A collaborative AI dataset creation for speech therapies

Vita Santa Barletta[1], Fabio Cassano[1], Alessandro Pagano[2] and Antonio Piccinno[1]

[1] *Department of Computer Science, University of Bari, Bari, Italy*

[2] *Department of Economics and Finance, University of Bari, Bari, Italy*

Abstract

Artificial Intelligence (AI) and Human-Computer Interaction are getting closer and closer in modern systems, leading to a slow but constant increasing synergy between the two topics. Text prediction and voice recognition are the most know application for AI techniques. Common examples are virtual keyboards, that suggests the next word to be used in the text, systems that recognise people's sentences, voice commands in voice assistants such as Google Home and Alexa, and many others. However, things get more difficult in contexts where a specific recognition, outside the "usual" models, are not the rule, but the exception, as in the case of the recognition of right and wrong phonemes in speech therapies. These difficulties often lie in the lack of the AI model generalisation abilities due to the small datasets used for training. In this position paper, we address this issue and we discuss the role that the culture of participation might have to support the dataset creation for speech therapy. Our aim is to investigate how combining the support of people in the creation of the phonemes samples, and the validation of those elements by the speech therapist, AI models' accuracy can improve.

Keywords

Artificial Intelligence, Speech Therapy, Culture of Participation, End-User Development, Adaptive learning

1. Introduction

In recent years it has been such a spread of techniques and methods of Artificial Intelligence (AI) in many contexts and systems, and today is difficult to identify a clear boundary between "AI systems" and "non-AI systems". AI is today considered a universal solution to solve all problems and exploit all opportunities in digital life. Some researchers think that the main goal of artificial intelligence is to replace human choices and behaviour; others believe that such systems remain speculative and will not be able to reason as a human being [1]. Presently, AI remains an engineering discipline that explores well-defined problems, and the most successful contributions occurred in AI achieving specific objectives, providing the basis for the current hype surrounding artificial intelligence. Differently, human involvement is not a relevant design criterion in this kind of approach [2].

Human-centred artificial intelligence (HCAI) focuses on the development of AI that enhances human performance in ways that make the system reliable, safe, and trustworthy.[3] The evolution of the HCAI allowed easy access to artificial intelligence systems even for end-users and has stimulated the culture of participation. Moreover, it has emerged as a result of the shift from consumer cultures (in which people are confined to passive recipients of artefacts and systems) to cultures in which users are actively involved in the development and evolution of solutions to their problems. Industry researchers identified groups of users at opposite ends of a continuum: professional programmers are excited by computers because they can create computer programs, while domain experts use computers because they allow them to perform their jobs. The goal of supporting domain experts in developing and/or evolving systems does not mean transferring responsibility for a valid system design to the end-user. Yet, if a tool does not satisfy its intended users' requirements, then end-user support should be used in conjunction with expert support to adapt and improve it. In cultures of participation, socio-technical environments should be designed to provide tools that make it easy to participate and continue generating value [4]. To facilitate participation, designers need to be aware of how much effort a user is willing to put into. This will affect how much he or she will contribute [5]. The participant effort can be minimised by providing the right kinds of meta-design tools, while the potential value one can create can be maximised by framing the problem and sharing results with others [6].

There is a design trade-off between AI and Human Computer Interaction (HCI) as they are (historically) on opposite sides of the spectrum of autonomy and control (AI provides for low user control as opposed to HCI) [7]. As a matter of fact, one of the goals of the ongoing world research is to explore the possible synergies that can lead AI to support HCI. Thereby, in recent years, several AI models and techniques have been successfully applied in different domains: for example, the user's support on virtual keyboards, suggesting the best possible "next word" to write in a message. This proves that the HCI world might find a little overlap with AI and that there is potential for deeper interaction between the two areas in more than one context. However, researchers not always agree on this idea. For example in his work, Grudin expressively describes that those two fields have a lot in common, however not yet found a common point [8], on the other hand, Harper et Al. discussed the current role of HCI and AI, explaining how difficult could be to design a human-centered AI [9]. Lissetti et Al. suggest a merge between HCI and social sciences through AI [10], meanwhile the analysis developed by Winograd proposes a way to let HCI and AI interact with each other [11].

Nowadays it is also interesting to see how AI can be applied to concepts such as End User Development (EUD). Myers et Al. propose an approach based on Natural Language Processing to improve the development of spreadsheets and support users in understanding why their program might fail [12]. On the other hand, Buono suggest how formal languages can be used to dynamically manipulate data from heterogeneous sources such as Internet Of Things devices [13]. However, AI is not only related to spreadsheets or electronics-related work-space. For example, it is also applied to multiple scenarios governed by the user's needs, such as smart home systems [14]. A proposal such as [15] describes how EUD can be applied to those kinds of systems, getting closer to the end-user needs to the smart home system management. Other approaches suggest how AI can be applied to improve the user experience with a mouse pointer, or how to improve the user's emotional response [16, 17].

2. The speech therapy context

Speaking well is one of the crucial aspects for a child during childhood. Currently, there are limited technological applications supporting speech therapists and patients (typically children below ten years old) during the therapy: some of them rely on the possibility of the patient to listen and repeat phonemes; others try to "understand" how the patient pronounces the words and return an evaluation score to the speech therapist. Despite this, there are many limitations to this approach.

- On the system side, it is not easy to automatically discriminate between close phonemes, leading to the misevaluation of the actual speech score;
- On the user side, engagement decreases as long as the therapy goes on, leading to an early quit of the therapy.

These critical issues can be resolved through the joint intervention of the disciplines of Human computer interaction and artificial intelligence. Currently, there is no structured and shared dataset for speech therapies capable of allowing a better assessment by the therapist and a better therapy for the patient. This paper proposes a collaborative approach for the creation and validation of a shared dataset to be applied to AI models with the aim to improve both children's speech evaluation and overall speech therapy.

We want to debate the support that AI can give to people in a specific scenario, such as speech therapies. In particular, we will discuss how AI can be exploited in systems used in this domain, and how it can be supported by cultures of participation (of the involved actors), through a collaborative approach. The final goal is to collaborative develop a comprehensive dataset for the most accurate training of AI models that are going to support speech therapies and therapists.

Usually, a dataset is a merge of heterogeneous data produced by, for example, sensors or people. AI models, which base their accuracy on the quality of the data, usually find it difficult to classify the "good" from the "bad"ones, leading to poor training and test performances. The rationale behind our idea is depicted in Figure 1. We can imagine the culture of participation on two different levels: the first one (called "Level 1") is made by users all over the world speaking the same language that, according to a given list of phonemes, are able to record and label them. The process between them is of collaboration, as each user independently record and send a phoneme to a shared database. This is going to be validated by the speech therapists (named "Level 2") who evaluate the recorded phonemes approving or rejecting them. The evaluation criteria might be personal (each speech therapist chose and validate one record) or cooperative, where two or more express their evaluation on the record. To improve the user's engagement in the process, each record is scored following a gamified approach (for example using a five-star rating widely used in gamification). After these steps, the "cooperative" dataset for the speech therapy, can be used to train the AI model, and used in real-world scenarios. One of the most important aspects of the data creation/validation for AI models is the human factor. As we have shown in our model, people are involved in two distinct steps: the creation of the entry and the validation of the labels. The failure of one of those implies a bad dataset creation and an overall failure of the entire system to support the speech therapists. For this reason, we can imagine applying to both the steps through some gamification techniques, which are highly tested to
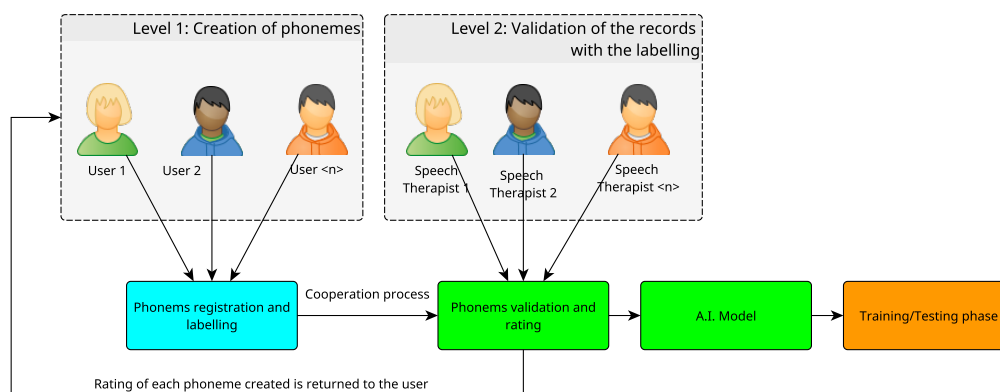
Figure 1: The representation of the collaborative dataset creation: the "Level 1" group represent the collaboration between people to create a phonemes dataset and its labelling. The "Level 2" one represent the collaboration between speech therapist that validate this dataset

be an effective stimulant for people to increment the creation or the validation of the dataset entries [18].

3. Conclusions

In recent years HCI and AI slowly but constantly have proved that they belong to two different faces of the same medal. A closer bond between the two, allowed an increasing number of works to merge the techniques from both HCI and AI.

This paper discussed the possibility to use a collaborative AI dataset creation to support speech therapies. In particular, we introduced how important it is, for modern AI models, to use well-curated datasets. Thus we proposed a model that could support this: by applying the culture of participation, people can support the basic creation of phonemes records; by using the same approach, speech therapists validate all the created entries. With this dataset, many AI models can be trained in order to correctly recognise phonemes during speech therapies, supporting therapists and patients.

The entire so-described approach could be supported by gamification, allowing a constant flow for new phonemes records and their validation. Furthermore: we have to validate the proposed approach and investigate any possible difficulty of the involvement of the people in the entire dataset creation and validation. Lastly, tests with real users and speech therapists/patients must be performed to validate our work.

References

[1] R. Kurzweil, The singularity is near : when humans transcend biology (2005) 652.
[2] E. Brynjolfsson, A. McAfee, The second machine age: Work, progress, and prosperity in a time of brilliant technologies., W W Norton & Co, 2014.

[3] B. Shneiderman, Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems, ACM Transactions on Interactive Intelligent Systems (TiiS) 10 (2020) 1–31.

[4] M. T. Baldassarre, V. S. Barletta, D. Caivano, A. Piccinno, A Visual Tool for Supporting Decision-Making in Privacy Oriented Software Development, Association for Computing Machinery, New York, NY, USA, 2020.

[5] A. Marengo, A. Pagano, L. Ladisa, Mobile gaming experience and co-design for kids: Learn german with mr. hut, Proceedings of the European Conference on e-Learning, ECEL 2016-January (2016) 467 − 475.

[6] G. Fischer, End-user development: Empowering stakeholders with artificial intelligence, meta-design, and cultures of participation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12724 LNCS (2021) 3–16.

[7] G. Fischer, Exploring design trade-offs for quality of life in human-centered design, Interactions 25 (2017) 26–33.

[8] J. Grudin, Ai and hci: Two fields divided by a common focus, Ai Magazine 30 (2009) 48–48.

[9] R. H. Harper, The role of hci in the age of ai, International Journal of Human−Computer Interaction 35 (2019) 1331–1344.

[10] C. L. Lisetti, D. J. Schiano, Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect, Pragmatics & cognition 8 (2000) 185–235.

[11] T. Winograd, Shifting viewpoints: Artificial intelligence and human−computer interaction, Artificial intelligence 170 (2006) 1256–1258.

[12] B. A. Myers, A. J. Ko, C. Scaffidi, S. Oney, Y. Yoon, K. Chang, M. B. Kery, T. J.-J. Li, Making end user development more natural, in: New Perspectives in End-User Development, Springer, 2017, pp. 1–22.

[13] P. Buono, F. Cassano, A. Legretto, A. Piccinno, Eudroid: a formal language specifying the behaviour of iot devices, IET Software 12 (2018) 425–429.

[14] V. S. Barletta, P. Buono, D. Caivano, G. Dimauro, A. Pontrelli, Deriving smart city security from the analysis of their technological levels: a case study, in: 2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS), 2021, pp. 1–6.

[15] P. Buono, F. Balducci, F. Cassano, A. Piccinno, Energyaware: a non-intrusive load monitoring system to improve the domestic energy consumption awareness, in: Proceedings of the 2nd ACM SIGSOFT International Workshop on Ensemble-Based Software Engineering for Modern Computing Platforms, 2019, pp. 1–8.

[16] K. E. Souza, M. C. Seruffo, H. D. De Mello, D. D. S. Souza, M. M. Vellasco, User experience evaluation using mouse tracking and artificial intelligence, IEEE Access 7 (2019) 96506–96515.

[17] C. C. Gomes, S. Preto, Artificial intelligence and interaction design for a positive emotional user experience, in: International Conference on Intelligent Human Systems Integration, Springer, 2018, pp. 62–68.

[18] A. Marengo, A. Pagano, L. Ladisa, Game-based learning in mobile technology, 17th International Conference on Intelligent Games on Simulation, GAME-ON 2016 (2016) 80–84.