Online Scent Classification by Ion-Mobility Spectrometry Sequences

Philipp Müller[1]*, Katri Salminen[2], Anton Kontunen[3], Markus Karjalainen[3], Poika Isokoski[1], Jussi Rantala[1], Joni Leivo[3], Jari Väliaho[3], Pasi Kallio[3], Jukka Lekkala[3] and Veikko Surakka[1]

[1] Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland, [2] Tampere University of Applied Sciences, Tampere, Finland, [3] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

For ion-mobility spectrometry (IMS)-based electronic noses (eNose) samples of scents are markedly time-dependent, with a transient phase and a highly volatile stable phase in certain conditions. At the same time, the samples depend on various environmental factors, such as temperature and humidity. This makes fast classification of scents challenging. The present aim was to develop and test an algorithm for online scent classification that mitigates these dependencies by using both baseline measurements and sequences of samples for classification. A classifier based on the K nearest neighbors approach was derived. The classifier is able to use measurements from both transient and stable phase, yields a label for the analyzed scent, and information on the trustworthiness of the returned label. In order to avoid the classifier being fooled by irrelevant features and to reduce the dimensionality of the feature space, principal component analysis was applied to the data. The classifier was tested with four food scents, each presented in two different ways to the IMS. By using baseline measurements, the misclassification rate was reduced from 20.0 to 13.3%. A second experiment showed that the used IMS type experiences device heterogeneity.

Keywords: ion-mobility spectrometry, electronic nose, K nearest neighbors, scent classification, time-series analysis, device heterogeneity

1. INTRODUCTION

Classifying scents is of interest in various applications. Examples include but are not limited to quality control in the food industry (e.g., monitoring beer quality in a brewery), detection of gas leaks and drugs, and analyzing curative and aromatic plants used in medicines, perfumes, and cosmetics (e.g., [1] and references therein). The ultimate goal of our research is to measure an original scent source using an electronic nose (eNose) at location A, and classify it based on the eNose data. The measurements and the obtained scent label will then be sent to location B, where a scent synthesizer generates a synthetic scent that is as similar as possible to the original scent at location A (see [2] for details). For this research on scent transfer over time and space, a method for quick and accurate scent measurement and classification needs to be developed.

To enable scent classification, a suitable method to acquire data needs to be selected. Various approaches for scent classification based on eNose measurements have been proposed and

tested recently. In general, the classification is based on a single sample, which consists of measurements for various quantities observed at the same time. The samples from scents can be taken with various types of eNoses [1, 3, 4]. Ideally, the chosen eNose will enable real-time monitoring, quick analysis, show fast response time, show little to no temporal performance degradation, be relatively affordable, commercially available, and portable. One technology that fulfills all these requirements is ion-mobility spectrometry (IMS) [3, 5, 6]. It is a time of flight analytical technique [7], which enables temporal or physical separation of ionized molecules [8]. IMS sensors are small and light, and therefore are used in commercially available and relatively affordable eNoses that were developed for field applications (i.e., outside laboratories). Furthermore, these sensors have low operating costs and power consumption and high sensitivity [9]. Finally, most IMS-sensors do not age significantly, as the rely on radioactive sources, with half-lives of several hundred years, for ionizing the molecules. This means, that changes in the IMS measurements over, for example, a month or a year are solely due to changes in the scent measured by the eNose, rather than due to changes in the IMS sensor.

In Müller et al. [10] 14 different food scents and three key odor components of jasmine were measured with an IMS-based eNose and successfully classified using a K nearest neighbor (KNN) approach. However, the classification accuracy strongly depended on the quality of the IMS measurements. For the experiments, the scents were presented to the eNose on a plate (i.e., there was surrounding room air) and in a flask (i.e., the source of scent was sealed in a flask so that the external air could not affect the measurement data). The time series analysis of IMS measurements showed a clear transient phase in the beginning, i.e., a time interval of $\approx$ 20–30 s in which the measured currents are gradually increasing or decreasing. Furthermore, strong temporal fluctuations were noticed when the scent source was presented on a plate. Thus, a suitable scent classifier has to account for these fluctuations. An often used approach to avoid issues related to the transient phases and strong fluctuations of IMS signals is to wait for the IMS readings to stabilize and to only measure scents from closed flasks, for which only small temporal fluctuations were recorded in the stable phase (e.g. in Mamat et al. [11], Giordani et al. [12], and Tang et al. [13]). However, in our project and for most practical applications (e.g., food quality control or detection of harmful substances in room air) neither of these approaches is an optimal solution. The scent needs to be classified quickly and it is impractical or even impossible to conserve the scent source in a sealed flask.

A second challenge that has to be addressed by the classifier, especially if the scent source is not presented in a flask, is the influence of environmental factors. These factors (e.g., temperature, barometric pressure, and air currents) affect the movement of molecules [7, 14] and hence the IMS measurements. Obviously, when measuring scents in the field the samples collected with the IMS device are noisier than in laboratory conditions. In order to classify scents correctly in conditions that differ from the conditions in which measurements for the training database were collected, there is a need for baseline correction (i.e. the measurements are corrected

for background noise). This is done, for example, for databases on hazardous gases and drugs. However, how baseline corrected measurements are exactly calculated is, in general, kept secret by the manufacturers of the eNoses.

Device heterogeneity is a third challenge that, in general, has to be considered. It describes the phenomenon that two devices, even of the same brand and model, yield different results in identical sensing conditions. Thus, the classifier has to correct for these differences in the signal.

In this paper the aim was to develop and test a simple classifier that addresses the first two challenges presented above. To mitigate the influence of signal fluctuations the proposed KNN-based classifier yields labels for analyzed scents online using time series of IMS measurements rather than single measurements. The classifier then uses distances to the K closest neighbors for computing how trustworthy the yielded label is. It constantly updates the label of the analyzed scent as well as the trustworthiness level as new measurements become available. This technique helps to reduce the influence of measurements from the transient phase on the final label, as these measurements are generally considered by the algorithm to be less trustworthy. On the other hand, it enables faster classification by using the smaller amount of information contained in measurements from the transient phase compared with approaches that have to wait for the signal to stabilize before starting the classification process. In order to achieve real-time classification, the IMS data is first transformed using principal component analysis (PCA), which enables to reduce the dimensionality of the data, and k-dimensional tree search is employed to find the K closest neighbors. Besides dimension reduction, PCA also removes any correlation from the data, which could degrade the performance of the classifier. The reason for correlation in IMS data is explained in section 2.1. Background noise corrections were done measuring the ambient room air before measuring the scent to get a baseline, which then was subtracted from the scent measurements. This means that relative measurements instead of absolute measurements were used to mitigate the influence of environmental factors.

The third challenge was not addressed by the classifier, because all measurements were taken with the same IMS sensor. However, further tests were done with that sensor and another IMS sensor of the same type to check whether these sensors experience device heterogeneity.

Therefore, the contribution of this paper is 3-fold. First, it presents and tests an online classifier that labels an analyzed scent and provides a measure for the trustworthiness of the returned label. Both label and trustworthiness label are constantly updated once new measurements are available. Second, the experiments with real data show that by using baseline measurements (i.e., correcting for the background signal) the accuracy of the classifier can be improved somewhat, but that additional information will be needed to distinguish between scents that share a large portion of odor-active compounds. Third, tests with two IMS devices of the same type show that device heterogeneity is an issue for IMS-based eNoses that needs to be considered when comparing measurements from two or more devices.

This paper is organized as follows. Section 2 briefly explains how IMS-based eNoses work and discusses the influence of time and environmental conditions on the IMS measurements. It then explains the studied machine learning problem, *K*NN-type classifiers, and derives the online scent classifier. Furthermore, this section describes how to use baseline measurements for mitigating the effects of environmental conditions on IMS measurements. The classification performance of the online classifier is tested in section 3 using data from scents presented to the eNose in different ways. In addition, data from two identical eNoses is analyzed to study if device heterogeneity affects IMS-based eNoses. Finally, the test results are discussed, conclusions are drawn, and an outlook on future research is given in section 4.

Notation: In this paper *a* denotes a scalar, **b** denotes a vector, and **C** denotes a matrix.

2. MATERIALS AND METHODS

2.1. Scent Analysis by Ion-mobility Spectrometry

IMS enables detection of scents at atmospheric pressure [15] even if their concentration is at parts per billion (ppb) level [16]. It separates the ions of a scent based on their velocity, under the influence of a constant electric field [8]. Due to differences in their molecular weight, charge and geometry between compounds the mobility of various ions differs [8], resulting, in general, in different "fingerprints" (i.e., measurements over all IMS channels) for different scents.

These IMS measurements were done with Environics' ChemPro100i handheld chemical detector [17]. The ChemPro100i uses an Americium-241 source for ionization, which is similar to the radioactive sources used in ionization type smoke detectors. The source of the analyzed scent is placed in front of the device. The device then sucks in the scented air and measures ions in the air with an aspirating system. A rotary vane pump generates a flow that carries the ions, and a perpendicular electric field picks them from the flow. The ChemPro100i constantly measures the ions as currents with 7 separate electrode pairs, and the electric field is constantly switched between positive and negative polarities. This means that instead of measuring the flight times of different ions, the ChemPro100i measures 14 groups of flight times. This means that ions with same charge and flight times within in a certain interval are measured by the same electrode. The current measured by any of the 14 electrodes indicates the presence of certain type of ions (more details can be found in Environics [18]). Each electrode yields one numerical value, which is the response of this channel (named so, because each electrode measures ions with a band of flight times). Together these values form a 14-dimensional IMS sample (so-called fingerprint).

It is important to note that the measurements from adjacent electrodes are (potentially) correlated. Imagine that electrodes A and B measure positively charged ions with flight times in $[a, b]$ and $[b, c]$, respectively. The flight time of the ions could be assumed to be somewhat noisy. Now a positively charged ion with flight time $\approx b$ might be measured by either of the electrodes
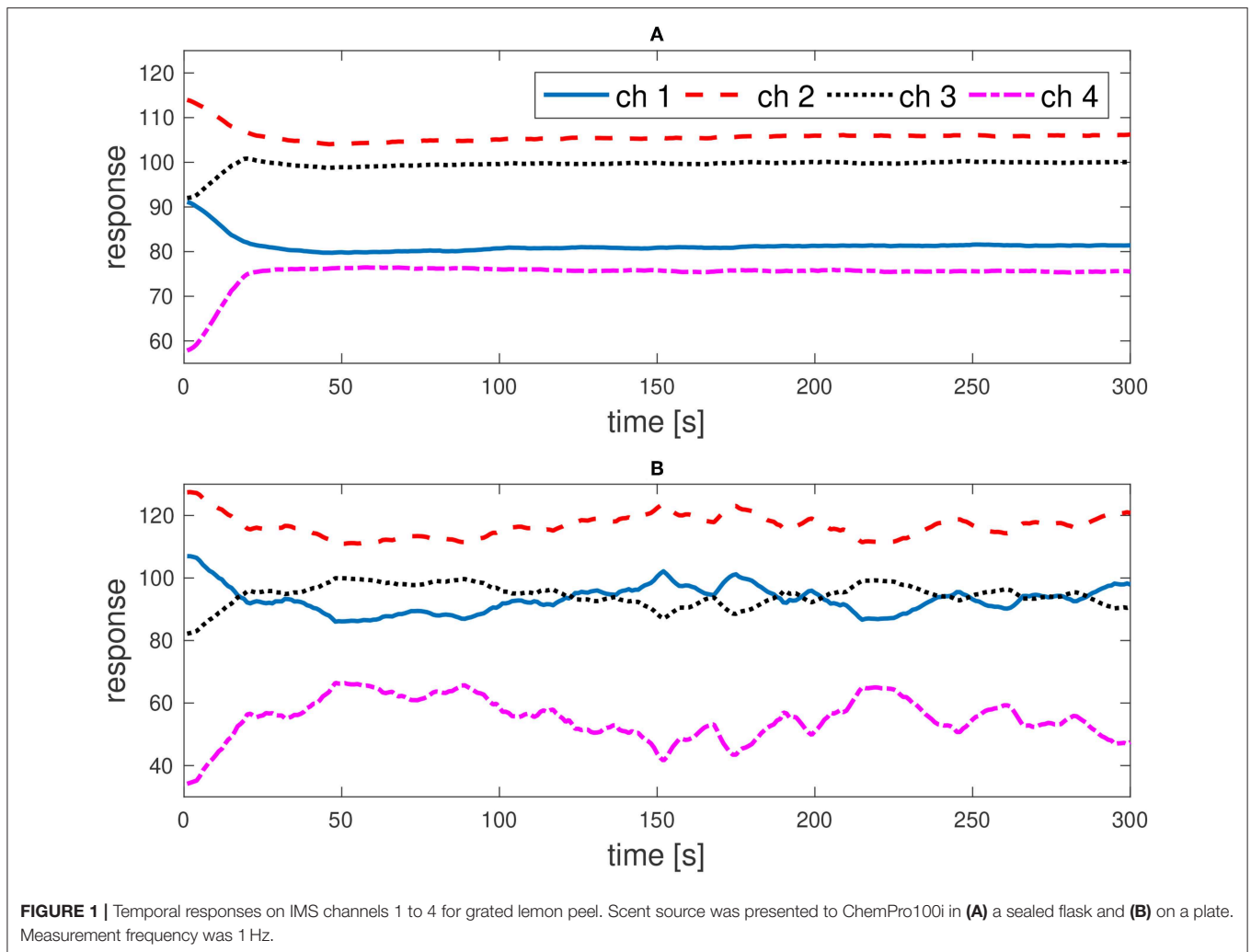
based on the noise level in its flight time. From this example, it becomes clear that measurements from electrodes are most likely correlated.

Figure 1A shows an example of temporal responses on the IMS channels 1 to 4 when presenting grated peel of ripe lemon in a flask to the eNose. Other channels show similar behavior. The channel responses take several seconds to stabilize. For each channel the time to stabilize, or in other words the length of the transient phase, can vary. In **Figure 1A** channels 1 to 4 each increase or decrease respectively for $\approx$ 20–30 s before stabilizing. All responses from these first 20–30 s are considered to be measurements from the transient phase. In the stable phase, i.e., after stabilizing, the channel responses in **Figure 1A** show only minor temporal fluctuations, which was observed for all scents presented to ChemPro100i in a sealed flask. If the scent source is presented to the device on a plate then the channel responses are, in general, different from the flask responses and show generally strong temporal fluctuations. **Figure 1B** shows responses on the IMS channels 1 to 4 of the grated lemon peel when presented on a plate. In Müller et al. [10] it was discovered that classifying a scent based on samples from the transient phase and/or presented on a plate instead of a sealed jar generally increases the risk of misclassification significantly. However, for real-world applications it is desirable to use already measurements from the transient phase for classification and to simply sniff the scent source from a plate rather than first preparing the scent source in a sealed flask. Thus, this paper describes an algorithm that enables that.

The air flow outside the device is not necessarily stable and affects how many molecules of the scent are sucked into the ChemPro100i. However, also the flow inside the measurement device varies because the flow is created with a rotary vane pump, which tends to create pulsation to the flow. Therefore, the IMS measurements fluctuate. **Figure 2** shows an example for the pulsating flow, measured for lemon peel presented to ChemPro100i on a plate. The flow fluctuates around 1.3 l/min with a small amplitude (standard deviation of 0.0018 l/min).

In order to mitigate the influence of the strong temporal fluctuations a sliding moving average was used in Müller et al. [10] for smoothing the channel responses. Another idea, which has been shown to be beneficial, for example, for magnetic location fingerprint-based localization [19], is to use temporal sequences (i.e., time series) of fingerprints instead of single fingerprints. In this paper both approaches are applied.

Besides the transient phase and the method for presenting the scent source, there are other factors that influence the channel readings. Especially humidity (i.e., moisture of the surrounding air) and temperature, but also barometric pressure, and air currents influence the mobility of molecules [20, pp. 250 ff.] and hence will influence the IMS readings. **Figure 3** shows the responses on IMS channels 1 and 3 for grated lemon peel on a plate. Data for set 1 was collected roughly 1 month before data set 2, at the same location. Although the conditions were quite similar [absolute humidity: 19.55 $\frac{g}{m^3}$ (set 1) vs. 19.90 $\frac{g}{m^3}$ (set 2); temperature: 27.82°C vs. 28.43°C; air pressure: 1002.68 mBar vs. 990.48 mBar] the responses on both channels

FIGURE 1 | Temporal responses on IMS channels 1 to 4 for grated lemon peel. Scent source was presented to ChemPro100i in **(A)** a sealed flask and **(B)** on a plate. Measurement frequency was 1 Hz.

differ significantly. It is important to note that the IMS sensor ages very slowly as the half-life of the radioactive source in ChemPro100i, Americium-241, is 432.2 years. Thus, aging is an unlikely reason for the observed differences. This needs to be considered when classifying scents using a training database collected under different conditions. In this paper baseline measurements are used to make fingerprints collected in different conditions comparable. Details can be found in subsection 2.5.

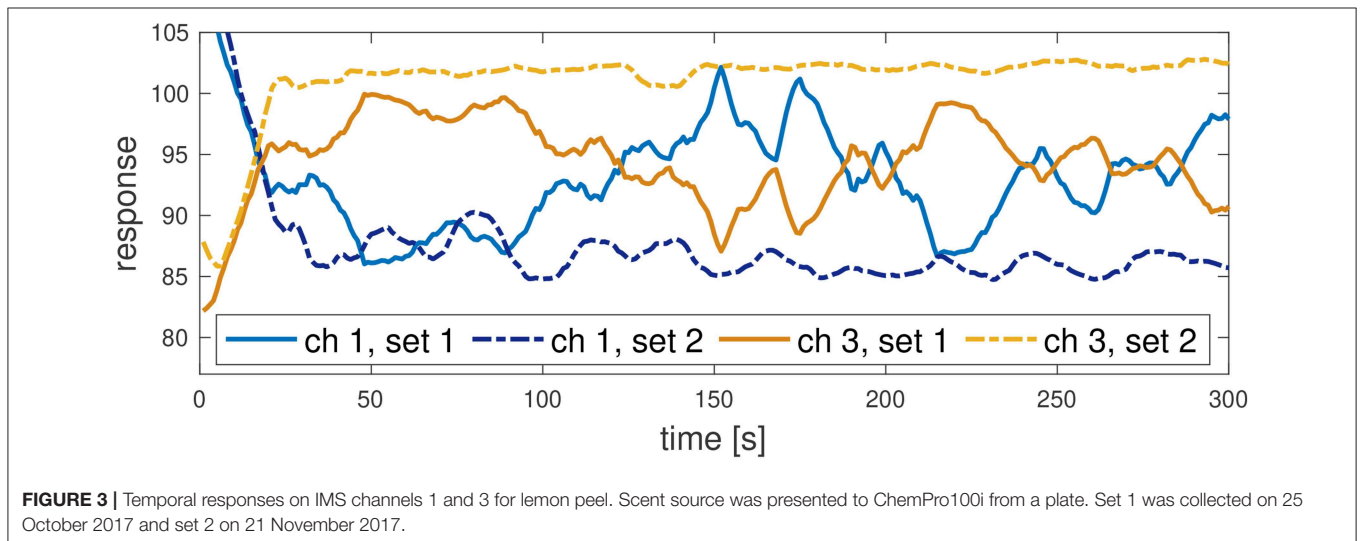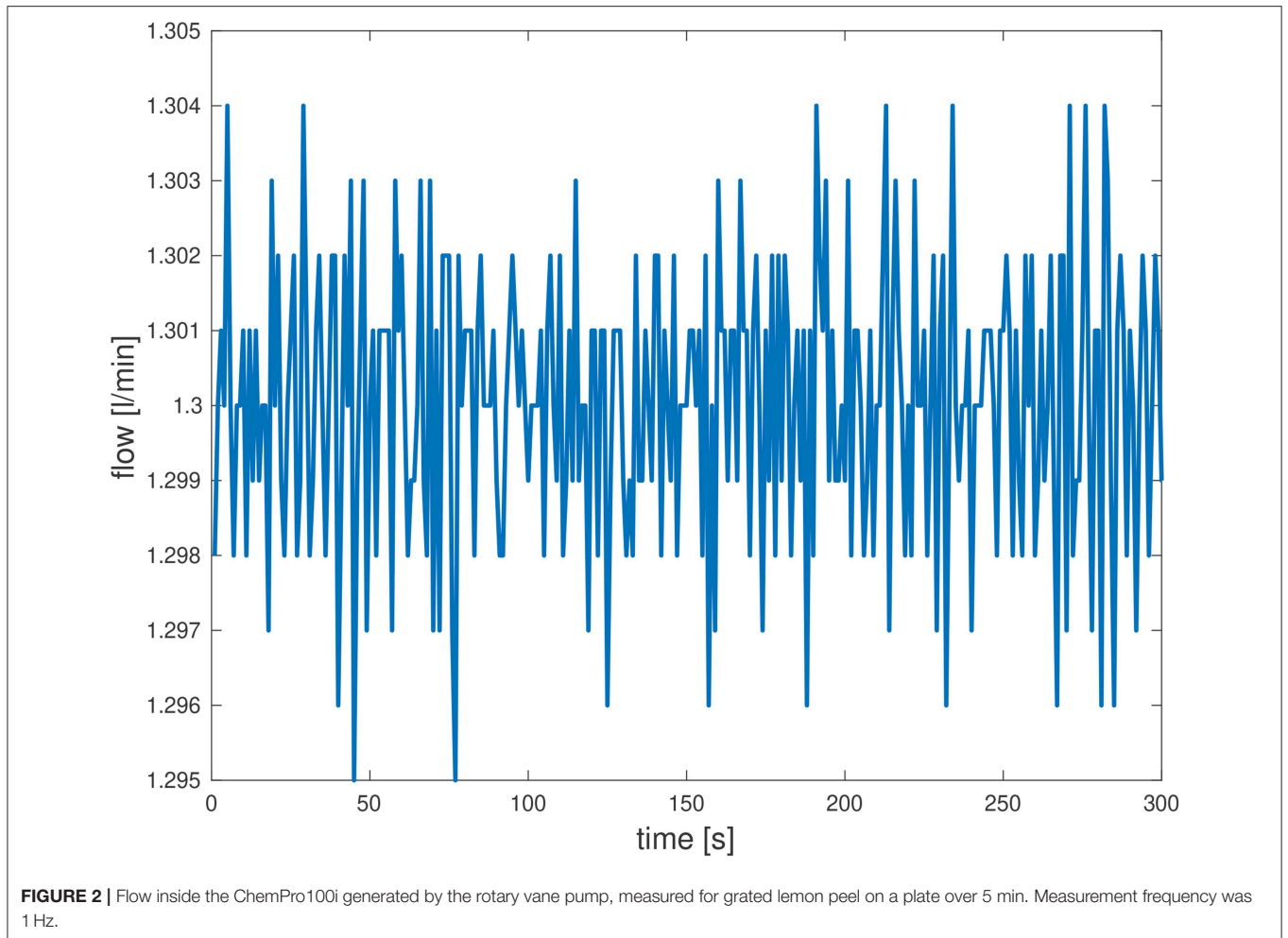## 2.2. Specification of the Machine Learning Problem

According to Jung [21, p. 7], the formal definition of any machine learning (ML) problem consists of (i) data characterized by features and labels, (ii) a hypothesis space containing feasible maps from feature to label space, and (iii) a loss function for measuring the quality of the predictor or classifier. For more details the reader is referred to Jung [21], on which this subsection is built on. In this paper, the data consists of 14-dimensional IMS samples $\mathbf{x} = [x_1 .. x_{14}]$ containing the currents measured by any of the 14 electrodes in the ChemPro100i. The 14 different channels could be used directly as the features of the ML

problem. However, in this paper the data will be transformed first by principal component analysis (PCA) and the first n_p principal components explaining at least 99% of the total variance in the training data are used as features. The feature space is then $\mathbb{R}^{n_p}$, meaning that each feature sample consists of n_p real numbers. The labels of the studied ML problem are the scent names related to each IMS sample and each feature sample.

In this paper, an approach based on the K nearest neighbors algorithm is used for mapping a PCA-transformed IMS sample to a predicted scent name. Because the label space is finite in the analyzed case, the hypothesis map is called a classifier [21, p. 18]. The quality of the classifier is evaluated by the percentage of correctly classified scents.

2.3. *K* Nearest Neighbors-Type Scent Classifiers

In Müller et al. [10] and Salminen et al. [22] a K nearest neighbors algorithm for classifying scents achieved high accuracy. The KNN classifier compares the 14-dimensional IMS sample $\mathbf{x}^{(us)} = [x_1^{(us)} .. x_{14}^{(us)}]$ of the unlabeled scent with (labeled) IMS training

FIGURE 2 | Flow inside the ChemPro100i generated by the rotary vane pump, measured for grated lemon peel on a plate over 5 min. Measurement frequency was 1 Hz.

FIGURE 3 | Temporal responses on IMS channels 1 and 3 for lemon peel. Scent source was presented to ChemPro100i from a plate. Set 1 was collected on 25 October 2017 and set 2 on 21 November 2017.

samples stored in a training database. The training database consists of N IMS samples $\mathbf{x}_i = [x_{i,1} .. x_{i,14}]$, $i = 1, .., N$, i.e., $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$. For the ith sample the corresponding scent label s_i is recorded and stored in $\mathbf{s} = \{s_i\}_{i=1}^N$. The closeness between the unlabeled sample $\mathbf{x}^{(us)}$ and the ith training sample is computed as the Euclidean distance between the two, which is defined as

$$d_{\mathrm{E}}(\mathbf{x}^{(\mathrm{us})}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{14} \left(x_{ij} - x_j^{(\mathrm{us})}\right)^2}. \tag{1}$$

The unlabeled scent is then classified as belonging to the same scent as the majority of its K closest neighbors (i.e., the samples for which $d_{\mathrm{E}}(\mathbf{x}^{(\mathrm{us})}, \mathbf{x}_i)$ is minimal). Parameter K should be odd to avoid ties in the majority vote [23, p. 183]. If there is nevertheless a tie then in this paper the label of the closest training sample is chosen as label for the scent.

Alternatives to KNN include Weighted KNN (WKNN) and fuzzy KNN. The WKNN uses normalized inverses of distances for weighting how much it trusts the labels of each of the K neighbors. The smaller the distance between $\mathbf{x}^{(\mathrm{us})}$ and a training sample the more the WKNN trusts the label of this training sample. Therefore, ties are almost impossible for WKNN classifiers. Fuzzy KNN classifiers adopt the principle of WKNN. Instead of focusing only on the K closest neighbors, they consider all training samples and the distances between them and $\mathbf{x}^{(\mathrm{us})}$, and yield probabilities for the unlabeled sample belonging to any of the scents stored in the training database.

A drawback of KNN-type classifiers is that they require a large number of training samples from each scent source that they should be able to classify, which will result in a large N and high computational demand [24]. Basic KNN-type classifiers will compare the unlabeled scent sample $\mathbf{x}^{(\mathrm{us})}$ with all N training samples in order to find the K closest training samples, which is called exhaustive search. In order to reduce the computational demand, here a pre-structuring technique called k-dimensional trees [25] is used. Note that k has no connection to K, and that generally $k \neq K$ will hold. The idea of k-dimensional (k-d) tree search is to split the training data into subsets, then use the binary k-d tree to address $\mathbf{x}^{(\mathrm{us})}$ to a certain subset, and choose the K nearest neighbors only from the training samples in this subset. The major drawback of k-d tree search is that it might miss the true nearest neighbors, because it is an approximate method. However, for large N it generally works well [23, p. 183], and in Müller et al. [10] the same classification accuracies were obtained with both exhaustive and k-d tree search.

KNN-type classifiers, like most other classifiers, can be fooled by irrelevant features. Therefore, within this paper principal component analysis is applied to the IMS data before the classification. PCA generates a new set of so-called principal components that represent the full data space (see e.g., [23, p. 183]). The advantage of PCA-transformed data, compared with the IMS data, is that its principal components (i.e., the transformed channels) are uncorrelated, which allows to remove irrelevant features. In addition, it allows to reduce the dimensionality of the data used for classification, which reduces the computation demand. Based on the results of Müller et al. [10], here the first n_p components explaining at least 99% of the total variance in the training data are used.

2.4. Online Scent Classification
Classifying a scent based on a single sample is possible and has been shown to provide reasonable classification accuracy [10]. However, in Müller et al. [10] it was also noticed that especially

samples that are taken within the first 20–30 s after the scent is presented to the IMS-device (i.e., samples from the transient phase) cause more often a misclassification than samples that were taken during the stable phase. In order to enable reliable yet quick scent classification, this paper introduces a classifier that decides the label of the scent based on a temporal sequence of samples. Furthermore, it adopts the idea of WKNN and fuzzy KNN to provide a measure for the trustworthiness of the returned label.

The idea of the classifier is to first collect n consecutive IMS samples from the unlabeled scent, average over the n samples to obtain the averaged sample $\bar{\mathbf{x}}^{(\mathrm{us})}$, and then transform the averaged sample into its PCA-transformed equivalent

$$\mathbf{y}^{(\mathrm{us})} = (\bar{\mathbf{x}}^{(\mathrm{us})} - \boldsymbol{\mu})\mathbf{C}, \tag{2}$$

where $\boldsymbol{\mu} = [\mu_1 .. \mu_{14}]$ is a vector containing the empirical means of the 14 IMS channels and $\mathbf{C}$ is a 14-by-14 matrix containing the principal component coefficients. Finally, it classifies the scent by comparing $\mathbf{y}^{(\mathrm{us})}$ with the training samples. If there is still uncertainty about the label then another n consecutive IMS samples will be collected and processed in the same way. This can be repeated indefinitely often, but usually after 5 to 10 iterations, depending of course on the choice of n and the length of the transient phase, the classifier should function reliably because at least some of the measurements will be from the stable phase and hence more trustworthy.

In order to avoid ties the principle of the WKNN is applied when determining the final label for the scent. Each of the m repetitions yields one scent label, to which a weight is addressed. For the tests in this paper the inverse of the average distance to the K nearest neighbors of each iteration is chosen as (unnormalized) weights. This way the classifier trusts those labels most for which the average distance between $\mathbf{y}^{(\mathrm{us})}$ and the K nearest neighbors is the smallest.

The pseudo-code for classification using this approach is given in Algorithm 1. Note that an overall classification label is returned after each iteration, together with the measure p_{label} that specifies how certain the classifier is that "it picked" the correct label for the analyzed scent. $\mathbf{1}_{\{s_k^l = \mathbf{s}^l\}}$ is one only if $s_k^l = \mathbf{s}^l$ holds, and otherwise zero. The normalized weights have to be recomputed in each iteration.

2.5. Baseline Measurements for Simple IMS Calibration
In order to mitigate the effect of varying environmental conditions in this paper baseline measurements are used for calibrating the samples collected from scent sources. Baseline measurements mean here that z consecutive IMS samples are collected from the air of the room in which samples for a scent source are collected.

The baseline measurements are used as follow. For each of the 14 IMS channels the averaged baseline sample is computed by

$$\bar{\mathbf{x}}^{(\mathrm{air})} = \frac{1}{z} \sum_{i=1}^{z} \mathbf{x}_i^{(\mathrm{air})}. \tag{3}$$

Algorithm 1: Pseudo-code of scent classifier

Input: PCA-transformed training data $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N}$ and corresponding scent labels $\mathbf{s} = \{s_i\}_{i=1}^{N}$, number of nearest neighbors K, number n of consecutive IMS samples taken from unlabeled scent, number n_p of principal components explaining at least $p_{\min}\%$ of total variance in the training data, minimum number $m_{\min}$ of iterations, minimum probability p_{sure} for final scent label

Set number of iterations $m \leftarrow 0$ and probability of label $p_{\text{label}} \leftarrow 0$

while $m < m_{\min}$ or $p_{\text{label}} < p_{\text{sure}}$ **do**

    Set $m \leftarrow m + 1$

    Collect n consecutive IMS sample from unlabeled scent

    Average over the n samples to obtain $\bar{\mathbf{x}}^{(\text{us})}$

    Transform $\bar{\mathbf{x}}^{(\text{us})}$ into $\mathbf{y}^{(\text{us})}$ using (2)

    Classify unlabeled scent by KNN using $\mathbf{y}^{(\text{us})}$ and compute average distance $\bar{d}_{\text{E}}^m$ to K nearest neighbors

    Use scent labels $[s_1^l, .., s_m^l]$ to compute final scent label $\mathbf{s}^l$ and compute p_{label} using normalized weights

    $[w_1, .., w_m] = {}^{[1/\bar{d}_{\text{E}}^1, .., 1/\bar{d}_{\text{E}}^m]} / \sum_{k=1}^{m} 1/\bar{d}_{\text{E}}^l$

    Return $\mathbf{s}^l$ and $p_{\text{label}} = \sum_{k=1}^{m} w_k \mathbf{1}_{\{s_k^l = \mathbf{s}^l\}}$

end while

Then, the averaged air sample $\bar{\mathbf{x}}^{(\text{air})}$ is subtracted from all samples of the scent to obtain the baseline-adjusted samples, i.e., $\mathbf{X}_{\text{bl}} = \{\mathbf{x}_i^{(\text{bl})}\}_i$ with

$$\mathbf{x}_i^{(\text{bl})} = \mathbf{x}_i - \bar{\mathbf{x}}^{(\text{air})}. \tag{4}$$

This means that the signals are corrected for the background "noise" caused by changing environmental conditions to make them comparable. The baseline-adjustment is done before the PCA-transformation.

3. RESULTS

3.1. Scent Classification With and Without Baseline Correction

In this subsection the classification performance of the classifier proposed in section 2 is checked. Furthermore, the effect of using baseline measurements on the algorithm's classification accuracy is analyzed.

For the test, training data for grape, jasmine oil (1% concentration), lemon peel (grated peel of ripe lemon), and vanilla (sliced dried vanilla fruit) was collected between 25 October and 1 November 2017 in our laboratory at the University of Tampere, Finland. Each of the four scent sources was presented to the ChemPro100i in a sealed flask and on a plate, meaning that the label space consisted of 8 elements [Grape (flask), Grape (plate), Jasmine (flask), ...]. Grape, lemon peel, and vanilla were chosen because they were used already in [10]. Jasmine was added because in Müller et al. [10] its three key odor components were analyzed and classified.

For each source and each presentation method five measurement sets were collected as follow. First, 5 min baseline from the air in the laboratory was measured. Subsequently samples of the scent source, presented in a flask or on a plate, were measured over 5 min. This was followed by a break of 10 min to ensure that the IMS readings returned to the levels of measuring room air only (instead of traces of the scent source that might have still lingered in the air or the sample gas inlet [16]). After the break, again 5 min of laboratory air for the baseline followed by 5 min of scent source were collected, before another break of 10 min. This cycle was repeated five times with a sampling rate of 1 Hz. Thus, for each scent source 1,500 training samples (five times 300 samples), which formed the training database $\mathbf{X}_{\text{nbl}} = \{\mathbf{x}_i^{(\text{nbl})}\}_i$ of non-baseline adjusted samples, and 1,500 baseline samples were collected. The same amount of training samples for each scent source was used to avoid skewed class distributions. In order to mitigate the effect of measurement noise, all training and test data sets were smoothed with a sliding moving average (MA) with window length 11, meaning that instead of raw IMS sample the average of the raw sample, its five predecessors, and its five successors is used. The smoothed data was then standardized by centering it and dividing it by the standard deviations of all training samples for any IMS channel (see [10] for details). Finally, PCA was applied to the IMS data to obtain $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\mathbf{C}$.

The test data, that is the samples from unlabeled scents, was collected in a similar way. Again, data for grapes, jasmine oil, lemon peel, and vanilla was collected continuously. However, this time each scent source was presented to the ChemPro100i in three different ways: in a sealed flask (22 November) and on a plate (21 November) in our laboratory, and on a plate in a meeting room (23 November 2017) at the University of Tampere, Finland. Due to the gap of 3 to 4 weeks between measuring training and test data the environmental conditions and hence the IMS readings changed significantly (compare **Figure 3**), which was necessary to check the influence of baseline measurements.

For each source and each presentation method two test sets à 5 min were collected, and 5 min for baseline before each measurement set. The baseline measurements were used as explained in subsection 2.5. Different numbers of baseline samples were tested: $z = \{0, 30, 300\}$, where $z = 0$ meant that no baseline was used to adjust IMS samples from the scent being analyzed.

The computations were done in Matlab R2016a. **Table 1** summarizes the results. In the test $n = 10$, which means that $\bar{\mathbf{x}}^{(\text{us})}$ was obtained by averaging over 10 consecutive IMS samples and that a maximum of 30 averaged samples could be used. In order to simulate online scent classification, blocks of 10 consecutive IMS samples were read-in and used for classification. If $m \geq m_{\min}$ and $p_{\text{label}} \geq p_{\text{sure}}$ then the classification was stopped. Otherwise, the next block of 10 consecutive IMS samples was read-in. The first block contained samples 1 to 10, the second 11 to 20, etc. For PCA $p_{\min} = 99\%$ was used. This resulted in using only the first four (for $z = \{0, 30\}$) or first five (for $z = 300$) principal components for classification (feature space was thus

TABLE 1 | Classification results for tests in 3.1 with ($z = 300$ and $z = 30$) and without ($z = 0$) using baseline measurements.

Scent	Location	Set	Label (z = 300)	p_{label} in %	m m	Label (z = 30)	p_{label} in %	m m	Label (z = 0)	p_{label} in %	m m
Grape	Flask	1	*Vanilla (f)*	81.3	8	Grape (f)	80.6	12	*Jasmine (f)*	80.3	10
		2	*Vanilla (p)*	100	5	*Vanilla (p)*	100	5	*Jasmine (f)*	80.9	19
	Plate	1	Grape (p)	86.9	5	*Vanilla (p)*	100	5	*Vanilla (p)*	82.1	6
	(lab)	2	*Vanilla (p)*	86.3	5	[Vanilla (p)]	61.5	30	*Vanilla (p)*	81.0	5
	Plate	1	Grape (p)	100%	5	Grape (p)	86.7	5	Grape (p)	80.3	17
	(meet.)	2	Grape (p)	100	5	Grape (p)	86.0	7	Grape (p)	80.1	7
Jasmine	Flask	1	Jasmine (f)	100	5	Jasmine (f)	100	5	Jasmine (f)	87.8	5
		2	Jasmine (f)	100	5	Jasmine (f)	100	5	Jasmine (f)	100	5
	Plate	1	Jasmine (p)	100	5	Jasmine (p)	100	5	Jasmine (p)	81.5	10
	(lab)	2	Jasmine (p)	100	5	Jasmine (p)	100	5	Jasmine (p)	100	5
	Plate	1	Jasmine (p)	100	5	[Jasmine (p)]	43.0	30	Jasmine (p)	81.6	5
	(meet.)	2	Jasmine (p)	100	5	[Jasmine (p)]	46.0	30	Jasmine (p)	80.5	5
Lemon	flask	1	Lemon (f)	100	5	Lemon (f)	80.0	10	Lemon (f)	100	5
		2	Lemon (f)	100	5	Lemon (f)	100	5	Lemon (f)	100	5
	Plate	1	Lemon (p)	84.0	5	Lemon (p)	85.2	5	Lemon (p)	80.3	7
	(lab)	2	Lemon (p)	100	5	Lemon (p)	100	5	Lemon (p)	86.1	5
	Plate	1	Lemon (p)	100	5	[Grape (p)]	37.7	30	Lemon (p)	82.5	7
	(meet.)	2	Lemon (p)	100	5	[Lemon (p)]	71.2	30	Lemon (p)	80.8	15
Vanilla	Flask	1	Vanilla (f)	80.7	13	Vanilla (f)	80.3	22	Vanilla (f)	80.8	5
		2	Vanilla (f)	100	5	Vanilla (f)	81.9	5	Vanilla (f)	81.7	10
	Plate	1	[Vanilla (f)]	51.1	30	[Lemon (f)]	49.0	30	*Jasmine (p)*	100	5
	(lab)	2	[Grape (p)]	45.7	30	*Vanilla (f)*	80.9	8	*Jasmine (p)*	80.1	12
	Plate	1	Vanilla (p)	80.2	29	*Grape (f)*	100	5	Vanilla (p)	100	5
	(meet.)	2	*Grape (p)*	80.8	21	Vanilla (p)	100	5	Vanilla (p)	80.2	23

For the test $n = 10$, which means that a maximum of 30 averaged samples could be used, $p_{min} = 99\%$, $m_{min} = 5$, and $p_{sure} = 80\%$. For the predicted labels f stands for flask and p stands for plate. Misclassified scents are underlined and cursive. An estimate in [ ] means that the classifier is not certain about it, i.e., $p_{label} < p_{sure}$, and that all 30 samples have been evaluated.

either $\mathbb{R}^4$ or $\mathbb{R}^5$, respectively). **Figure 4** shows the percentages of the total variance explained by each principal component as well as their cumulative sum for $z = 30$ (for $z = \{0, 300\}$ no clear differences are visible). Components 5/6 to 14 do not contain any new information, and can therefore be discarded. In addition, at least $m_{\min} = 5$ averaged samples were classified for each unlabeled scent and $p_{\text{sure}} = 80\%$. If $p_{\text{label}} < 80\%$ after evaluating all 30 samples then the classifier did not decide on the scent's final label.

The classifier had problems to correctly classify grapes for all z-values. However, without using baseline measurements for calibration (i.e. $z = 0$) four out of six test sets were misclassified. When using baseline measurements only two ($z = 30$) or three ($z = 300$) sets got misclassified; all of them as vanilla. One reason why the classifier struggled to correctly classify grapes might be that it was impossible to ensure that grapes for both training and test data were of the same cultivation and/or ripeness level. However, explaining why grapes got always misclassified as vanilla is more difficult. The most likely explanation is that both scents contain ionized molecules with similar weights. Hence, these molecules would end up in the same channels, resulting in

similar fingerprints. One has to keep in mind that IMS measures volatile organic compounds (VOCs). Thus, if there is one or more VOCs in large quantities having similar molecular properties, and hence similar time of flight, in both scents, then IMS readings will be similar to each other [5]. For example, grapes contain β-Damascenone in significant quantities, which is the most powerful scent that is not fermentation related [26]. At the same time β-Damascenone is one of the four most powerful odor-active compounds (the other three being Vanillin, Guaiacol, and Ethyl-E-cinnamate) in crushed vanilla [27]. Therefore, additional information/ measurements will be needed for distinguishing between grape and vanilla scent, which is left for further research. However, the test showed that using the baseline to correct for background "noise" at least helps to avoid unreasonable classification as Jasmine.

For vanilla the results were ambiguous. While the two sets measured from flask were correctly classified independent on z, the sets measured from plate were often misclassified or no final decision was made after evaluating all 30 samples. However, without using baseline calibration the results were worse, as in the case of grapes. With $z = 0$ the two sets measured from the

FIGURE 4 | Importance of principal components. Portion of total variance in the training data explained by each of the 14 principal components as well as cumulative sum (red line) for $z = 30$. Dashed line shows used $p_{min} = 99\%$.
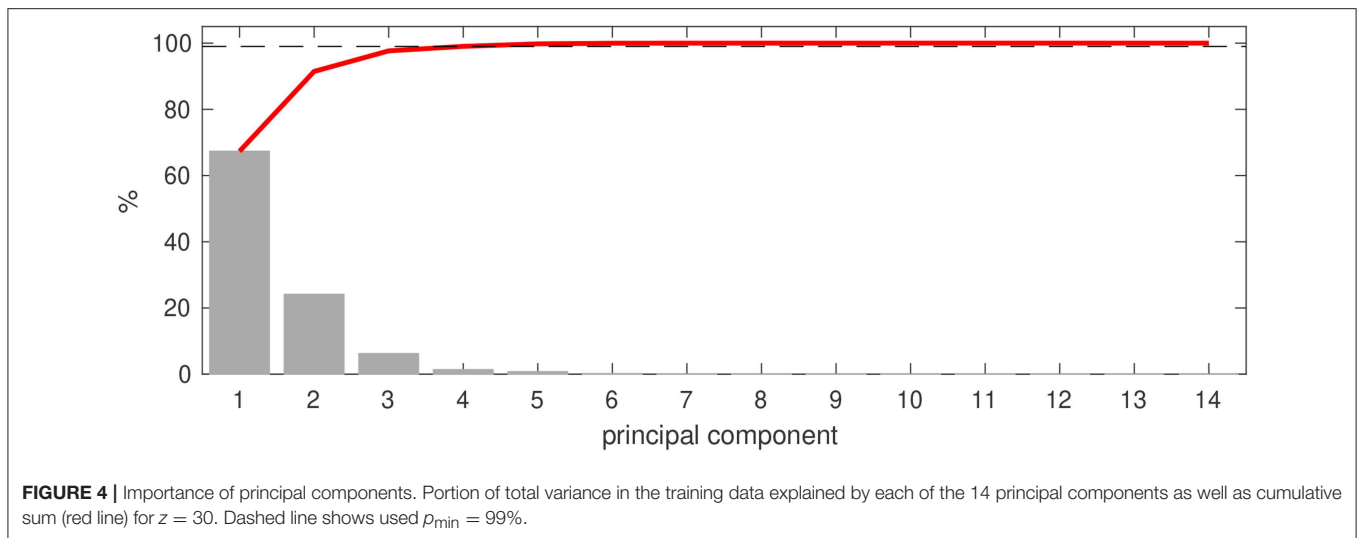
plate in the lab were misclassified as jasmine. Also for $z = 30$ two sets were misclassified and for one no decision was reached, but one of the misclassified sets was still classified as vanilla and the other as grape, which again can be explained by β-Damascenone being contained in both vanilla and grape. For $z = 300$ one set was misclassified as grape, which is in line with the explanation why grape got misclassified as vanilla. However, for the first set the most probable label after evaluating all 30 samples was vanilla and for the second set it was grape, which supports aforementioned explanation.

Jasmine oil and lemon peel were always correctly classified. Only for measurements from plate in the meeting room the classifier with $z = 30$ failed to make a final decision based on the 30 samples. However, for all three of the four sets it suggested the correct label but failed to achieve $p_{label} \geq 80\%$.

Thus, it can be concluded that using baseline measurements somewhat helps in reducing the effect of background noise on scent classification but further research is needed to find ways to reduce the noise even further. In addition, additional features, besides IMS measurements, should be tested for distinguishing between scents sharing dominant odor-active compounds.

The improvement in classification accuracy using baseline measurements comes at the cost of larger computational demand. The average number of samples analyzed by the classifier was 9.2 ($z = 300$), 12.7 ($z = 30$), or 8.5 ($z = 0$), respectively. However, considering only the cases in which the classifier labeled the scent correctly the differences in average used samples were smaller when using baseline measurements, namely 6.8 ($z = 300$), 7.2 ($z = 30$), and 8.1 ($z = 0$). This means, that the computational demand was on a similar level and the classification took slightly longer when using the baseline because for the first z seconds room air was measured.

3.2. Device Heterogeneity

Device heterogeneity (sometimes also called signal shift) describes the phenomenon that two devices, even of the same brand and model, yiel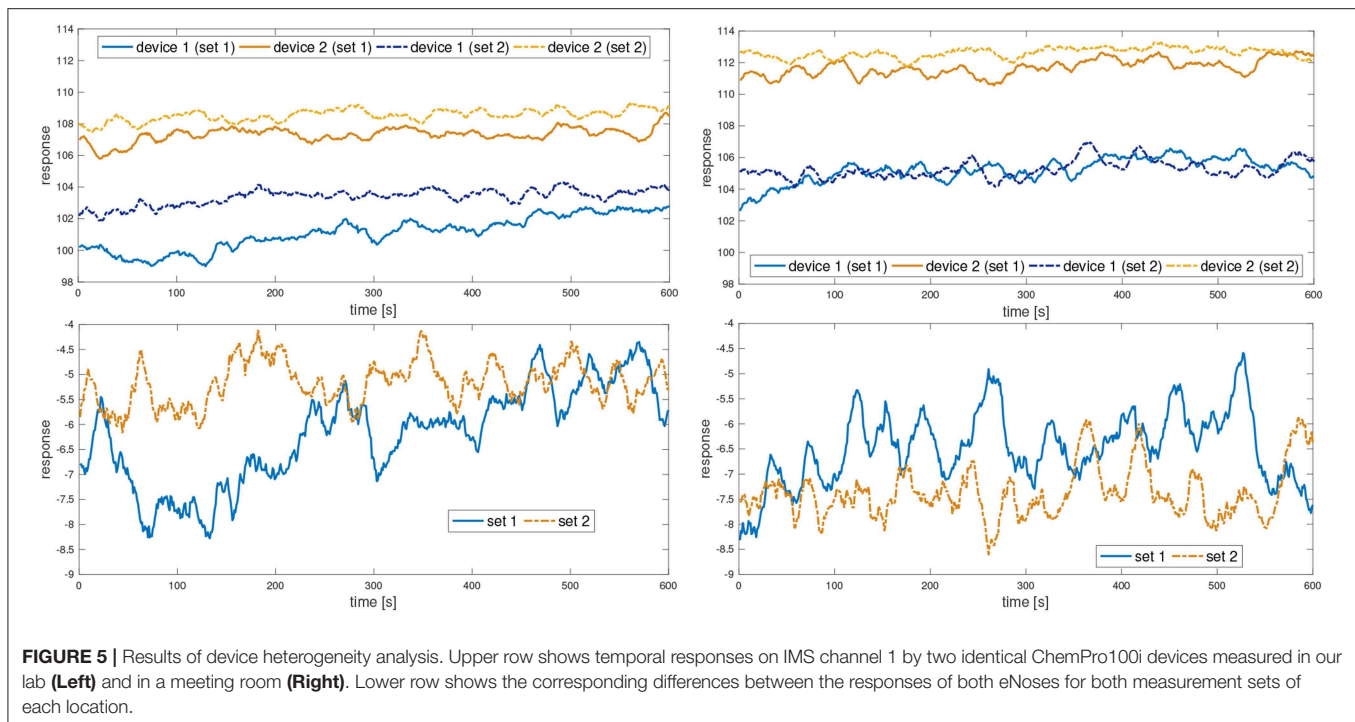d different results in identical sensing conditions. It can be observed, for example, for metal-oxide semiconductor (MOS) gas sensors (see e.g., [28]).

In order to check whether device heterogeneity is an issue that needs to be addressed for scent classification, IMS samples with two ChemPro100i devices were collected simultaneously at our lab and the meeting room from subsection 3.1. At each location two sets à 10 min were collected, with a break of 2 min between the sets.

The upper row in **Figure 5** shows the responses on channel 1 in both locations as an example. Device 1 is the IMS that was used for collecting training and test data for section 3.1, and device 2 is an identical ChemPro100i. However, the readings on channel 1 of both devices differ significantly for both sets. Comparing the raw IMS data over the 10 min similar differences were observable for all channels and in both locations. This means that device heterogeneity is an issue that one needs to be aware of when collecting samples, either for training or testing, with different devices. The differences in the readings of sets 1 and 2 show furthermore the time-dependence of channel responses and the short-term fluctuations mentioned earlier.

In the lower row of **Figure 5** the differences between the IMS readings on channel 1 are depicted. For the remaining 13 channels the plots are similar. The difference in IMS readings of both devices is not constant, which means that simply determining a value for the offset of device 2 compared with device 1 and then subtracting it all the time from the second device's channel response to obtain data comparable with the first device's channel response will not be the best possible solution. One reason is the pulsating air flow in the IMS. Because each pump is unique and the flow sensor's accuracy is limited, it is impossible to ensure that both devices have exactly the same flow at any time. This explains, beside the difference, also the fluctuations seen in the lower row of **Figure 5**.

It is important to note the difference between using baseline measurements and device calibration. Baseline measurements are used to make data collected with the same device but at different time (and location) comparable. Device calibration is used to

FIGURE 5 | Results of device heterogeneity analysis. Upper row shows temporal responses on IMS channel 1 by two identical ChemPro100i devices measured in our lab **(Left)** and in a meeting room **(Right)**. Lower row shows the corresponding differences between the responses of both eNoses for both measurement sets of each location.

make data collected with different devices at the same time and location comparable. However, the necessity for calibration needs to be studied further. One hypothesis that has to be tested is, that the difference between two devices when measuring baseline can be also observed when both devices measure a scent. In this case using baseline measurements could already remove or at least mitigate the influence of device heterogeneity.

4. DISCUSSION

In this paper a KNN-based scent classifier that labels scents based on PCA-transformed temporal sequences of ion-mobility spectrometry samples measured by an electronic nose was presented and tested. The results showed that the classifier enables online classification using also samples from the transient phase, rather than waiting first for the signals to stabilize, and provides both a label for the scent and its confidence level. This enables the users to decide how much uncertainty they are willing to accept.

Furthermore, by measuring the air for a few seconds before presenting the scent source to the IMS and using it as baseline, the misclassification rate was reduced from 20.0 to 13.3% in the test. The remaining misclassification could be explained by the fact that for both grape and vanilla β-Damascenone is one of the their most powerful odor-active compounds. The reduction in the misclassification rate was independent on whether 30 or 300 s of samples were used for the baseline measurement. Thus, the time it takes to measure the baseline could be chosen shorter (e.g., 30 s) than the time it took for signal stabilization in Mamat et al. [11], Giordani et al. [12], and Tang et al. [13]. Furthermore, using baseline samples reduced the number of samples required by the

classifier to provide a trustworthy label for a scent. This saved, on average, ≈ 10 s compared with classification without baseline (for correctly classified scents). Thus, classification using a baseline of 30 s effectively takes only ≈ 20 s longer than classification without baseline. As such, the results show great promise for developing scent measurement and classification technologies needed to transfer scents in real time. Next, we will test the scent transfer paradigm by using a system architecture to communicate the scent measurement data.

In the test a training database of four scent sources presented in two ways was used. Therefore, the classifier had to distinguish between 8 different scents. For future research, more scents will be added to the training database and the classifiers performance will be rechecked. In order to be able to differentiate also between scents containing ionized molecules with similar weights additional measurements will be added to the analysis, thus adding more features to the training database. To handle the larger training databases alternative approaches for classifying single averaged samples will be studied and compared with the KNN-based classifier presented in this paper. However, the concept of using sequences of IMS samples and providing uncertainty information will be kept.

Although, using baseline measurements to mitigate background signals improved the classifier's accuracy, there are potential solutions for further improvements. For example, one could try to develop a mathematical model that allows predicting the change in IMS responses due to changes in measurement conditions (e.g., temperature, humidity, etc.).

This paper also showed that IMS-based eNoses are affected by device heterogeneity. However, its influence on the classification accuracy still needs further research. If the influence is significant

then individual eNoses have to be calibrated to enable the use of different devices for collecting reliable training and test samples. Calibration could be done (1) by using one eNose as master device and transforming the readings of all other devices to resemble the master device, (2) by updating a calibration model developed for the master device using measurements taken with another IMS-based eNose, or (3) by transforming the target values used in the calibration [29]. For non-IMS eNoses the first approach is the most commonly used (see [29] and references therein). In addition, existing manual (e.g., [30, 31]) and automatic calibration methods (e.g., [32, 33], which have been successfully applied for calibrating devices measuring WiFi signal strengths, could be studied.

All in all, the present study showed significant potential for developing online scent classification using IMS-based eNose technology. Although device heterogeneity might be an issue to be resolved in the future, the main interest of this study was scent classification, which proved to be successful. At the same time the tests showed that additional information, besides IMS samples, should be used for scent classification. In this respect it is likely that the classification is functional in other contexts facing changes in the volatile compounds in air such as warning systems for gas leaks. For the classification of a wider selection of scents the use of information from additional sensors will be studied.

DATA AVAILABILITY

The datasets analyzed for this study can be found in the Etsin Research data finder at https://etsin.avointiede.fi/dataset/urn-nbn-fi-csc-kata20190506161504626222 under "Supplements." By clicking on the link right to Supplements' you will be able to download a zip-file that contains the datasets measured with the eNose as well as all scripts needed for redoing the tests discussed in this manuscript.

AUTHOR CONTRIBUTIONS

PM contributed to conceptualization, data curation, software, formal analysis, investigation, methodology, software, and visualization. He wrote the original draft. KS contributed to conceptualization, data collection, methodology, and funding acquisition. AK contributed to conceptualization, formal analysis, and methodology. MK contributed to methodology. PI contributed to conceptualization and formal analysis. JR contributed to investigation, methodology, resources, and software. JLei contributed to formal analysis. JV contributed to conceptualization, and methodology. PK contributed to conceptualization, supervision, funding acquisition, and project administration. JLek and VS contributed to conceptualization, supervision, funding acquisition, and project administration. All authors reviewed the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Wilson AD, Baietto M. Applications and advances in electronic-nose technologies. *Sensors.* (2009) **9**:5099–148. doi: 10.3390/s90705099
2. Nieminen V, Karjalainen M, Salminen K, Rantala J, Kontunten A, Isokoski P, et al. A compact olfactometer for IMS measurements and testing human perception. *Int J Ion Mobil Spectrom.* (2018) **21**:71–80. doi: 10.1007/s12127-018-0235-1
3. Loutfi A, Coradeschi S, Mani GK, Shankar P, Rayappan JBB. Electronic noses for food quality: a review. *J Food Eng.* (2015) **144**:103–11. doi: 10.1016/j.jfoodeng.2014.07.019
4. Kiani S, Minaei S, Ghasemi-Varnamkhasti M. Application of electronic nose systems for assessing quality of medicinal and aromatic plant products: a review. *J Appl Res Med Aromatic Plants.* (2016) **3**:1–9. doi: 10.1016/j.jarmap.2015.12.002
5. Li J, Gutierrez-Osuna R, Hodges RD, Luckey G, Crowell J, Schiffman SS, et al. Using field asymmetric ion mobility spectrometry for odor assessment of automobile interior components. *IEEE Sensors J.* (2016) **16**:5747–56. doi: 10.1109/JSEN.2016.2568209
6. Baether W, Zimmermann S, Gunzer F. Pulsed ion mobility spectrometer for the detection of Toluene 2,4-Diisocyanate in ambient air. *IEEE Sensors J.* (2012) **12**:1748–54. doi: 10.1109/JSEN.2011.2179117

7. Maestre RF. Calibration of the mobility scale in ion mobility spectrometry: the use of 2,4-lutidine as a chemical standard, the two-standard calibration method and the incorrect use of drift tube temperature for calibration. *Anal Methods.* (2017) **9**:4288–92. doi: 10.1039/C7AY01126A
8. Zamora D, Blanco M. Improving the efficiency of ion mobility spectrometry analyses by using multivariate calibration. *Anal Chim Acta.* (2012) **726**:50–6. doi: 10.1016/j.aca.2012.03.023
9. Mäkinen MA, Anttalainen OA, Sillanpää MET. Ion mobility spectrometry and its applications in detection of chemical warfare agents. *Anal Chem.* (2010) **82**:9594–600. doi: 10.1021/ac100931n
10. Müller P, Salminen K, Nieminen V, Kontunten A, Karjalainen M, Isokoski P, et al. Scent classification by K nearest neighbors using ion-mobility spectrometry measurements. *Expert Syst Appl.* (2019) **115**:593–606. doi: 10.1016/j.eswa.2018.08.042
11. Mamat M, Samad SA, Hannan MA. An electronic nose for reliable measurement and correct classification of beverages. *Sensors.* (2011) **11**:6435–53. doi: 10.3390/s110606435
12. Giordani DS, Castro HF, Oliveira PC, Siqueira AF. Biodiesel characterization using electronic nose and artificial neural network. In: *Proceedings of European Congress of Chemical Engineering (ECCE-6).* Copenhagen (2007).
13. Tang KT, Chiu SW, Pan CH, Hsieh HY, Liang YS, Liu SC. Development of a portable electronic nose system for the detection and classification of fruity odors. *Sensors.* (2010) **10**:9179–93. doi: 10.3390/s101009179

14. Majid A, Speed L, Croijmans I, Arshamian A. What makes a better smeller? *Perception.* (2017) **46**:406–30. doi: 10.1177/0301006616688224

15. Guharay SK, Dwivedi P, Hill HH Jr. Ion mobility spectrometry: ion source development and applications in physical and biological sciences. *IEEE Trans Plasma Sci.* (2008) **36**:1458–70. doi: 10.1109/TPS.2008.927290

16. Barth S, Baether W, Zimmermann S. System design and optimization of a miniaturized ion mobility spectrometer using finite-element analysis. *IEEE Sensors J.* (2009) **9**:377–82. doi: 10.1109/JSEN.2009.2014411

17. ChemPro100i. (2018). Available online at: http://www.environics.fi/product/chempro100i/ (accessed October 16, 2018).

18. Environics. (2018). Available online at: https://www.environics.fi/product-categories/chemical-detection-technologies/ (accessed October 16, 2018).

19. Gao C, Harle R. Sequence-based magnetic loop closures for automated signal surveying. In: *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN).* Calgary, AB (2015).

20. Eiceman GA, Karpas Z, Hill HH Jr. *Ion Mobility Spectrometry.* 3rd ed. CRC Press (2014).

21. Jung A. *Machine Learning: Basic Principles.* (Espoo) (2019). Available online at: https://arxiv.org/pdf/1805.05052.pdf (accessed June 17, 2019).

22. Salminen K, Rantala J, Isokoski P, Lehtonen M, Müller P, Karjalainen M, et al. Olfactory display prototype for presenting and sensing authentic and synthetic odors. In: *20th ACM International Conference on Multimodal Interaction (ICMI'18).* Boulder, CO (2018).

23. Duda RO, Hart PE, Stork DG. *Pattern Classification.* 2nd ed. New York, NY: Wiley-Interscience (2001).

24. Moreno-Seco F, Micó L, Oncina J. A modification of the LAESA algorithm for approximated k-NN classification. *Pattern Recogn Lett.* (2003) **24**:47–53. doi: 10.1016/S0167-8655(02)00187-3

25. Bentley JL. Multidimensional binary search trees used for associative searching. *Commun ACM.* (1975) **18**:509–17. doi: 10.1145/361002.361007

26. Sun Q, Gates MJ, Lavin EH, Acree TE, Sacks GL. Comparison of odor-active compounds in grapes and wines from Vitis vinifera and non-Foxy American grape species. *J Agric Food Chem.* (2011) **59**:10657–64. doi: 10.1021/jf2026204

27. Takahashi M, Inai Y, Miyazawa N, Kurobayashi Y, Fujita A. Key odorants in cured Madagascar vanilla beans (vanilla planiforia) of differing bean quality. *Biosci Biotechnol Biochem.* (2013) **77**:606–11. doi: 10.1271/bbb.120842

28. Yan K, Zhang D. Improving the transfer ability of prediction models for electronicnoses. *Sensors Actuat B Chem.* (2015) **220**:115–24. doi: 10.1016/j.snb.2015.05.060

29. Marco S, Gutiérrez-Gálvez A. Signal and data processing for machine olfaction and chemical sensing: a review. *IEEE Sensors J.* (2012) **12**:3189–214. doi: 10.1109/JSEN.2012.2192920

30. Vaupel T, Seitz J, Kiefer F, Haimerl S, Thielecke J. Wi-Fi positioning: system considerations and device calibration. In: *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN).* Zurich (2010).

31. Haeberlen A, Flannery E, Ladd AM, Rudys A, Wallach DS, Kavraki LE. Practical robust localization over large-scale 802.11 wireless networks. In: *MobileCom'04.* Philadelphia, PA (2004).

32. Laoudias C, Piché R, Panayiotou CG. Device signal strength self-calibration using histograms. In: *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN).* Sydney, NSW (2012).

33. Koski L, Perälä T, Piché R. Indoor positioning using WLAN coverage area estimates. In: *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN).* Zurich (2010).