Personal Primer Prototype 1: Invitation to Make Your Own Embooked Speech-Based Educational Artifact

Daniel D. Hromada[1], *Hyungjoong Kim*[2]

[1] [2]Berlin University of the Arts / Einstein Center Digital Future, Germany

daniel@udk-berlin.de, hjk@udk-brlin.de

Abstract

In our show & tell contribution, we present more closely seven properties of a digital educational artifact known as Personal Primer. These are: speech-based, embooked, voluminous, modular, circadian, SDG-compliant, edge-computing. Furthermore, we provide link to code repository as well as enumerate off-the-shelf components which can be combined together in order to yield personal Primer prototypes of the first generation. Thus, this show& tell contribution can be understood as an invitation addressing any motivated teacher, student, parent or engineer willing to make an own copy of a Primer for oneself or others.

Index Terms: personal primer, speech recognition, human-computer interaction, make-your-own-device

1. Introduction

This speech&tell contributions is part of an article series [1, 2, 3, 4] describing the materialization of first prototypes of a family of digital educational instruments known as "personal Primer" (PP). Departing from didactic principles of instrumental cognitive enhancement [5] and technical intuitions of [6, 7] PP is an AI-supported "Bildungsinstrument" aiming to harness advantages and possibilities offered by digital technologies for the purpose of education of elementary school pupils, most notably the acquisition of basic literacy (i.e. reading, writing, arithmetic) and promotion of informatic thinking.

2. Implemented properties

As of April 2023, circulating PP prototypes implement following properties motivated by the original road-map [1]

2.1. Speech-based

The Primer provides audio-text support, starting from syllabic level of language acquisition through words and simple sentences all the way to more complex texts like fairy-tales and fables. We currently use DeepSpeech [8] models and our own exercise-specific language models (e.g. "scorers") for domain-constrained speech recognition process.

Since it cannot be expected that our young learners could authenticate themselves with a password, PP also implements ECAPA-TDNN embeddings [9] for speaker verification and identification.

A child can start to speak the word which is given on the screen. Once the child speaks the given text, the audio is sent to locally (e.g. in-class, in-school, in-house) hosted Nvidia Jetson whose speech recognition engines execute necessary inferences in real time and send the result back to PP by using websocket technology. Note that in certain specific, domain-constrained scenarios, it is also possible to execute ASR inferences on a RaspberryPi Zero itself [10].

2.2. Embooked

The term "embooked" means that primer looks like a book shape and shares certain properties with a classical, paper-based book. These are, among other:

1. Primer uses reflected and not emitted light (i.e. e-ink and not LED or OLED) to display content
2. ideally consists of multiple content-displaying sheets which can be "turned around" and which are bound together with specific binding techniques
3. is endowed with both content-encoding interior as well as external front cover, back cover and rücken all potentially covered with another level of protection (e.g. a "dust jacket")
4. has a standardized format (e.g. A5) which allows it to be easily carried around in existing wearables (e.g. school bags) and stored or archived in existing mobiliar (e.g. bookshelves)

2.3. Voluminous

Volume can be useful. As is already clear to mathematicians and engineers, many problems which do not have a solution on a planar surface are easily solvable in 3-D. Theoretical considerations aside, designing educational instruments which are moderately voluminous brings further advantages:

1. using 3D gives us the PP's builders more freedom in dealing with energy-transformation and heat-related issues
2. exterior of a voluminous object offers more than 2 planes on which the content can be displayed
3. a voluminous object can serve as a vessel, i.e. contain / carry things in its interior

Thus, more than an e-book reader with front and back display, PP can display content (e.g. pupil's name) also on the 3rd plane provided by its front or back. And a pupil is invited to use the interior of the DP as a sort of pencil-box and carry in it tools he/she needs during his school practice (pen, paper, ruler, rubber) or plug & play sensors, peripherals or replace components of the Primer itself.

2.4. Modular

The modularity property implies that an idealized PP is to be conceived as a holistic, "more than sum of its parts´´ complex of an extendable set of loosely coupled components. Stated differently, it should be possible - both for the student who creates the PP as well as for the pupil who creates it - to plug and re-plug, position and reposition diverse components in different constellations.

Hence, one can overcome certain limits inherent to technology and protocols used (e.g. I2C for sensor-controller communication) by being able to replace, on-the-fly, one set of sensors with another set of sensors. Additionally, the longevity of the *DP as an entity* could be assured by possibility to replace the broken component with a spare back-up copy of the same component (or its updated variant). s should comply with the policy on pre-prints, which can be found on the conference web site.

2.5. Circadian

Primer is a circadian device [2]. That is, a device with rhythms. For this reason, Primer implements a WittyPi component which allows to create hardware-driven shutdown and booting sequences.

Thus, device can only run in certain period, so that it is not possible for children to use the device whole day.

2.6. SDG-compliant

It is imperative to raise awareness about United Nations' Sustainable Development Goals (SDG) [11] whenever it is possible. To do so, the Primer:

1. consume less amount of energy than alternative solutions (PiZero)
2. are directly chargeable and rechargeable by a solar panel
3. can be and should be used in outdoor environments (e-ink)

It is not to be excluded that future iteration of the Primer will further reduce the battery-related ecological costs by combining the "circadian" and "solar" features in a way that Primer will be active only in presence of solar energy and non-active otherwise

2.7. Edge-computing

All training of speech recognition and voice synthesis data takes place either on the device itself, or on local devices like home- or school- server. The sampe applies for inferencing.

Thus, private child speech data is not shared with big tech cloud and stays where it should: on the edge.

3. Hardware

Main computational unit of PP v1.0 is first version of Raspberry Pi Zero WH. The choice of a relatively Pi Zero v1 is principally motivated by its low price, small form factor (65 mm x 30 mm x 5 mm) and low energy consumption. Being endowed with 512 MB DDR2 ram and 1GHz CPU with ARM11 architecture, Zero's computational capabilities are still highly sufficient for tasks which are to be performed, including visual content generation, adaptive learning as well and potentially even domain-specific automatic speech recognition [10]. Also, being constrained by limits of the Zero at the very beginning, potential hardware upscale in the future will automatically lead to boost in performance.

Asides Pi Zero, DF1.0 integrates functions offered by following off-the-shelf components:

1. two 6 inch (800x600px) e-ink displays with IT8951 controller with ¡1s full refresh rate
2. ReSpeaker dual-microphone expansion board with WM8960 stereo audio codec and Grove I2C ports
3. Grove I2C gesture recognition sensor (PAJ7620U2) for touchless, gesture-based interaction
4. wittyPi3 for clock, voltage & current sensing and implementation of circadian boot and shutdown sequences

Audio output is generated by a 1W 8 Ohm audio transducer which makes vibrate the e-ink display to which it is attached.

The power section consists of a battery, a USB charging circuit and a 5v linear regulator to provide stable voltage. Even when performing complex tasks raspberry Pi Zero consumes around 250mA so a 2500mAh battery is sufficient for at least 10 hours of operation.

4. Code

The core code for the Personal Primer Prototype described in this project is publicly available at `https://github.com/hromi/illustratedprimer`.

For inferences which cannot run on Primer itself `https://github.com/hromi/lesen-mikroserver` project is used to provide multi-scorer, muti-model capabilities, including the possibility to train user-specific models issued from "human-machine peer learning" interactions [4].

Instance of Knowledge Management System providing audiotext contents and other open educational resources used by the Primer is accessible at the address `https://fibel.digital`.

5. References

[1] D. D. Hromada, "After smartphone: Towards a new digital education artefact," *Enfance*, no. 3, pp. 345–356, 2019, presses Universitaires de France.

[2] ——, "Circadian and eutark reduction of the energy trace of a digital school." *Synergie*, no. 7, pp. 74–75, universität Hamburg.

[3] D. D. Hromada, P. Seidler, and N. Kapanadze, "Bauanleitung einer digitalen fibel von und für ihre schüler," *Mobil mit Informatik*, vol. 9, p. 37, 2020.

[4] D. D. D. Hromada and H. Kim, "Proof-of-concept of feasibility of human-machine peer learning for german noun vocabulary learning," in *Frontiers in Education*, vol. 8. Frontiers, p. 48.

[5] R. Feuerstein, R. S. Feuerstein, L. Falik, and Y. Rand, "Part ii, practical applications of the instrumental enrichment program." in *Creating and enhancing cognitive modifiability: The Feuerstein Instrumental Enrichment Program*. ICELP Publications, 2006.

[6] A. C. Kay, "A personal computer for children of all ages," in *Proceedings of the ACM annual conference-Volume 1*, 2011.

[7] N. Stephenson, *The diamond age*. Penguin UK, 1998.

[8] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates *et al.*, "Deep speech: Scaling up end-to-end speech recognition," *arXiv preprint arXiv:1412.5567*, 2014.

[9] B. Desplanques, J. Thienpondt, and K. Demuynck, "Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification," *arXiv preprint arXiv:2005.07143*, 2020.

[10] F. Brodbeck, P. Seidler, and D. Hromada, "Power consumption of diverse speech command classification methods on the raspberry pi zero," 2021.

[11] D. o. E. United Nations and S. A. S. Development, "Transforming our world: the 2030 agenda for sustainable development," p. 16301, 2015. [Online]. Available: https://sdgs.un.org/2030agenda