The inter-breath-interval was one minute. The participants were invited for three visits on three days to see if there are changes in the breath over time that affect the classification results. For each participant we had three datasets with each 10 exhales (a total number of 120 breath samples). Each visit took less than 30 minutes. No nose clip, mouthpiece, or VOC filter was applied during sample collection.

III. Data Analysis and Results

Figure 1 shows four examples for the 14-dimensional "breath prints". The first two breath prints (blue and red) are consecutive samples from the first measured exhale of the first participant, measured one second apart. There are considerable differences in the currents measured by most electrodes. The third and fourth breath prints (yellow and purple) are consecutive samples from an exhale of the second participant. The measured currents are noisy and react to exhales by either increasing or decreasing. Once the participant had stopped exhaling, the measured current returned to the pre-exhale level. To capture the reaction to the exhale and the return to baseline, sequences of 10 seconds per exhale were used even though exhales took only five seconds. Two analyses were carried out to check whether a person can be identified. Five widely applied, basic machine learning classifiers were used: Classification Decision Tree (CDT), K Nearest Neighbors (KNN), Naïve Bayes (NB), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis. The analyses used MATLAB R2018b.

The classifiers can be characterized as follow: CDT (MATLAB function fitctree) generates a binary tree based on the predictors (here readings from the 14 electrodes), and then follows a path in the tree from the root to a leaf to label a breath sample from the participant that has to be identified. The label is, e.g., ‘participant 1’. Training time for a CDT classifier is proportional to the number of training samples n (here the number of breath prints), and prediction time, i.e. time to find a label for an unlabeled breath print, is fast. For CDT the model containing the optimal sequence of pruned subtrees was used. The idea behind KNN classifiers (fitcknn) is to find the K labeled breath prints most similar to an unlabeled breath print and label that breath print based on the label that is most common amongst the obtained K labeled breath prints. In this paper Euclidean distance was used for measuring the similarity and K=3. The used KNN classifier does not require any training and prediction time is proportional to n^2. NB (fitcnb) assumes that the data has an underlying probability distribution, which can reduce the influence of outliers on the classifier. For the analyses kernel was used as assumed underlying distribution. A crucial feature of NB is that it assumes that the predictors are independent within each class. However, in [13] it was shown that this assumption does not hold for measurements from the ChemPro100i. Discriminant analysis, similar to NB, assumes that observations in each prediction class (here participant ID) can be modeled with a Gaussian distribution. However, the assumption of independence in each predictor is dropped. Thus, a multivariate Gaussian distribution is fitted to each class. Training time for the classifier and prediction time are both proportional to the size of the data set. LDA (fitcdiscr with 'DiscrimType' set to 'pseudolinear') assumes the covariance of each prediction class to be the same, which results in linear boundaries between the classes.

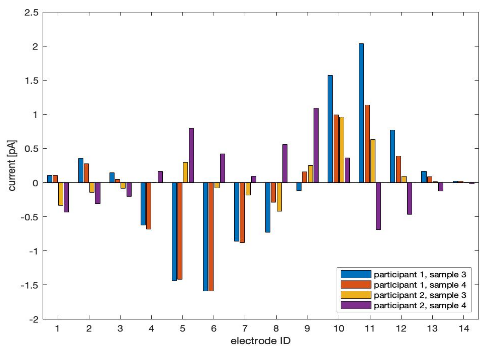

Fig. 1: Examples of 14-dimensional breath prints

QDA (fitcdiscr with 'DiscrimType' set to 'pseudoQuadratic') removes the assumption of equal covariances, which results in quadratic boundaries between classes.

In the first analysis, only data from the first measurement session of each participant was used. All five classifiers were trained using five breath samples per participant (i.e. 20 breath samples in total). The remaining 5 samples per participant were used for testing the classifiers. A leave-p-out cross-validation approach was used, meaning the procedure was repeated C(10,5)=252 times. Thus, each time different training and test sets were used. Each classifier returned for each exhale in the test set a label, which then was compared with the true, known participant identifier. Table 1 shows the percentage and average number of correctly classified breath samples as well as the corresponding standard deviations over the 252 repetitions. LDA was the best method as it classified on average 17.16 of the 20 test samples correctly. It also had with 1.60 the second-smallest variation in the number of accurately classified test samples. KNN and NB also performed well, while CDT showed only mediocre identification performance. QDA showed to be an inappropriate choice for identifying a person based on a short breath sample.

TABLE I. RESULTS OF THE FIRST DAY

Classifiers	***CDT***	***KNN***	***NB***	***LDA***	***QDA***
Accuracy	59.8%	75.2%	78.3%	85.8%	29.7%
Mean	11.96	15.04	15.65	17.16	5.90
Stand. dev	2.18	1.85	2.03	1.60	1.44

Since the results of the first test were encouraging, at least for three of the five tested classifiers, a second test was done. The aim of this analysis was to examine if persons could be identified reliably even if the breath samples used for testing were collected on a different day than the samples used for training the classifiers. The recorded currents that were used for identification depend on the mobility of ionized molecules, which is affected by factors such as temperature and humidity [14]. Because it could not be ensured that the environmental conditions were the same for datasets measured on different days, an approach to correct for background noise was used. For each dataset, the average of IMS values for samples 10 to 40 was computed and subtracted from each sample of the exhales. Measurements from the first 9 seconds were ignored to ensure that the IMS samples used for calculating the baseline had stabilized.

For the test, two sets per participant were used for training and the remaining set was used for testing the classifiers. This means that the training data consisted of 80 breath samples (20 per participant) and 40 breath samples were then classified. This test was repeated three times. Each time different datasets were used for training and testing. For example, in the first test datasets 1 and 2 of all participants were used for training the classifiers and datasets 3 were used for testing them. Table 2 shows the percentage of correctly classified samples in each repetition as well as average accuracies over all three repetitions. The results showed that the number of correctly classified breath samples depended significantly on the training sets. Furthermore, none of the tested classifiers provided a satisfying identification accuracy. Therefore, more sophisticated approaches were tested, namely Dynamic Time Warping (DTW) and Recurrent Neural Networks (RNNs).

The RNNs used Sobel gradient magnitude or gradient direction matrices as input. However, neither DTW nor the RNNs yielded higher accuracies than the algorithms presented in Table 2. This suggests that the limiting factor was the data. Thus, further research is needed to find methods that ensure that breath samples collected under different environmental conditions are comparable.

TABLE II. SECOND ANALYSIS

Train Data Sets	Test Data Sets	*Performance of the classifiers when data from all three sessions are used*				
		CDT	***KNN***	***NB***	***LDA***	***QDA***
1,2	3	40%	57.5%	42.5%	40%	42.5%
1,3	2	57.5%	55%	45.0%	50%	25%
2,3	1	20.0%	32.5%	50.0%	30%	30%
Average accuracy		39.2%	48.3%	48.5%	41.7%	32.5%

IV. DISCUSSION

The results show that out of five tested classifiers, three (LDA, KNN, and NB) can be used to classify breath samples between four participants even though none of the classifiers reached 100% classification accuracy. These results are in line with the previous studies considering human breath samples and other odors [e.g., 4, 12], which indicate that 100% accuracies are rarely achieved. CDT and QDA were less successful. One reason for the poor performance of CDT is that it splits the data first on the first measured current, then on the second measured current, and so on. Thus, this approach heavily depends on the order of currents in the breath prints, making it unsuitable for the studied task. With respect to functionality of the QDA, the low identification rate may indicate that the covariances of all prediction classes are more or less the same.

Unlike previous studies [3, 4], chemical analysis or any special sampling procedures to collect the data was not used. Therefore, the data collected was more sensitive to external distractions. LDA, KNN and NB functioned best if training and test data were collected in identical environmental conditions (within the same session). However, if the environmental conditions vary significantly (within different sessions) then the classification accuracy decreases considerably. Overall, no classification accuracy higher than 57.5% was achieved. This means that most of the time the classifiers performed better than chance alone (25%), but not well enough to reliably be used to identify all the participants based on the breath sample. More sophisticated approaches (DTW and RNNs) did not improve the accuracies, suggesting that the focus should be on normalizing samples such that the impact of environmental conditions on the breath samples could be mitigated during analysis. However, it should be noted that to use breath samples in real-life applications the classifier only needs to be able to determine if the person giving the sample is, for example, the owner of the computer or suggest a visit to a dentist based on breath sample. The response of the algorithm can be simple yes / no instead of identifying samples as A, B, C, D, etc.

We presented a preliminary study investigating the potential to use scent as an input. The amount of data, somewhat standard environment (i.e., laboratory) and the use of only five classifiers with standard baseline correction are clear limitations of the current study. To continue developing scent-based input, more data and testing of different analytical methods are needed so that it will be possible to identify conditions where the proposed idea can work and where it cannot (e.g., if breath sample can increase the identification accuracy together with facial recognition). Further, it can be feasible to collect the data so that the environmental factors are manipulated on purpose. The additional information measured by the eNose (e.g., humidity) and preprocessing techniques to remove background noise from the breath samples, and/or improving the algorithm, would likely obtain better results (see, for example, [2] for time-related variations in chemical compounds in breath).

ACKNOWLEDGMENT

This work was supported by Academy of Finland projects DIGITS (295433 and 295432) and ProScents (323498 and 323529).

REFERENCES

[1] de Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., ... & Ratcliffe, N. M. (2014). A review of the volatiles from the healthy human body. Journal of breath research, 8(1), 014001.

[2] Sinues, P. M. L., Kohler, M., & Zenobi, R. (2013). Human breath analysis may support the existence of individual metabolic phenotypes. PloS one, 8(4), e59909.

[3] Wang, X. R., Lizier, J. T., Berna, A. Z., Bravo, F. G., & Trowell, S. C. (2015). Human breath-print identification by E-nose, using information-theoretic feature selection prior to classification. Sensors and Actuators B: Chemical, 217, 165-174.

[4] Westhoff, M., Rickermann, M., Franieck, E., Littterst, P., & Baumbach, J. I. (2019). Time series of indoor analytes and influence of exogeneous factors on interpretation of breath analysis using ion mobility spectrometry (MCC/IMS). International Journal for Ion Mobility Spectrometry, 22(1), 39-49.

[5] Wongchoosuk, C., Lutz, M., & Kerdcharoen, T. (2009). Detection and classification of human body odor using an electronic nose. Sensors, 9(9), 7234-7249.

[6] Bikov, A., Pako, J., Kovacs, D., Tamasi, L., Lazar, Z., Rigo, J., ... & Horvath, I. (2011). Exhaled breath volatile alterations in pregnancy assessed with electronic nose. Biomarkers, 16(6), 476-484.

[7] Bajtarevic, A., Ager, C., Pienz, M., Klieber, M., Schwarz, K., Ligor, M., ... & Hilbe, W. (2009). Noninvasive detection of lung cancer by analysis of exhaled breath. BMC cancer, 9(1), 348.

[8] Righettoni, M., Tricoli, A., & Pratsinis, S. E. (2010). Si: WO3 sensors for highly selective detection of acetone for easy diagnosis of diabetes by breath analysis. Analytical chemistry, 82(9), 3581-3587.

[9] Ruzsányi, V., Kalapos, M. P., Schmidl, C., Karall, D., Scholl-Bürgi, S., & Baumann, M. (2018). Breath profiles of children on ketogenic therapy. Journal of breath research, 12(3), 036021.

[10] Sinues, P. M. L., Kohler, M., & Zenobi, R. (2013). Human breath analysis may support the existence of individual metabolic phenotypes. PloS one, 8(4), e59909.

[11] Utriainen, M., Kärpänoja, E., & Paakkanen, H. (2003). Combining miniaturized ion mobility spectrometer and metal oxide gas sensor for the fast detection of toxic chemical vapors. Sensors and Actuators B: Chemical, 93(1-3), 17-24.

[12] Müller, P., Salminen, K., Nieminen, V., Kontunen, A., Karjalainen, M., Isokoski, P., ... & Lekkala, J. (2019). Scent classification by K nearest neighbors using ion-mobility spectrometry measurements. Expert Systems with Applications, 115, 593-606.

[13] R.F. Maestre. 2017. Calibration of the mobility scale in ion-mobility spectrometry: the use of 2,4-lutidine as a chemical standard, the twostandard calibration method and the incorrect use of drift tube temperature for calibration. Analytical Methods, 9, 4288-4292

[14] Müller, P., Salminen, K., Kontunen, A., Karjalainen, M., Isokoski, P., Rantala, J., ... & Surakka, V. (2019). Online Scent Classification by Ion-mobility Spectrometry Sequences. Frontiers in Applied Mathematics and Statistics, 5, 39.