Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 9-12, 2020, Leuven, Belgium

Comprehensive insights into the TrackYourTinnitus database

Robin Kraft[a,b,*], Michael Stach[a], Manfred Reichert[a], Winfried Schlee[c], Thomas Probst[d], Berthold Langguth[c], Marc Schickler[a], Harald Baumeister[b], Rüdiger Pryss[e]

[a]*Institute of Databases and Information Systems, Ulm University, Germany*
[b]*Department of Clinical Psychology and Psychotherapy, Ulm University, Germany*
[c]*Clinic and Policlinic for Psychiatry and Psychotherapy, University of Regensburg, Germany*
[d]*Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Austria*
[e]*Institute of Clinical Epidemiology and Biometry, University of Würzburg, Germany*

Abstract

The ubiquity of smart mobile devices facilitates data collection in the healthcare domain. Two of the concepts, which can be applied in this context, are mobile crowdsensing (MCS) and ecological momentary assessment (EMA). TrackYourTinnitus (TYT) is an advanced mobile healthcare platform that combines both concepts enabling the monitoring and evaluation of the users' individual variability of tinnitus symptoms. This paper describes the underlying data set and structure of the TYT mobile platform and highlights selected issues whose investigation provides advanced insights into the users of this mobile platform as well as their data.

Keywords: Mobile crowdsensing (MCS); ecological momentary assessment (EMA); mobile healthcare application; chronic disorder; tinnitus

1. Introduction

As smart mobile devices (e.g., smartphones and smartwatches) are becoming ubiquitous, new opportunities for collecting data emerge in the healthcare domain. Mobile crowdsensing (MCS) and ecological momentary assessment (EMA) constitute two fundamental concepts that benefit greatly from these advancements [6, 4]. Both concepts can be used in combination in the form of mobile applications to correlate EMA data with sensor measurement data, enabling valuable additional insights into the patients' situation [3]. The *TrackYourTinnitus (TYT)* mobile platform utilizes MCS and EMA to track the individual tinnitus of a user as well as to monitor and evaluate its variability over

* Corresponding author.
E-mail address: robin.kraft@uni-ulm.de

time. This work describes the data set and structure that underlies the TYT platform, highlights insights that we gained from the gathered data, and discusses technical issues and limitations.

A recent review on mobile health crowdsensing [12] has confirmed that, although an increasing amount of research is conducted in the field, only few experience reports exist that were gained over multiple large-scale, long-running projects like TYT and related MCS-EMA endeavors. Although several MCS and EMA mobile applications with research background (e.g., *Moodpath*[1]), open source platforms (e.g., *PACO*[2]), and commercial tools such (e.g, *ilumivu*[3]) exist, only little information is publicly available about the data structures underlying these apps as well as the insights that can be gained from them. This makes it quite challenging to compare existing approaches or to start new projects and research in this area. The work at hand addresses this issue by providing the following contributions:

- A comprehensive description of the TYT data set and its structure is provided. This allows (1) better understanding and reproducing the results obtained from this data set, (2) comparing these results with the data from other projects, and (3) serving as basis for new projects.
- Descriptive statistics for different aspects in the data set are provided. Particularly, insights that can be gained from these data and that crave for additional investigations are presented.
- Future analyses, technical issues, and limitations of the data set are discussed.

The remainder of this paper is organized as follows. In Section 2, background information on the TYT platform as well as the underlying data set is provided. Descriptive statistics and the insights gained from these data are presented in Section 3. Section 4 discusses future evaluations and analyses, technical issues and limitations of the current data set. Finally, Section 5 concludes the paper with a summary and an outlook.

2. Background information

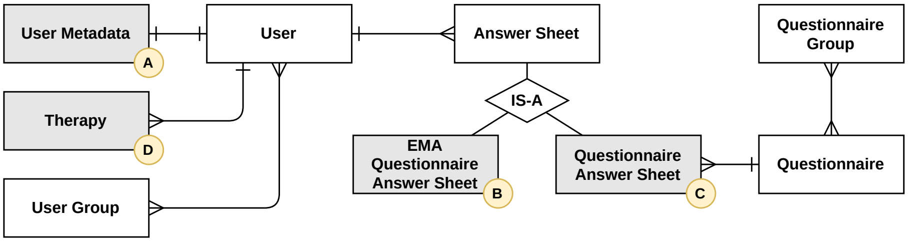
In this section, background information on the TYT mobile platform is provided and the underlying data set as well as its structure is described.

The TrackYourTinnitus (TYT) platform has been in operation since 2014 and been continuously evolved. It is composed of a registration and information website[4], two native mobile applications (iOS and Android), and a central backend that stores the collected data in a relational database. The mobile applications track the individual tinnitus perception of users by asking them to complete ecological momentary assessment (EMA) [11] questionnaires for tinnitus assessment at randomly selected times of the day [8]. The exact procedure was described in [3]. Tinnitus is the perception of an internal sound when there is no corresponding external noise. The symptoms are subjective and vary over time. Therefore, TYT was implemented to monitor and evaluate the variability of symptoms over time based on EMA and MCS [10].

Before describing the data set itself, the data collection & processing procedure is briefly outlined. The data is first collected by the mobile applications, stored on the mobile device and, if the device is connected to the internet, eventually transmitted to the backend. The backend then validates and assigns the transmitted data to the model classes, and finally stores the data in a relational database. The relational data model has been chosen to enforce data consistency based on the well-known ACID transaction guarantees. For this work, a snapshot of the data set was extracted on 23 March 2020 and, in order to reduce noise in the data, subsequently cleaned. This snapshot represents the basis of the descriptive statistics of this work.

In order to illustrate the data set, an excerpt of the TYT data model representing the entered user data is depicted as Entity-relationship model in Fig. 1. Since TYT uses an MCS approach to collect data, the user table is either in direct or indirect relation with all other entities. To separate operational data and user data, table *User Metadata* (see Fig. 1, Ⓐ) is responsible for storing demographic data (e.g., country, mobile device information, etc.); currently, it comprises data of 4929 unique users. To conduct studies, users may be assigned to one or more user groups.

[1] https://mymoodpath.com/
[2] https://pacoapp.com/
[3] https://ilumivu.com/
[4] https://www.trackyourtinnitus.org/



Data Set

Table	Entries	Unique Users
A	4929	4929
B	77488	2632
C	9262	3193
D	143	84

Fig. 1: Entity-relationship model (ERM) of the TYT database. Tables A–D are considered in detail in the scope of this work. The respective number of entries in the database and the number of unique users who submitted these data are shown in the table to the right of the figure.

As explained in Section 1, TYT utilizes MCS and EMA to collect data on the user's individual tinnitus. To be more precise, standardized questionnaires (see Section 3.1) are used as a data collection instrument. The answers, in turn, are stored in *Answer Sheets*. The latter can be divided into an answer sheet for (1) a fixed questionnaire related to the momentary assessment and (2) a dynamically composable questionnaire. The first type is denoted as *EMA Questionnaire Answer Sheet* (see Fig. 1, Ⓑ). It consists of the user's answers, which enable the tracking of the individual tinnitus, an environmental sound level measurement, information about the triggered system notification on the device, and the current user agent. Note that the EMA questionnaire is a predefined questionnaire that has not been modified since the start of the TYT platform in order to ensure comparability of the answers. Since the EMA questionnaire is triggered on a daily basis, it holds the largest share of the answers sheet with 77488 stored entries of 2632 users. The second type of answers sheets, denoted as *Questionnaire Answer Sheet* (see Fig. 1, Ⓒ), holds the answers of dynamically composable questionnaires. To be more precise, the latter include all other questionnaires that, for example, gather demographic user data. Those questionnaires can be grouped into complex data collection schemes if needed. Currently, the TYT data set consists of 9262 *Questionnaire Answer Sheets* of 3193 unique users.

Additionally, TYT users may store information about their on-going treatments and therapies. The *Therapy* table (see Fig. 1, Ⓓ) contains information about the therapy type (e.g., dental treatment, sport, hearing aid, etc.), the current medication of the patient, personal notes, and the start as well as end dates of the treatment. Note that this feature is only available on the registration and information website, which is reflected in the small number of therapies in the data set (n = 143).

3. Potential insights into the TrackYourTinnitus database

Based on the extracted data set described in Section 2, different aspects of the TYT platform are investigated and explored for potential insights in the following.

3.1. Socio-demographics and sub-populations

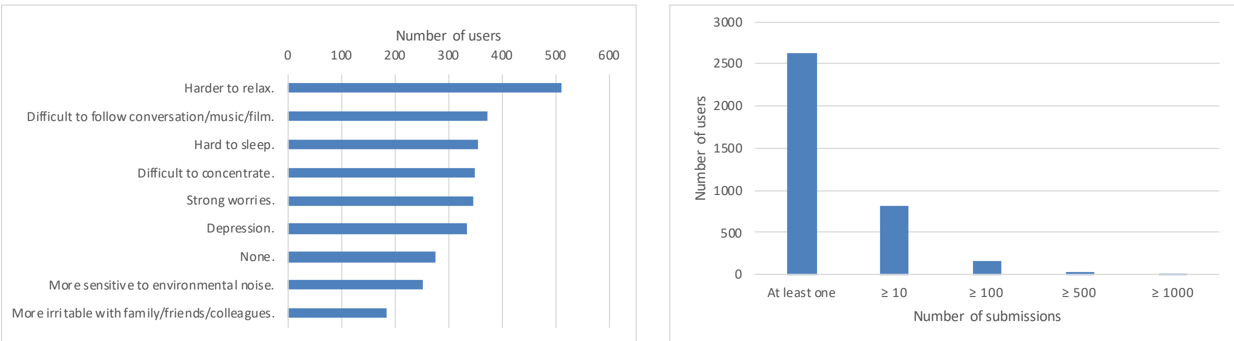
In order to assess the socio-demographic data of the user population of TYT, the data collected with the three initial questionnaires when users start the mobile application for the first time was evaluated. The *Mini Tinnitus Questionnaire* (n = 3179) assesses the initial distress of users, whereas the *Tinnitus Sample Case History Questionnaire (TSCHQ)* (n = 3009) corresponds to a standardized questionnaire for tinnitus case history [5], including several questions on the socio-demographic background of patients as well as their tinnitus history. Finally, the *Worst Symptom* questionnaire (n = 3072) asks the users to select their subjectively worst symptom from a list of symptoms (or to indicate that none of the symptoms apply to them).

The answers for the Mini Tinnitus Questionnaire have a mean sum score of 13.64 (s = 6.11). The correlation between the individual sum scores and the answers of the EMA questionnaires can then be investigated (e.g., the perceived tinnitus loudness or distress). Furthermore, socio-demopraphic data assessed with the TSCHQ are shown in Table 1. For different questions, it can be examined if there are differences between the sub-populations. For instance, it can be investigated whether there are differences in the perception of tinnitus between female and male users. Finally, the distribution of the subjectively worst symptom of each user is shown in Fig. 2a. It can be investigated whether there is a correlation between this worst symptom and the course of the tinnitus assessed with the EMA questionnaire. Additionally, as the user is asked in the EMA questionnaire whether he perceives this worst symptom at that very moment, it can be investigated how the various symptoms evolve over time.

Characteristic	**Value**	**Frequency** (n = 3009)	**%**
Gender	Male	2032	67.96
	Female	958	32.04
Handedness	Right	2176	72.7
	Left	364	12.16
	Both Sides	453	15.14
Familiy history of tinnitus complaints	Yes	720	24.11
	No	2266	75.89

Table 1: Socio-demographic data assessed with the *Tinnitus Sample Case History Questionnaire (TSCHQ)*

Another aspect worth considering is the number of submissions for the EMA questionnaire. As can be seen in Fig. 2b, 2632 of the 4929 users (53.4%) submitted at least one answer sheet, but only 30.85% of these users have 10 or more submissions (812), with only a few users having 100 or more (163), 500 or more (26), or even 1000 or more (8) submissions. Note that the latter group of 8 users submitted 14.37% of the total EMA answer sheets in the database, i.e., a few users make up a large part of the total data set. For these users, so-called *n-of-1* studies [2] could be conducted to gain in-depth insights on the course of their tinnitus. Furthermore, it can be investigated whether there are differences in the course of the tinnitus between these sub-populations of users. Finally, as adherence and incentives constitute critical topics in EMA and MCS [13, 1, 3], it could be investigated what the difference between sub-populations is, also with regard to other aspects discussed in this paper, and ultimately, what might motivate users to submit more data over a longer period of time.



(a) Number of users by their reported worst symptom (n = 3072).

(b) Number of users with different numbers of EMA submissions. Users with no submissions are omitted in the figure.

Fig. 2: Number of users by (a) their worst symptom and (b) their number of EMA submissions.

3.2. Therapies

Another secondary data set that has yet to be evaluated are *Therapies*. A feature on the website allows users to document treatments that might influence their tinnitus in any way. For each treatment, a *therapy type* (e.g., medical treatment, physical therapy, auditory stimulation), name, time period (i.e., start and end date (optional)) and a personal note can be entered. As this feature was primarily implemented for the user himself, this data is currently not interpreted in any way. The data set contains 143 therapy entries from 84 different users. The ten most frequent therapy types are shown in Fig. 3. The mean therapy duration is 139.17 (s = 335.3). It can be investigated whether there is a correlation between the course of the tinnitus and a specific accompanying therapy or type of therapies.

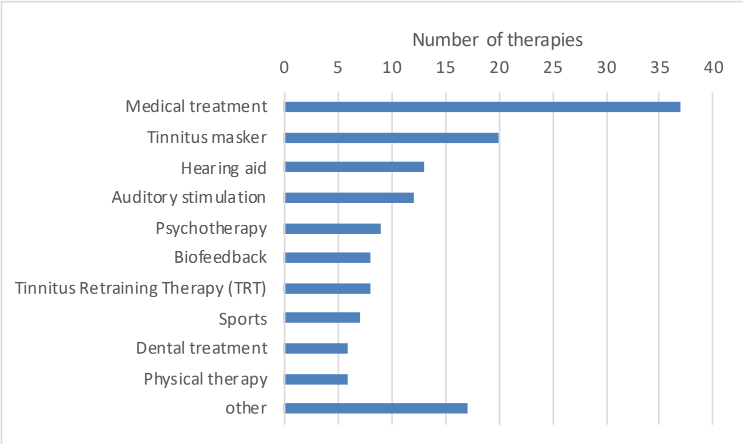


Fig. 3: The ten most frequent therapy types. The remaining therapy types are summarized with 'other' (n = 143).

3.3. *Countries and regions*

For each user, the mobile application stores the country extracted from the device locale on registration. Users that register via the website are asked to select their country from a list. The database contains records for users from 193 different countries. The distribution of the ten countries with the most users are shown in Fig. 4. It can be seen that the largest number of users is from Germany, followed by the United States, the Netherlands and Great Britain. This can be explained by the fact that TYT was initially launched in Germany and has been available in German and English. Since 2017, the mobile application has been available in Dutch as well. It can now be investigated whether there are differences in the perception of tinnitus between countries and regions.

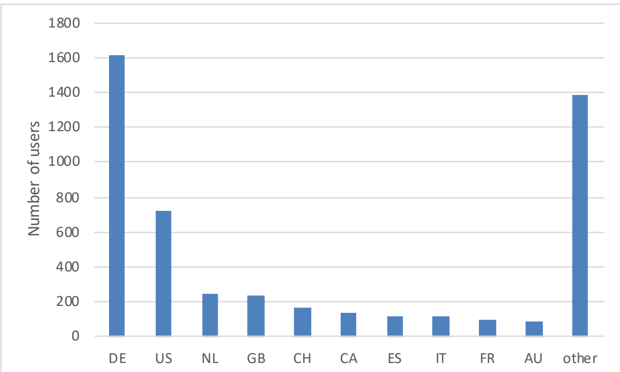
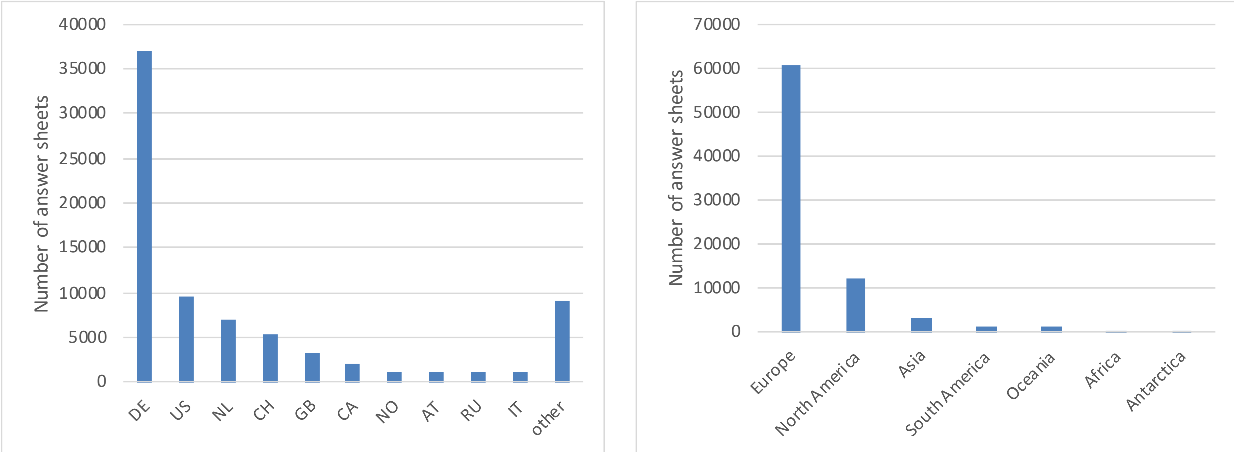


Fig. 4: Number of users by country code. The ten countries with the most users are displayed, users from other countries are grouped under 'other' (n = 4929).

To this end, for each EMA answer sheet, the country of the respective users who submitted the sheet was recorded. As only users who submitted at least one answer sheet are considered, this results in 101 different countries (of the initial 193). The distribution of countries for answer sheets is shown in Fig. 5a. The countries were then grouped by their continent, as shown in Fig. 5b. For the resulting sub-populations, it can be investigated whether there are differences in the course of the tinnitus between countries and continents respectively. Similarly, other implications, e.g., for the hemispheres (northern vs. southern, western vs. eastern), could be investigated.



(a) Countries

(b) Continents

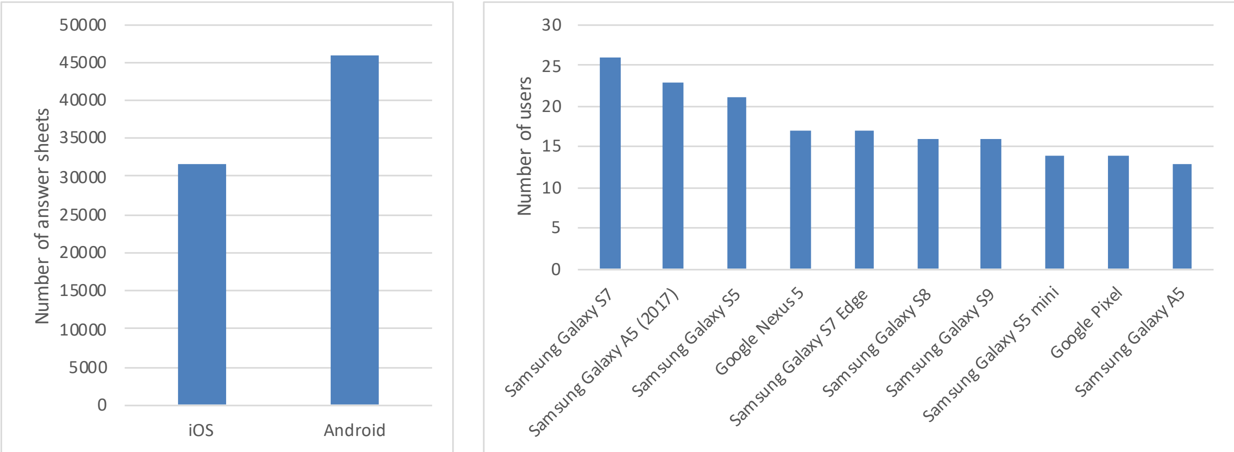
Fig. 5: Number of answer sheets by (a) country and (b) continents (n = 77488). For (a), the ten countries with the most answer sheets are displayed, answer sheets from other countries are grouped under ’other’.

3.4. Smart mobile devices

Another meta data that is automatically extracted and stored for each user is the mobile user agent, which is used for (a) the submission of the initial socio-demographic questionnaires (cf. Section 3.1) and (b) each time an EMA questionnaire is submitted. For the initial questionnaires, 27.67% (1364) of the users use an iOS device, whereas 21.57% (1063) use an Android device. For the remaining 50.76% (2502) of users, there is no information on the mobile operating system, indicating that they did not complete the initial questionnaires (1736) or submitted the questionnaires before introducing this feature (766). As shown in Fig. 6a, 40.78% of EMA answer sheets are submitted by an iOS device, whereas 59.22% are submitted by an Android device. It can be investigated whether there is a difference in the perception and course of tinnitus between users of these operation systems. In addition, for Android, the exact device model is stored. The ten most used Android device models are shown in Fig. 6b. Note that 32 users have changed their device one or more times during the use of the TYT mobile application and are therefore included multiple times in this consideration. For the latter users, it could be further investigated whether the device model directly influences their EMA data.

3.5. Notifications

Notifications are a crucial part of the EMA data collection methodology, since they remind users to fill out the questionnaires and, therefore, help to establish a continuous sampling rate. In order to correlate the triggered system notifications with the questionnaire filled out, the TYT mobile application stores meta data about the notification (i.e., timestamp of the trigger event), meta data about the answer sheet (i.e., timestamp of the moment of storage), and the user’s current notification scheme. In TYT, the latter is either a schedule of fixed points in time (i.e., *fixed*) or the result of an algorithm that produces random points in time (i.e., *random*). The notification scheme can be changed by the user at any time [9]. Currently, the TYT data sets consists of 77488 *EMA Questionnaire Answer Sheets*. With respect to the notification settings, 60100 answer sheets (77.5%) are stored with a random notification setup and 17388 answer sheets (22.5%) with a fixed scheme of points in time. In addition to the number of answer sheets, Fig. 7 illustrates the distribution of answer sheets that can be attributed to triggered notifications and the non-attributable answer sheets. With less than 25%, both of the notification schemes show a rather low amount of attributable answer



(a) Operating systems

(b) Android device models

Fig. 6: (a) Number of answer sheets by operating systems and (b) number of users by the ten most used Android device models (n = 77488).

sheets. Interestingly, the amount of attributable answer sheets is 10% higher for a fixed notification scheme compared to a random notification scheme. The effect of the notification scheme for the notification adherence could be subject of further investigations.

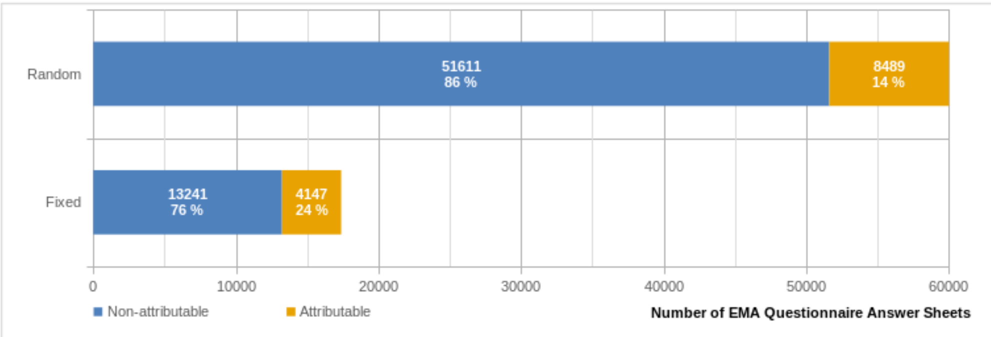


Fig. 7: Numbers of answer sheets that are associated with a fixed or random notification

4. Discussion

Based on the aspects and sub-populations considered in this work, more advanced analyses and evaluations could be conducted in order to gain in-depth insights on TYT, its users and, ultimately, the tinnitus disorder. For instance, *machine learning (ML)* can be applied to the large data set of different features for both *supervised* and *unsupervised learning* (e.g., [7]). Furthermore, advanced statistical analyses, e.g., *multilevel analyses* on the hierarchical EMA data or n-of-1 studies on the large data sets for some of the users (cf. Section 3.1), can be applied to the data set. In addition, as the same (or similar) structure of data can be found in other related projects like *TrackYourHearing (TYH)*, *TrackYourDiabetes (TYD)*, *TrackYourStress (TYS)* or *TinnitusTipps* [3], the aforementioned analyses can be applied to these data sets equivalently.

There are several peculiarities and technical considerations about what data is stored and when it is collected. For instance, when recording the user agent (i.e., the operating system and smart mobile device models) of users, one should take into account that users might change their device in the course of using the application. For this purpose, it is good practice to store the user agent during registration and each time data is collected thereafter. Furthermore, the TYT platform and its data structure has several limitations. For example, in case of notifications (cf. Section 3.5), the assessment of the notification adherence (i.e., how often users answer the EMA questionnaire after receiving a notification) is aggravated by the fact that notification schedules are only stored locally on the devices and, therefore, no information on how many notifications the user has received in total is available during the analysis. Furthermore, it is noteworthy that notifications are only optional reminders and users are able to switch or close the TYT application at any point in time. In both cases, the answer sheet can no longer be attributed to a notification with the consequence that the data set does not contain any notification meta data.

5. Summary & outlook

In this work, we presented the underlying data set and its structure of the TrackYourTinnitus (TYT) platform. We then provided descriptive statistics on various aspects of the data set and highlighted potential insights that could be gained from this data. Finally, we discussed in-depth analyses and evaluations that could be conducted on the data set based on the highlighted aspects as well as technical issues and limitations of this and other similar data sets.

When running an MCS and EMA project like TYT, increased insights can be gained if additionally, socio-demographic and meta data about the users of the platform is stored. This allows identifying various sub-populations and investigate correlations between characteristics of users and their EMA data in order to gain a better understanding of the assessed phenomenon (e.g., a disorder like tinnitus). As an outlook, we intend to conduct further analyses as discussed in this work. For instance, we are currently investigating the environmental sound level measured with the mobile application and how it interacts with other characteristics described in this paper. We hope that providing a comprehensive description of the data set to the community allows to better understand and reproduce the results from our research that are based on the TYT data set or similar projects. Finally, we believe that the insights gained in the scope of this work can serve as a basis for other and future projects and facilitate the comparability between their results.

References

[1] Agrawal, K., Mehdi, M., Reichert, M., Hauck, F., Schlee, W., Probst, T., Pryss, R., 2018. Towards incentive management mechanisms in the context of crowdsensing technologies based on trackyourtinnitus insights. Procedia computer science 134, 145–152.

[2] Duan, N., Kravitz, R.L., Schmid, C.H., 2013. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. Journal of clinical epidemiology 66, S21–S28.

[3] Kraft, R., Schlee, W., Stach, M., Reichert, M., Langguth, B., Baumeister, H., Probst, T., Hannemann, R., Pryss, R., 2020. Combining mobile crowdsensing and ecological momentary assessments in the healthcare domain. Frontiers in Neuroscience 14, 164.

[4] Kubiak, T., Smyth, J.M., 2019. Connecting domains—ecological momentary assessment in a mobile sensing framework, in: Digital Phenotyping and Mobile Sensing. Springer, pp. 201–207.

[5] Langguth, B., Goodey, R., Azevedo, A., Bjorne, A., Cacace, A., Crocetti, A., Del Bo, L., De Ridder, D., Diges, I., Elbert, T., et al., 2007. Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus research initiative meeting, regensburg, july 2006. Progress in brain research 166, 525–536.

[6] Pryss, R., 2019. Mobile crowdsensing in healthcare scenarios: Taxonomy, conceptual pillars, smart mobile crowdsensing services, in: Digital Phenotyping and Mobile Sensing. Springer, pp. 221–234.

[7] Pryss, R., John, D., Reichert, M., Hoppenstedt, B., Schmid, L., Schlee, W., Spiliopoulou, M., Schobel, J., Kraft, R., Schickler, M., et al., 2019. Machine learning findings on geospatial data of users from the trackyourstress mhealth crowdsensing platform, in: 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), IEEE. pp. 350–355.

[8] Pryss, R., Reichert, M., Herrmann, J., Langguth, B., Schlee, W., 2015a. Mobile crowd sensing in clinical and psychological trials–a case study, in: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, IEEE. pp. 23–24.

[9] Pryss, R., Reichert, M., Langguth, B., Schlee, W., 2015b. Mobile Crowd Sensing Services for Tinnitus Assessment, Therapy and Research, in: IEEE 4th International Conference on Mobile Services (MS 2015), IEEE Computer Society Press. pp. 352–359.

[10] Schlee, W., Pryss, R.C., Probst, T., Schobel, J., Bachmeier, A., Reichert, M., Langguth, B., 2016. Measuring the moment-to-moment variability of tinnitus: the trackyourtinnitus smart phone app. Frontiers in aging neuroscience 8, 294.

[11] Stone, A.A., Shiffman, S., 1994. Ecological momentary assessment (ema) in behavorial medicine. Annals of Behavioral Medicine .

[12] Tokosi, T.O., Scholtz, B.M., 2019. A classification framework of mobile health crowdsensing research: A scoping review, in: Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019, ACM. p. 4.

[13] Zhang, X., Yang, Z., Sun, W., Liu, Y., Tang, S., Xing, K., Mao, X., 2015. Incentives for mobile crowd sensing: A survey. IEEE Communications Surveys & Tutorials 18, 54–67.