Privacy-Preserving Automatic Collection of Acoustic Voiding Events

Laura Arjona[1], Yasha Iravantchi[2], Alanson Sample[2], Marcos L. Alvarez[1], Alfonso Bahillo[3], Elba Canalon[4]

***Abstract*— Uroflowmetry is a non-invasive diagnostic test used to evaluate the function of the urinary tract. Despite its benefits, it has two main limitations: high intra-subject variability of flow parameters and the requirement for patients to urinate on demand. To overcome these limitations, we have developed a low-cost ultrasonic platform that utilizes machine learning (ML) models to automatically detect and record natural in-home voiding events, without any need for user intervention. This platform operates outside of human-audible frequencies, providing privacy-preserving, automatic uroflowmetries that can be conducted at home as part of daily routines. After evaluating several machine learning algorithms, we found that the Multi-layer Perceptron classifier performed exceptionally well, with a classification accuracy of 97.8% and a low false negative rate of 1.2%. Furthermore, even on lightweight SVM models, performance remains robust. Our results also showed that the voiding flow envelope, helpful for diagnosing underlying pathologies, remains intact even when using only inaudible frequencies.**

***Clinical relevance*— This classification task has the potential to be part of an essential toolkit for urology telemedicine. It is especially useful in areas that lack proper medical infrastructure but still host ubiquitous embedded privacy-preserving audio capture devices with Edge AI capabilities.**

I. INTRODUCTION

New strategies are needed to transition from in-person and reactive healthcare systems to remote and proactive ones that provide continuous and non-intrusive care to patients. The greatest focus is on developing systems that allow the patient to self-monitor, thereby transitioning medical-grade sensing from the hospital to the home. By collecting, monitoring, and analyzing several bio-signals with various sensors, some diseases, like those related to the urinary tract, can be prevented or curbed, potentially avoiding severe damage which may have otherwise gone undetected. A problem frequently associated with aging is voiding dysfunction. This highly prevalent issue significantly impacts the quality of life for a large number of individuals (more than 60% of men over 60 years of age) [1]. *Uroflowmetry* is a non-invasive diagnostic test widely used to assess how well the urinary tract functions by tracking how fast urine flows, how much flows out, and how long it takes. Current uroflowmetry tests are unsuitable for continuous health monitoring in a nonclinical environment, as they are often distressing, costly, and burdensome for the public. It is carried out on an outpatient basis, at specified procedure areas, and involves having the person urinate into a uroflowmeter. This process is unnatural and requires "on-demand" voiding, often with either low or very high bladder filling. This leads to significant test-to-test variability as the situational stress of the patient can affect the flow rate.

Remote assessment of people with voiding dysfunction may allow for the capture of unique voiding features, thus facilitating the prompt diagnosis of a disease. Recent works have demonstrated the feasibility of using mobile devices, such as smartphones [2] and smartwatches [3], in a home environment, to characterize urinary flow patterns by capturing the sound generated when the urine stream hits the water in a toilet bowl. This test is known as audio uroflowmetry. The main limitation of these devices is that they need active user interaction (patients need to interact with an App), and their battery needs to be recharged often, especially in the case of commercial smartwatches. Furthermore, the presence of a listening microphone in the toilet may raise significant privacy concerns to users.

To overcome these limitations, we seek to explore sounds outside of human hearing and their utility for automatic sound-based uroflowmetries at home. We have developed a proof-of-concept acoustic Raspberry Pi (RPi) based platform that runs a novel machine learning (ML) algorithm to automatically detect and record voiding events at home that can be easily incorporated into a normal daily routine requiring no human intervention. The human audible frequencies are removed from the audio signals to preserve user privacy. Detecting and recording voiding events automatically at home preserving the users' privacy is a must for the implementation of audio uroflowmetries.

II. RELATED WORK

A. Acoustic Uroflowmetry Events at Home

Audio uroflowmetry represents an alternative approach to characterize urinary flow patterns by capturing the acoustic energy generated when the urine stream hits the water in the toilet. Most existing works use smartphones as a platform for audio uroflowmetry [2]. Recently, smartwatches have been used to perform audio uroflowmetries at home with promising results [3]. However, to detect and record voiding events automatically (without user intervention) both the smartphone and the smartwatch have challenges:

- They require active user interaction, which often results in limited data collection at night time.
- They may reduce patient compliance over time due to the inconvenience of having the device while urinating.

[1]Laura Arjona and Marcos L. Alvarez are with the University of Deusto, Bilbao, Spain laura.arjona@deusto.es

[2]Yasha Iravantchi and Alanson Sample are with the University of Michigan, Ann Arbor, MI yiravan@umich.edu, apsample@umich.edu

[3]Alfonso Bahillo is with the University of Valladolid, Valladolid, Spain, alfonso.bahillo@uva.es

[4]Elba Canelon is with the Hospital Universitario de Puerto Real, Cádiz, Spain elbacanelon@gmail.com

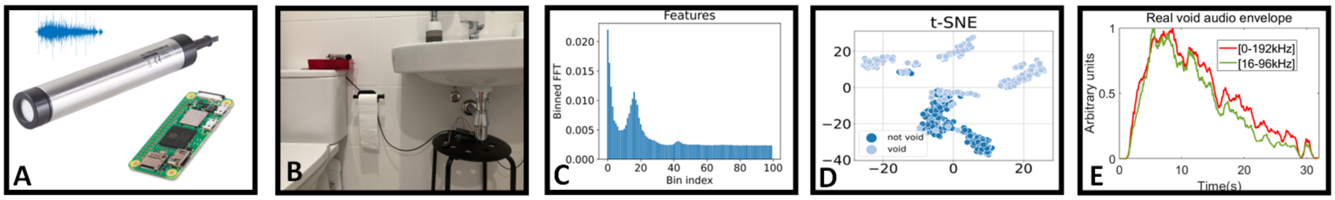

Fig. 1. (A) Audio collection hardware (B) deployed in a home toilet to build our database. (C) For each audio clip we extract 100 linear spaced FFT bins to generate the ML model features. (D) t-SNE plot shows how the void (light blue) and no-void (dark blue) classes can be well separated. (E) Finally we show that the removal of the audible frequencies (green line) do not affect the envelope extraction needed for sound uroflowmetries.

- They may evoke privacy concerns due to the presence of a microphone in the toilet and other sensitive in-home areas.
- The smartphone has additional requirements, such as fixed placement to obtain consistent results across recordings.

In [4], authors propose an algorithm to automatically recognize the movements involved in urination by analyzing movement data collected from smart bands. Nevertheless, to automatically recognize a urination event, they need the user to perform a forward movement with the wrist to prepare for urination and then a backward movement for cleaning up. Because they detect the voiding event after the event has ended, the event cannot be recorded. Authors in [5] presented what they call a smart toilet that calculates the flow rate and volume of urine using computer vision. However, it the video violates the users' privacy.

B. Machine Learning for Sound Classification

Sound classification algorithms typically attempt to predict the content of an audio segment. Conventionally, many approaches employ Fast Fourier Transforms (FFTs) and Mel Frequency Cepstral Coefficients (MFCCs) as features and use classical supervised learning algorithms for classification. MFCCs have been widely used with both classical and deep learning approaches with high accuracy [6]. A growing number of works use deep learning approaches which have shown exceptional performance in sound labeling tasks. In fact, deep learning models are becoming the *de facto* standard in mobile and embedded applications [7]. Sound has also been used for scene and object recognition, as presented in SoundNet. [8] have also exploited the temporal nature of sound and have used convolutional recurrent neural networks to create sound models. Additionally, within the last 20 years, compiled and crowdsourced-labeled datasets, such as AudioSet, have become increasingly available. Consequently, the use of ML algorithms for sound classification seems to be a good choice.

C. Microphones and Privacy

As the number of microphone-containing devices increases, there is a growing number of works analysing privacy issues and concerns [9]. Systems that look to attack microphones can inject audio to initiate unauthorized commands to voice agents using lasers, ultrasonic side channels, non-linearities in commodity smartphone microphones. Human hearing, and subsequently microphones, have been tuned to capture human sounds and speech. However, nature contains valuable information in the ultrasound band. For example, some animals can hear in ultrasound for communication, navigation, and echolocation, (e.g., dogs to 44kHz, cats to 77kHz, dolphins to 150kHz), and certain human daily activities can be recognized in this band [9].

This work explores the use of ultrasound band to detect void events with the goal of privacy-preserving automatic sound uroflowmetries at home. There is also no previous development of a database of voiding events recorded in the ultrasound frequency band, using inaudible frequencies.

III. Materials and Methods

A. Hardware Platform

To facilitate easy deployment, we use a Raspberry Pi (RPi) Zero 2 W because it presents a good compromise between cost, footprint, and computational power. We generate a novel dataset with the Ultramic384K, a high quality, professional digital audio and ultrasonic microphone (see Fig. 1.A). We capture the full spectrum allowed by the microphone with a sampling rate (fs) of 384kHz.

B. Dataset Description

In order to classify audio events into void or no-void, we developed a dataset with 501 5-second audio signals comprised of two classes: void (*Class 1*) and no-void (*Class 0*). Class 1 represents 46% of the total dataset, while Class 0 represents the remaining 54%. Each audio event is a 5-second audio clip. Next, we detail how we built the dataset for each class. Fig. 1.B shows the experimental set up used to collect the audio signals in a home toilet. The experimental procedures conform to the provisions of the Declaration of Helsinki (as revised in Edinburgh 2000).

- *Class 1* consists of 231 5-second voiding sounds collected with the Ultramic384K from 4 different volunteers at 4 different Spanish home toilets, with an average distance of the water level to the floor of 15cm. The first and last second of each voiding session is removed to avoid discontinuities and is then split into 5-second audio clips.
- *Class 0* consists of 270 5-second non-voiding sound events that typically occur in a traditional home toilet.

To visualize the separability between the two classes, we apply the dimensionality reduction technique T-distributed

TABLE I

SOUND COMPOSITION OF CLASS 0 ("NO-VOID")

Label	# samples	Label	# samples
Silence	60	Toilet flush	60
Sink	20	Bathtub	104
Wash hands	10	Brush teeth	16

neighbor embedding (t-SNE), that converts similarities between data points to joint probabilities. Results are shown in Fig. 1.C., demonstrating two distinct classes.

C. Sound Classification Model

In this section, we detail the development of a supervised machine learning model to automatically detect voiding events from outside human audible frequencies. We built a model for real-time classification of incoming audio clips into two classes: void and no-void. Our model must meet three main requirements: run inference in real-time, run on a low-power device with constrained computation capabilities, and constrain the number of false negatives to reduce missing voiding events. Considering the size of our dataset, we explore traditional ML models for the classification task. To select the best model that meets the requirements presented before, we have trained 3 different ML models to compare the performance of the classification task: a Support Vector Machine (SVM), a Random Forest (RF) with 1000 trees, and a Multi-layer Perceptron (MLP) classifier with 1 hidden layer of 100 hidden units.

D. Feature Selection

The first step is to generate features to characterize each audio sample. For each audio clip, we extract a linear-binned FFT with the SciPy python library. Because the audio signals were recorded with fs=384kHz, we can extract acoustic information up to 192kHz. To compute the FFT, we set the lower frequency limit to 16kHz to remove human audible frequencies. To select the upper limit, we quantify the importance of different spectral bands by employing feature selection methods that rank each band by its information power. We perform supervised feature selection and classification using Random Forests, which are robust and can build a model using the Gini impurity-based metric [9]. Using Gini impurity to measure the quality of our split criterion, we can quantify the weighted impurity of each feature in the tree, indicating its importance. More importantly, Gini impurity decreases the importance of features that contain information that overlaps with other features thus quantifying the "solo" predictive power for each frequency band. Fig. 2 shows each frequency component's predictive power, as measured by Gini Impurity.

This figure shows that the bins around 1kHz, 52kHz, and 70kHz contain the highest predictive power to distinguish between void and no-void acoustic events. Since the vast majority of important bands are inaudible, our system can omit speech (red) and other audible frequencies (green) entirely to preserve user privacy. However, it is also important

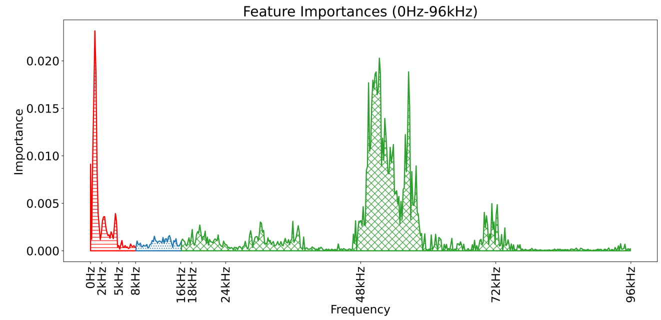

Fig. 2. Each frequency component's predictive power. While "unsafe" frequencies (red(-) speech, blue(.) high audible) have some predictive power, the vast majority are found in "safe" regions (green(x) ultrasound), especially in high ultrasound ($> 48kHz$).

to select an upper-frequency limit and characterize the trade-offs. In our experiments, we used a specialized ultrasound microphone, but most commercial microphones are designed with sampling rates of 16kHz, 48kHz, and 96kHz. Thus, we will evaluate the 3 different ML models and vary the upper-frequency limit, presenting a compromise between model performance and microphone cost and availability.

IV. RESULTS AND DISCUSSION

We next evaluate 3 different ML models, using linear-spaced FFT bins as features. We first compute 1000 bins for the full range 0-192kHz, presenting a full-spectrum high-resolution set of baseline features. We then compute a "lightweight" 100 bin feature set for two other ranges: 16-48kHz (which represents a 96kHz device) and 16-96kHz (which represents a 192kHz device). We evaluate each model for these 3 different frequency ranges. We use stratified 10-fold validation to ensure that each fold of the dataset is class balanced across labels. For each model, we report the following metrics: classification accuracy, False Positive Rate (FPR), and False Negative Rate (FNR). The FPR provides information about what proportion of the Class 1 was incorrectly classified (no-void events classified as void events). The FNR provides information of what proportion of the Class 0 was incorrectly classified (void events classified as no-void events). It is important to note that regarding the end application of our model, it is desired to have a relatively low FNR to decrease the probability of missing a voiding event. Results are shown in Table II.

Results show that for each individual model, similar performance is achieved regardless of which frequency range is used. This suggests two things: A) that voiding events can be detected robustly without the need for privacy-invasive audible frequencies, and B) expensive, specialized hardware such as the Ultramic384K ($\sim$ 400€) is not required and future deployments can utilize cheaper more readily available components. While the MLP model presents the highest accuracy (97.8%), the lightweight SVM model remains robust with an accuracy of 91.2%, presenting a viable path for performing all operations even in significantly compute-limited environments. To conclude, we select the MLP as our final

TABLE II

MODELS EVALUATION BY FREQUENCY RANGE IN TERMS OF THE CLASSIFICATION ACCURACY, STD, FNR, AND FNR.

	Frequency range	Accuracy (%)	STD	FPR (%)	FNR (%)
SVM	[16kHz - 48kHz]	91.2	3.2	1.6	7.2
SVM	[16kHz - 96kHz]	91.2	3.7	1.8	6.9
SVM	[0kHz - 192kHz]	91.2	3.4	1.8	6.9
RF	[16kHz - 48kHz]	94.7	2.0	1.2	3.9
RF	[16kHz - 96kHz]	95.8	3.5	0.9	3.2
RF	[0kHz - 192kHz]	95.8	2.2	0.9	3.2
MLP	[16kHz - 48kHz]	97.6	2.1	0.9	1.3
MLP	[16kHz - 96kHz]	97.6	1.8	0.9	1.2
MLP	[0kHz - 192kHz]	97.8	1.6	0.7	1.4

model for real-time inference and event classification, with an upper limit of the frequency bins of 96kHz (fs=192kHz). This configuration provides a compromise between model accuracy (97.8%), low false negative rate (1.2%), and cost (does not require a specialized recording device).

A. Envelope Extraction

The scientific literature has shown a good correlation between the standard and the sound-based flowmetry test in terms of the shape of the visual trace and some of the parameters obtained from the voided flow [3]. According to [10] non-bell flow shapes are carefully associated with underlying pathologies. Previous work focused on voiding events recorded with human audible frequencies. Thus, in this work we explore if we can maintain a good accuracy obtaining the voiding audio when human audible frequencies are removed without loosing the envelope shape information. We have followed the pipeline to extract the envelope presented in [3]. We process the original and filtered signal to extract the audio envelope. Results are shown in Fig. 3.

Next we quantify the difference between the 2 envelopes by measuring the mean squared error (MSE). Fig. 1.D presents the comparision. The value obtained is MSE=0.0050. We also obtain the relative error in calculating the area under the curve, obtaining 0.149%. The low values indicate a good similarity between the 2 signals. Thus it is safe to remove human-audible frequencies in the recorded audio signal to accurately obtain the voiding audio envelope.

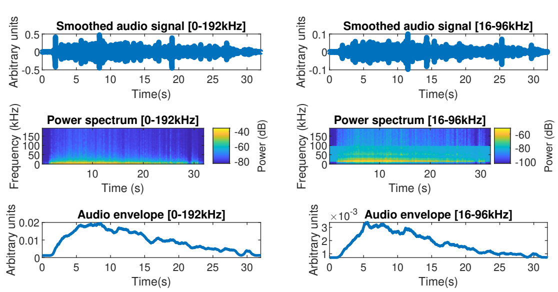

Fig. 3. Comparison of the audio uroflowmetry signals. Left figures are obtained for 0-96kHz. Right figures are obtained after removing the human audible frequencies 16-96kHz. a.u. refers to arbitrary units.

V. CONCLUSIONS

This work aims to address problems associated with current in-clinic uroflowmetry tests where the patients are required to unnaturally void "on-demand". It also aims to address the privacy concerns related to current experimental sound-based uroflowmetry platforms surrounding recordings with a human-audible microphone. This work presented a proof-of-concept acoustic Raspberry Pi-based platform that runs a novel ML classification model to detect and record voiding events automatically while preserving user privacy. To build the classification model, we built a novel dataset containing 501 instances of 5-second audio clips related to void events and other typical in-home sounds recorded with an ultrasound microphone. Results show that the MLP classifier using a frequency range of [16kHz-96kHz] to compute the FFT-based input features from the audio clips, obtains a 97.6% classification accuracy with 1.2% of FNR to distinguish between void and no-void acoustic events. As future work, we are currently working on expanding the dataset and in-hardware approaches to removing human audible frequencies.

ACKNOWLEDGMENT

This research was supported by the Spanish Ministry of Science and Innovation under the PeaceOfMind project (ref. PID2019-105470RB-C31) and the Basque Government under the UROSOUND project (ZL-2021/00649).

REFERENCES

[1] J. Cambronero Santos and C. Errando Smet, "Prevalencia de síntomas del tracto urinario inferior de llenado en pacientes varones que acuden a consulta de urología en España. La urgencia urinaria como predictor de calidad de vida," *Actas Urológicas Españolas*, vol. 40, no. 10, pp. 621–627, Dec. 2016, publisher: Elsevier.

[2] Y. J. Lee, M. M. Kim, S. H. Song, and S. Lee, "A Novel Mobile Acoustic Uroflowmetry: Comparison With Contemporary Uroflowmetry," *Journal of the Korean Continence Society*, Jan. 2021, publisher: Korean Continence Society.

[3] L. Arjona, L. Enrique Diez, A. Bahillo Martinez, and A. Arruza Echevarria, "Urosound: A smartwatch-based platform to perform non-intrusive sound-based uroflowmetry," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2022.

[4] W. Taeg-Keun, E. Sung-Jong, J. Eun-Young, P. D. Kyun, K. S. Jin, K. C. Hee, C. K. Jin, and K. Khae-Hawn, "Personalized urination activity recognition based on a recurrent neural network using smart band," *Int Neurourol J*, vol. 22, no. Suppl 2, pp. S91–100, 2018.

[5] B. J. L. e. a. Seung-min Park, Daeyoun D Won, "A mountable toilet system for personalized health monitoring via the analysis of excreta," *Nature biomedical engineering*, vol. 4, no. 6, pp. 624–635, 2020.

[6] B. Vimal, M. Surya, Darshan, V. Sridhar, and A. Ashok, "MFCC based audio classification using machine learning," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–4.

[7] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in *2018 15th Learning and Technology Conference (L&T)*, 2018, pp. 29–34.

[8] K. M. Z. M. e. a. Gupta, G., "Comparing recurrent convolutional neural networks for large scale bird species classification," *Cognitive behaviour therapy*, vol. 11, no. 17085, 2021.

[9] Y. Iravantchi, K. Ahuja, M. Goel, C. Harrison, and A. Sample, "Privacymic: Utilizing inaudible frequencies for privacy preserving daily activity recognition," ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021.

[10] R. Li, Q. Zhu, M. Nibouche, and A. Gammie, "Urine flow rate curve shapes and their descriptors," *Neurourology and Urodynamics*, vol. 37, no. 8, pp. 2938–2944, 2018.