

Improving hearing healthcare with Big Data analytics of real-time hearing aid data

Jeppe H Christensen
 Eriksholm Research Centre
 Oticon A/S
 Snekkersten, Denmark
 jych@eriksholm.com

Niels H Pontoppidan
 Eriksholm Research Centre
 Oticon A/S
 Snekkersten, Denmark
 npon@eriksholm.com

Marco Anisetti, Valerio Bellandi, Marco Cremonini
 Department of Computer Science
 Università degli Studi di Milano
 Milan, Italy
 name.surname@unimi.it

Abstract—Modern hearing aids are not simple passive sound enhancers, but rather complex devices that can log (via smartphones) multivariate real-time data from the acoustic environment of a user. In the evotion project (<http://h2020evotion.eu>) such hearing aids are integrated with a Big Data analytics platform to bring about ecologically valid evidence to support the hearing healthcare sector. Here, we present the background of the Big Data analytics platform and demonstrate that modeling of longitudinally sampled data from hearing aids can support clinical investigations with hypotheses about hearing aid usage prognosis, and support public health decision-making within the hearing healthcare sector by simulation techniques. We found, that distinct characteristics of the acoustic environment significantly modulate how hearing impaired individuals use their hearing aids. Higher sound levels and an increased sound diversity but degraded signal quality all predicts more minutes of use per hour. By simulation, we show that a projected increase in the overall sound levels by 10dB followed by a 4dB increase in noise exposure will increase the need for hearing aid use by an additional 1 hour/day across a population of hearing impaired hearing aid users.

Keywords—hearing aids, Big Data analytics, mixed models, multilevel clustered data, evidence-based public-health policies

I. INTRODUCTION

Hearing loss (HL) is a condition that affects approximately one-third of people over 65 and 5% of the world's population [1]. Disabling HL is associated with early cognitive decline in adults [2], and when unaddressed, HL restricts social integration and reduces employment and educational opportunities, hampers emotional well-being and, thus, poses an economic challenge at both the individual and national level [3]. Moreover, HL prevalence is on the rise worldwide, primarily due to increased noise exposure and increase of the aging population [1]. The leading management strategy of people suffering from HL is the provision of hearing aids (HAs). The use of HAs improve general health-related quality of life and hearing-specific quality of life associated with participation in daily activities and listening abilities [4]. However, HA users still face significant challenges, such as listening in noisy environments with poor sound qualities, and difficulty to

select among predefined HA programs and settings [5]. These challenges pose significant risks to the uptake and successful usage of HAs, which to a large part can be ascribed to ineffective or poorly fitted HAs. Ideally, HA fitting should adapt automatically to changing environmental conditions on a continuous basis rather than apply “one-size fits all” settings. On the other hand, advances in generating health-related data by devices and sensors together with the development of technology and methodologies for processing large data-sets (i.e. so-called Big Data) now permits the realization of data-driven solutions better adapted to the treatment needs of individuals [6], [7]. In this paper, we show how modern hearing aids with acoustic sensors combined with a platform for processing Big Data could improve the treatment of HL and enable evidence-based decision making within the hearing healthcare sector. We first introduce our Big Data analytics (BDA) platform by describing the architecture and the implemented catalogue-based management of analytics tasks and workflows. Next, we present a case study that demonstrates how our BDA workflows and modeling applied on longitudinally sampled real-life data from hearing aids can 1) inform the prognosis of HA uptake and usage, and 2) simulate outcomes for public-health relevant stakeholders. Finally, some relevant works in the context of Big Data platforms for healthcare and of hearing healthcare are described and our conclusions are discussed. The work presented in this paper (data collection, BDA platform implementation, and analytics) is part of the on-going research project, EVOTION, which aims at building the evidence base for the formulation of public health policies related to the prevention, early diagnosis, long-term treatment and rehabilitation of patients affected by hearing loss

II. BIG DATA ANALYTIC PLATFORM

Typically, domain experts (e.g., policy-makers, public-health experts, healthcare professionals) are not familiar with data science methods and are not always supported by data scientists to help produce analytical evidence from Big Data. Thus, there is a need for an easy to use platform that enables the potential and enlarges the real application scenarios of Big Data for such relevant stakeholders. Our

This work was partly supported by EU-funded project EVOTION (No. H2020-727521)

Big Data analytics (BDA) platform is tailored for trial-and-error style of analysis in order to let domain experts execute complex analyses intuitively or choose among pre-defined workflows to support proposals with numerical evidence.

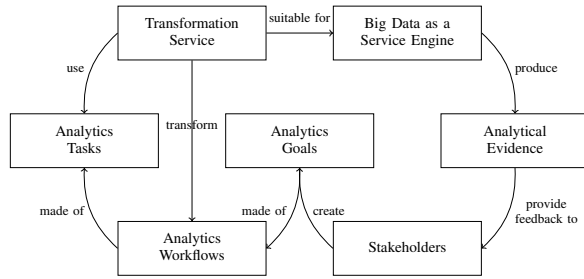


Figure 1. A methodological view of our Big Data Analytics as a service.

More specifically, the approach depicted in Figure 1 represents a stakeholder that defines an analytical goal by means of a set of workflows made of combined pre-defined analytics tasks. The compossibility of tasks are obtained due to the fact that they share the same interface exposing three methods: *init*, *run* and *post-processing*. The *init* method is used to process the parameters requested by the task's algorithm, default values, and the input data flow. It is followed by the *run* method used for implementing the algorithm and the *post-processing* method preparing the output, for example to make it suitable for consecutive tasks within the same workflow. The tasks are implementations of either readily available or customized algorithms (e.g. for data preparation, cleaning, statistics, mining). The workflows can be defined by the stakeholders in a intuitive declarative way and are automatically converted into executable workflows by a transformation of their declarative description into code, which can be executed via our BDA engine (Transformation Service). The output of an executable workflow can be used as feedback to a consecutive workflow or to refine the current workflow structure and parameters. To support the above process, the BDA platform has to offer i) a declarative way to specify a workflow made of analytics tasks, ii) Big Data analytics as a service engine to execute the workflow, iii) a set of commonly used analytics tasks usable as building blocks, iv) simple paradigms to define additional analytical tasks that can be easily embedded into complex workflows, v) a simple and effective way to aggregate and present results and summary statistics. Our BDA platform is a concrete advanced incarnation of the Toreador Model-based Big Data Analytics as a Service [8], where the landing platform is based on the Apache Foundation ecosystem [9]. It is enhanced with additional custom services in order to support the above requirements, the trial-and-error methodology and to be offered as a service to stakeholders.

Our BDA platform is composed of:

- *Dashboard*: Enables a stakeholder to interactively define a declarative workflow and visualize results.

- *BDA Engine*: An Ambari cluster of 5 nodes based on the Apache ecosystem (Spark, Oozie, Hive, Hadoop etc.).
- *Task catalogue*: A lists of tasks for which an executable implementation is available. For instance, an entry in the Task Catalogue could be `Spark_k-mean`, representing the specific Spark-based implementation of K-mean. Every task is linked to the executable implementation deployed within the BDA cluster and to a meta description of environmental variable and parameters requested to be executed. It offers APIs for managing the stored task implementations (add, delete, modify).
- *Workflow catalogue*: Composed by three sub-components: i) a Workflow repository containing the set of available executable workflows, ii) a workflow scheduler for the execution of workflows according to the given scheduling (i.e., periodic, upon request, or driven by data changes), and iii) the workflow manager, which is responsible for keeping track of the running workflows.
- *Transformation Service*: to transform the declarative workflow to an executable workflow.

A stakeholder can design, using our Dashboard, an analytical goal (e.g. hypothesis about dependencies between data variables) building an analytics workflow based on a set of tasks provided by the platform's task catalogue. Available tasks can be selected and possibly combined, sequentially, in parallel, or as alternative, to form a declarative workflows. A declarative workflow is defined using Oozie language (based on xml). The dashboard graphically assist stakeholders in building the Oozie workflow. The data flow is represented as a subsequence of intermediate csv files on the HDFS that are passed as arguments to the relevant tasks following the workflow structure. Even if this approach introduces a latency due to the use of intermediate csv files, performances are not a strong requirement for the application considered in this paper and csv files allow having i) a more easy to use workflow representation, ii) intermediate results to be inspected aligned with the spirit of a trial-and-error analytics.

Prior to execution, the declarative Oozie workflow requires to be transformed into an executable workflow. This transformation requires the definition of environmental variables for the execution engine, mapping of tasks' arguments and mapping of data flow to the BDA deployment. During workflow execution, notifications are sent to the dashboard to inform the stakeholder about the ongoing analytics processing. When the execution terminates, results are visualized on the dashboard. At that point, the user can decide if the returned results should be used as input for a new workflow (e.g., build a model and use it as input for prediction or simulation).

III. HEARING AID CASE STUDY

In the following section, we demonstrate how our analytics methodology could be applied to real-life longitudinal data concerning hearing aid (HA) users. The achieved results will i) shed light on individualized HA usage and preferences; ii) enable a better understanding of factors influencing HA outcomes of a population; and iii) support public-health policy makers with ecologically valid evidence and simulations.

A. Background

Several studies have found that self-reported factors predict the success of HA uptake as measured by hours of usage per day [10]. However, these studies are often biased by the fact that people tend to overestimate their actual HA usage [11]. Therefore, it is important to use objective data-logging of HA usage for accurate analyses. In addition, patterns of HA usage (i.e., “how” the HA has been used) are at least as important in predicting HA outcomes as the duration of HA use (i.e., “how long” the HA was used) [12]. Combined, these observations advocate the use of real-time longitudinal data-logging to enable a more detailed analysis of which environmental factors affect the daily usage patterns (i.e., “how” the HA has been used).

B. Methods and Data

In the clinical study part of EVOTION [13], participants with varying degrees of HL are provisioned a pair of Oticon EVOTION hearing aids (based on the Oticon OpnTM) and a Bluetooth connected smart-phone. During normal HA use, the smart-phone logs a data-vector of 15 acoustic parameters (recorded by the HA) and a timestamp every minute. The acoustic parameters covers the momentary sound pressure level (*SPL*); the signal-to-noise ratio (*SNR*); the noise floor (*Nf*); the modulation index (*MI*); and the modulation envelope (*ME*) in three frequency bands: 0-1.3kHz; 1.3-4.1kHz; and 4.1-10kHz.

Using data from EVOTION, we hypothesize that HA usage of a hearing impaired population is predicted by variables of the acoustic environment. Specifically, HA users proactively use their HAs more if they are faced with loud acoustic environments but low quality signals. In addition, corroborating previous studies [14], we also hypothesize that the diversity of the sound environment is positively related to usage times, which would indicate that individuals who use their HAs more also lead a more diverse lifestyle.

Note that, to preserve privacy and enable data-sharing, the data sample for this study is synthesized from the EVOTION data-repository [15] using DataSynthesizer [16]. DataSynthesizer synthesizes new data based on Bayesian networks, thus, higher order statistical and co-variance structures are preserved. In addition, prior to synthesizing, all acoustic parameters are aggregated for each hour, day, and participant as both the mean and the variance. This means that absolute

times and unique data-points cannot be inferred from the synthesized data sample. In total, the data source included 530,167 observations of 25 variables from 54 participants spanning up to 54 days and a total of approximately 5,000 hours of HA usage.

C. Analytical workflow

The following computational methods (*tasks*) were applied to test the hypothesis. Also see Algorithm 1 below for a pseudo-code description of the first three tasks.

Data selection and aggregation: Relevant grouping parameters (i.e. participant ID, age) together with usage logs (i.e. timestamps) and parameters of the acoustic environment are retrieved and aggregated according to the needs of the hypotheses being tested.

Dimensionality reduction: The acoustic environments encountered by HA users are described by 30 parameters (hourly mean and hourly variance of the five acoustic characteristics, each measured in three frequency band). To reduce the amount of dimensions, and to ensure identifiability in subsequent modeling, factor analysis is applied to identify distinct and interpretive latent factors. Each factor is defined as a weighted mean of the standardized parameters exhibiting a factor loading larger than 0.6

Model selection: The predictive variables (here, the factors resulting from dimensionality reduction) are included as fixed effects and model selection identifies the most parsimonious parameterization of the random effects using generalized linear mixed-modeling (GLMM, see next section). Each combination of random effects (i.e. random intercepts and slopes) are compared with regards to Akaike’s Information Criteria (*AIC*). The model with the lowest *AIC* is chosen for further evaluation and predictions.

Model evaluation: The quality of the model fit is evaluated by inspecting the residuals and the predictive power of the final model is evaluated by the amount of explained variance. In addition, main effects of the regression coefficients for each predictive variable are tested for significance by a MANOVA using Wald’s chi-square tests (see Table I).

Model simulation: The parameterization of the final model is re-fitted to the data with a binominal GLMM. This enables simulation from the model to new data (since the predicted response is bound by 60 minutes).

D. Modeling details

We used generalized linear mixed models (GLMMs) as a unified framework for testing the significance of predictive variables and to simulate outcomes of HA usage for new, hypothetical, acoustic data. In GLMMs, inter-observational correlations and group-differences are dealt with by including random effects besides the independent predictive variables, which are treated as fixed effects with regression coefficients [17]. Fixed effects are contributed

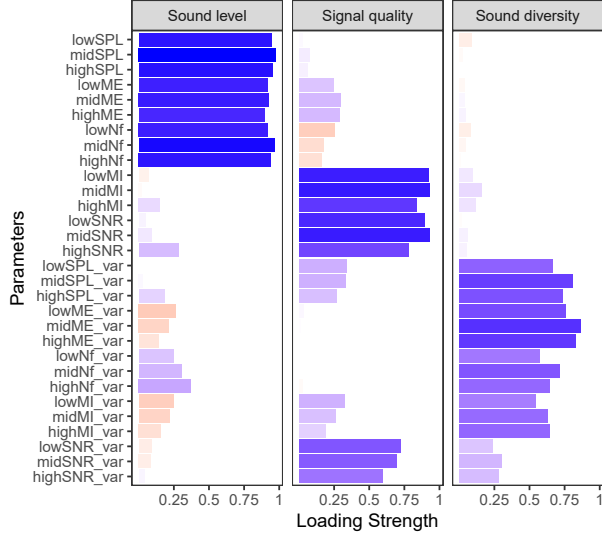


Figure 2. Factor loadings for the parameters describing the acoustic environment with five characteristics. The characteristics were: SPL = Sound pressure level; ME = Modulation Envelope; Nf = Noise floor; MI = Modulation Index; SNR = Signal-to-noise ratio. The indexing "low", "mid", and "high" refers to the three measured frequency bands. "_var" indicates that the corresponding parameter represents the hourly variance. The three factors explained 29%, 23%, and 21%, respectively, of the total variance between the parameters. Only parameters with factor loading larger than 0.6 were included with the associated factor for subsequent modeling workflow.

to variables that are assumed to exhibit constant slopes and intercepts with the response variable regardless of any hierarchical grouping - that is, fixed effects are considered within-participants effects that are general for the whole tested population. On the other hand, random effects allow the slope and/or intercept to vary due to grouping variables (such as age or gender). For example, we might expect that HA usage within a day vary between participants due to unmeasurable variables such as preference, which would suggest including participant IDs as a random variable for the intercept (i.e. a between-participant effect). We might also believe that the strength of the associations between acoustic parameters and HA usage vary between individuals due to differences in sensitivity, which in turn would suggest modeling the fixed effects of the acoustic parameters with random slopes per individual. Thus, individual predictions can be obtained by combining the fixed with individual random effects coefficients, and population-wise predictions can be obtained by inspecting the fixed effects and the distribution of random effects.

Formally, the dependent response variable (here, HA usage), y , is modeled on the following form:

$$y = X\beta + Z\gamma + \epsilon, \quad (1)$$

where y is a $N \times 1$ column vector of N observations; X is a $N \times p$ matrix of the p predictor variables (see first column in Table I); β is a $p \times 1$ column vector of the fixed effects

Algorithm 1 Modelling workflow's tasks pseudocode

Require: Matrix of static data ($\Psi_{r \times c}$); matrix of longitudinal data ($\Omega_{n \times k}$); hypothesis $H = [\hat{a}, \hat{b}]$ with $\hat{a} = (a_1, \dots, a_j)$ set of j predictor labels and $\hat{b} = (b_1, \dots, b_q)$ set of q grouping labels.

- 1: T = timestamp column of Ω
- 2: S = ID column of Ω
- 3: $Z = \Psi_{*, c \cup \hat{b}}$
- 4: **procedure** DATA AGGREGATION
- 5: $X_1 = \emptyset, X_2 = \emptyset, y = \emptyset, N = 0$
- 6: **for all** $i \in \text{hour}(T), ii \in \text{day}(T), iii \in S$ **do**
- 7: $N = N + 1$
- 8: $n' \subset n$ where $\text{hour}(T) = i \wedge \text{day}(T) = ii \wedge S = iii$
- 9: $X_1(N) = \text{rowMean}(\Omega_{n', k=\hat{a}})$
- 10: $X_2(N) = \text{rowVariance}(\Omega_{n', k=\hat{a}})$
- 11: $y(N) = \text{rowSum}(\Omega_{n', k=T})$
- 12: $X_{N \times 2j} = X_{N \times j} = \text{join longFormat}(X_1 X_2)$
- 13: **procedure** DIMENSIONALITY REDUCTION
- 14: $\Lambda_{F \times J} = \text{factorAnalysis}(X, F = 3)$
- 15: **for** $f = 1, \dots, F$ **do**
- 16: $\hat{X}_{N, f} = J^{-1} \sum_{p=1}^J X_{N, p} \times \Lambda_{f, p}$
- 17: **procedure** MODEL SELECTION
- 18: $AIC = \emptyset, \hat{d} = \emptyset$
- 19: **for each** $b \in \hat{b}$ **do**
- 20: $m_{re} := \text{glmfit}(y \sim \hat{X}\beta + Z_{*, \hat{b}=b}\gamma + \epsilon)$
- 21: to AIC add $\text{aic}(m_{re})$
- 22: $\hat{b} = \text{rank}(\hat{b})$ ascending by AIC
- 23: $m_0 := \text{lm}(y \sim \hat{X}\beta)$
- 24: $m_{full} = m_0$
- 25: **for each** $b \in \hat{b}$ **do**
- 26: to \hat{d} add b
- 27: $m_{test} := \text{glmfit}(y \sim \hat{X}\beta + Z_{*, \hat{b} \cup \hat{d}}\gamma + \epsilon)$
- 28: **if** $\text{aic}(m_{test}) < \text{aic}(m_{full})$ **then**
- 29: $m_{full} = m_{test}$
- 30: **else** from \hat{d} remove b

regression coefficients (see second column in Table I); Z is the $N \times q$ design matrix for the q random effects; γ is a $q \times 1$ vector of random effects; and ϵ is the $N \times 1$ column vector of the residuals, i.e. the part of y not explained by our model. We implemented the identified factors of the acoustic environment from dimensionality reduction (see Figure 2) as fixed effects, and allowed for random intercepts and slopes due to participant ID, hour-of-day, and number of days since HA fitting (i.e., days since study enrollment).

E. Results

Factor analysis of the 30 acoustic parameters yielded three distinct factors describing the acoustic environment by the *sound level*, *signal quality*, and *sound diversity* (see Figure 2), and explaining 73% of the total variance between the parameters. Sound level is given in dB, signal quality is given in a dB ratio (i.e. strength of an acoustic signal compared to noise), and sound diversity is given in dB². Time-courses of the factors are shown as grand averages in Figure 3 (right). Both the Sound level and the signal quality exhibit noteworthy fluctuations over time: the sound level is highest between 8am and 4pm and then

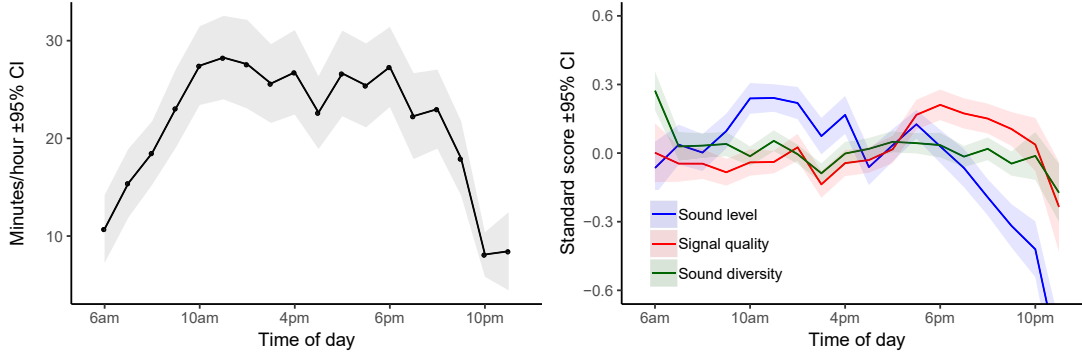


Figure 3. Grand average HA usage (left) and the factors describing the acoustic environment (right). Each acoustic environment factors is defined in standardized score, that is - the numerical values represents how far from the mean (mean = 0) an observation is in standard deviations. Shaded area represents the 95% CI.

drops dramatically after 6pm. In addition, the signal quality peaks from 6pm to the end of the day, but is otherwise stable. These characteristics are based on population-wise aggregation of the data, and thus, reflects the "acoustical day" of a typical hearing aid user, that is - they indicate routine activities taking place at roughly the same time every day (such as lunch, dinner, working, and leisure activities).

Figure 3 (left) shows the grand average HA usage (from 6am to midnight) with the 95% CI indicated by the shaded areas. Usage is fairly stable just below 30 minutes/hour from 10am to 8pm. This indicates that the HAs were turned off for around 50% of all pooled participant days. Model selection identified the optimal (most parsimonious) model by comparing each parameterization of the random effects with a *NULL* hypothesis that does not allow random variation in neither slopes nor intercepts. The best-fitting model improved the *AIC* from the *NULL* model with $\Delta AIC = 307$. In total the model captured 32% of the variance between observed HA usage times. The most contributing random effects were the intercepts for participant ID and hour-of-day, explaining a total of 16% of the variance. The regression coefficients of the optimal model for predicting HA usage are listed in Table I. All three factors are significantly associated with HA usage. The sound level and sound diversity are positively related to usage and the signal quality is negatively associated to usage. The constant in Table I indicates that, on average, the HAs were used for 41.6 minutes every hour they were turned on (i.e. the number of observations = 7,665).

In Figure 4, the distribution of total daily HA usage per individual is plotted together with the rather well-fitting predictions of the GLMM models. The grand average total daily HA usage in the observed data was 6.16 (SD = 3.7) hours/day.

F. Simulation for policy making scenario

Evidence-based decision-making, whether in healthcare sectors or in government planning of indoor working environments, relies on inferences from reliable statistics of population norms. In the current study, the identified optimal model of HA usage versus the acoustic environment represents the ground truth (i.e., population norms), that is - the structure observed in a general population of hearing impaired hearing aid users. We can now simulate what the expected daily total HA usage might be for a particular sub-population of hearing impaired hearing aid users exposed to a more extreme acoustic environment than what is normal. For example, say urban planning projects an increase in everyday acoustic noise due to changed requirements for noise prevention initiatives. This will lead to an overall increased sound level and worsened signal quality due to more noise. In Figure 4, the blue distribution reflects the

Table I
REGRESSION COEFFICIENTS AND SIGNIFICANCE FOR PREDICTING HA USAGE FOR EACH (SCALED AND CENTERED) PREDICTOR.

<i>Dependent variable:</i>	
HA usage (minutes/hour)	
<i>Sound level</i>	1.398* (0.700)
<i>Signal quality</i>	-3.046** (0.659)
<i>Sound diversity</i>	1.288* (0.625)
Constant	41.629** (1.511)
Observations	7665
<i>Note:</i>	*p<0.05; **p<0.01

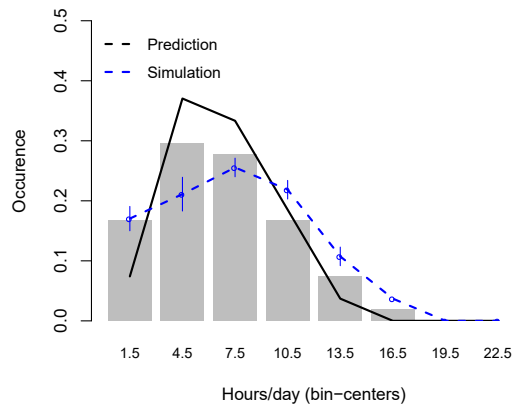


Figure 4. Observed versus predicted and simulated distribution of daily total HA usage among individuals in the data-sample. The simulation assumes a +10dB and -4dB change to the sound level and signal quality, respectively.

expected daily HA usage in a population that is exposed to 10dB higher sound levels and 4dB worse signals. The mean expected daily HA usage is now 7.33 (SD = 4.07) hours/day. This indicates that the specific change in acoustic environment have lead to an increased need for HA treatment by ≈ 1 hour.

IV. DISCUSSION AND CONCLUSIONS

Using a general linear mixed-models approach, we demonstrated the capabilities of our BDA platform, and we identified significant relationships between the acoustic environment of a HA user and absolute usage times in patients with hearing loss. More HA usage is associated with a higher-than-normal sound level, less-than-normal signal quality (i.e. more noise) and higher-than-normal sound diversities. This indicates that users are protectively engaging with their HAs in potentially difficult acoustic situations. Moreover, the diversity of the sound environment can be considered a proxy for life-style activity levels [18]. Thus, a more active day seems to be associated with increased HA usage. This finding corroborates earlier studies, showing that when using the HAs more often, and reporting greater satisfaction, older adults indicated more diverse listening situations [14]. One possible future exploitation could be to link HA usage and the sound diversity with the movement activity of individual hearing aid users measured through relative GPS coordinates and by this establish more general health-related evidence. The proposed work enables evidence-informed public-health policy making within the hearing healthcare sectors. For example, from the current results, a stakeholder could argue that the provision scheme of HAs should not only be guided by age and/or hearing loss but also by the acoustical environment a HA user are

expected to be exposed to. Moreover, healthcare professionals could use the information about how the acoustic environment affects HA usage to better guide and personalize consultation. Finally, simulations can help to estimate the expected effect of provisioning HAs to specific populations, or to estimate the change in need for hearing loss treatment due to changes in environmental factors. To conclude, in this work we have presented one of the first so far experimental application of a Big Data analytics as a service platform designed to process health data related to hearing impaired patients. The work demonstrates the possibilities that a data-driven approach to healthcare could provide by considering a restricted sample of user data. Our findings are aligned with earlier studies and confirm the suitability of a generalized linear mixed-models approach.

V. RELATED WORKS

Evidence-based decision making have attracted remarkable attention in the last years both in the medical and in the data science communities [19]–[22]. Big Data techniques are considered useful tools in translating personalized medicine initiatives into clinical practice. For example, medical analyses could be personalized by linking health-related data (e.g., medication list and family history) to personal data (e.g., income, education, place of living, dietary habits, sport activity, entertainment), and to environmental data (e.g. polluted or noisy workplaces). Moreover, novel streams of data are increasingly available from sensors and medical devices, providing opportunities to study correlations between multiple factors related to healthcare [23], [24]. However, as also indicated in a report by the Institute of Medicine, several open problems still remains [25]. With respect to hearing healthcare, Big Data approaches are still underexplored. The most relevant work, up to now, was presented by Mellor, Stone, & Keane [26] as "proof-of-concept" examples of how BDA methods could be used to extract useful information from healthcare data, and they went on showing how clustering methods applied to a large data-set containing a number of variables concerning hearing aid users could be used to mine for interesting patterns. For example, the authors found that specific hearing aid settings (i.e. gain-reduction profiles) could be associated with specific distributions of sound pressure levels of the acoustic environment. However, no other attempts at integrating a BDA framework with real-time data-feeds from hearing aids have been made to date.

ACKNOWLEDGMENT

This work was partly supported by EU-funded project EVOTION (contract n. H2020-727521) and by the program "piano sostegno alla ricerca 2015-17" funded by Università degli Studi di Milano.

REFERENCES

- [1] World Health Organization, *Global costs of unaddressed hearing loss and cost-effectiveness of interventions.*, 2017, oCLC: 975492198.
- [2] B. O. Olusanya, K. J. Neumann, and J. E. Saunders, "The global burden of disabling hearing impairment: a call to action," *Bulletin of the World Health Organization*, vol. 92, no. 5, pp. 367–373, May 2014.
- [3] B. S. Wilson, D. L. Tucci, M. H. Merson, and G. M. O'Donoghue, "Global hearing health care: new findings and perspectives," *The Lancet*, vol. 390, no. 10111, pp. 2503–2515, Dec. 2017.
- [4] M. A. Ferguson, P. T. Kitterick, L. Y. Chong, M. Edmondson-Jones, F. Barker, and D. J. Hoare, "Hearing aids for mild to moderate hearing loss in adults," *Cochrane Database of Systematic Reviews*, Sep. 2017.
- [5] A. McCormack and H. Fortnum, "Why do people fitted with hearing aids not wear them?" *International Journal of Audiology*, vol. 52, no. 5, pp. 360–368, May 2013.
- [6] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [7] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, no. 1, p. 3, 2014.
- [8] C. Ardagna, V. Bellandi, M. Bezzi, P. Ceravolo, E. Damiani, and C. Hebert, "Model-based big data analytics-as-a-service: Take big data to the next level," *IEEE TSC*, 2018.
- [9] M. M. Rathore, A. Paul, A. Ahmad, M. Anisetti, and G. Jeon, "Hadoop-based intelligent care system (hics): Analytical approach for big data in iot," *ACM Transactions on Internet Technology (TOIT)*, vol. 18, no. 1, p. 8, 2017.
- [10] B. Gopinath, J. Schneider, D. Hartley, E. Teber, C. M. McMahon, S. R. Leeder, and P. Mitchell, "Incidence and Predictors of Hearing Aid Use and Ownership Among Older Adults With Hearing Loss," *Annals of Epidemiology*, vol. 21, no. 7, pp. 497–506, Jul. 2011.
- [11] E. Perez and B. A. Edmonds, "A Systematic Review of Studies Measuring and Reporting Hearing Aid Usage in Older Adults since 1999: A Descriptive Summary of Measurement Tools," *PLoS ONE*, vol. 7, no. 3, p. e31831, Mar. 2012.
- [12] A. Laplante-Lévesque, C. Nielsen, L. D. Jensen, and G. Naylor, "Patterns of Hearing Aid Usage Predict Hearing Aid Use Amount (Data Logged and Self-Reported) and Overreport," *Journal of the American Academy of Audiology*, vol. 25, no. 2, pp. 187–198, Feb. 2014.
- [13] G. Dritsakis, D. Kikidis, N. Koloutsou, L. Murdin, A. Bibas, K. Ploumidou, A. Laplante-Lévesque, N. H. Pontoppidan, and D.-E. Bamiou, "Clinical validation of a public health policy-making platform for hearing loss (EVOTION): protocol for a big data study," *BMJ Open*, vol. 8, no. 2, Feb. 2018.
- [14] B. Williger and F. R. Lang, "Hearing Aid Use in Everyday Life: Managing Contextual Variability," *Gerontology*, vol. 61, no. 2, pp. 158–165, 2015.
- [15] B. Ye, I. Basdekis, M. Smyrlis, G. Spanoudakis, and N. Koloutsou, "A big data repository and architecture for managing hearing loss related data."
- [16] H. Ping, J. Stoyanovich, and B. Howe, "DataSynthesizer: Privacy-Preserving Synthetic Datasets," in *Proc. of SSDBM '17*. Chicago, IL, USA: ACM Press, 2017, pp. 1–5.
- [17] J. Fox and J. Fox, *Applied regression analysis and generalized linear models*, third edition ed. Los Angeles: SAGE, 2016, oCLC: ocn894301740.
- [18] Y.-H. Wu, E. Stangl, X. Zhang, and R. A. Bentler, "Construct Validity of the Ecological Momentary Assessment in Audiology Research," *Journal of the American Academy of Audiology*, vol. 26, no. 10, pp. 872–884, Nov. 2015.
- [19] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England journal of medicine*, vol. 375, no. 13, p. 1216, 2016.
- [20] M. J. Khoury and J. P. Ioannidis, "Big data meets public health," *Science*, vol. 346, no. 6213, pp. 1054–1055, 2014.
- [21] M. Anisetti, V. Bellandi, M. Cremonini, E. Damiani, and J. Maggesi, "Big data platform for public health policies," in *Proc. of IEEE SmartWorld*. IEEE, 2017, pp. 1–6.
- [22] M. Anisetti, C. Ardagna, V. Bellandi, M. Cremonini, F. Frati, and E. Damiani, "Privacy-aware big data analytics as a service for public health policies in smart cities," *Sustainable Cities and Society*, vol. 39, pp. 68–77, 2018.
- [23] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Transactions on information technology in biomedicine*, vol. 10, no. 1, pp. 119–128, 2006.
- [24] R. K. Pathinarupothi, P. Durga, and E. S. Rangan, "Iot based smart edge for global health: Remote monitoring with severity detection and alerts transmission," *IEEE Internet of Things Journal*, 2018.
- [25] *The Global Use of Medicines: Outlook Through 2016*. MS Institute, 2016.
- [26] J. Mellor, M. A. Stone, and J. Keane, "Application of Data Mining to a Large Hearing-Aid Manufacturer's Dataset to Identify Possible Benefits for Clinicians, Manufacturers, and Users," *Trends in Hearing*, vol. 22, Jan. 2018.