Paje: Deployment of an Speech Recognition Based Framework for Community Networks

Emerson Oliveira, Aldebaro Klautau and Francisco Muller

***Abstract*— Rural areas in Brazil, especially in the Amazon region, suffer from the lack of provision of basic services, such as health, communications and transportation. Community communication networks can play an important role on these regions by providing a low-cost communication solution by deploying mobile telephony and Internet on remote rural areas. The present work presents an application which aims to assist professionals in the process of gathering data using an voice-based system which calls users inside the community network. The proposed solution provides a software framework which uses open communication and speech recognition software deployed locally, bypassing the need of continuous Internet access and assuring that sensitive data is not used unfairly. Preliminary laboratory test results of the proposed software are presented.**

***Keywords*— Community Networks, Automatic Speech Recognition, E-health, Open-Software.**

I. Introduction

The lack of basic infrastructure, specially in rural areas, affects the everyday life of millions of people in Brazil. For instance, [1] shows the poor provision of telecommunication services in Brazil, where most of the digital exclusion occurs on remote regions. To diminish the digital gap, effort is being made by deploying low-cost community mobile networks, such as the CELCOM project [2]. This network provides mobile telephony and low speed internet connectivity using second generation GSM (Global System for Mobile Communications) technology that can be accessed by community members using regular mobile cellphones, not requiring a smartphone.

Providing connectivity may not be enough. Many people in these communities, especially elder people, may have very low digital literacy. This suggests the use of a more active approach to connect people in these communities and help them access basic services. One approach is the use of speech recognition technologies as an interface for various services in a community network. In order to work effectively, the proposed solution should not rely on reliable internet connection, which is not always available on these rural communities.

One aspect that needs improvement in most of rural communities is access to health services. In Brazil, [3] describes how SUS (Sistema Único de Saúde) lacks enough funding and resources, which impacts Brazil's population entirely, and the Amazon region in particular. These vulnerability issues have endorsed several E-Health projects around the globe. [4] intends to harness ICTs (Information Communication Technologies) to significantly improve the delivery of healthcare services in public primary and secondary healthcare facilities across their country. In [5], Cystic Fibrosis patients and caregivers to manage the disease with an innovative app.

This paper presents a solution called Paje, a software framework developed to provide an audio interface using ASR (automatic speech recognition) for community mobile networks. The Paje framework considers that these networks may have limited access to the internet. ASR can be more accurate when executed on the server (such as Google's ASR service), but executing on the client's device, as suggested in this work, is advantageous when the connectivity is limited. The authors also implemented an e-health application for health professionals and doctors to assist residents from remote areas. Section II describes the framework architecture, detailing functionalities, software components and usage. Section III outlines the preliminary results achieved on practical tests and simulations. Finally, Section IV concludes the work.

II. Framework Architecture

Figure 1 defines the overall architecture of the application, showing each of its most important components on a clear-visual way. The yellow hexagons presents components which need to request any kind of external party. The blue ones, on the other hand, categorizes the self-contained application components, all of its functionalities run locally.

Essentially, Paje comprises a speech recognition engine, Coruja [6], including an acoustic model trained and specially developed for Brazilian Portuguese; a Python [7] developed API with Flask framework using as back-end database SQLite and a script API caller that queries Oddcast API [8] for voice recording on-the-fly. It also makes use of Google Chart API [9] for generation of resulting graphics. Lastly, two modules were developed for VoIP applications, one for Asterisk use-case and the other one for Freeswitch. Paje is designed to be integrated with CELCOM pilots. The gateway connection for both applications depends on the VoIP software used.

A proposed use case for Paje is an e-health application. The application's workflow is as follows: health professionals register on the application, via the web front-end, being able to get started and create their forms. A form is a set of questions regarding health conditions, symptoms and whatever may the professionals wish to know about community's health context. Then, the user can select one form and launch an automated telephony survey.

On a survey execution, all the online mobile stations connected to the CELCOM pilot will be dialed, and as soon as they answer, questions from the form will be synthesized and played to the receiver. Users will listen to and then answer the questions. This activates the automatic speech recognition

LASSE - Grupo de Pesquisa em Engenharia Humanitária, Universidade Federal do Pará, Belém-PA, E-mails: {emersonjr, aldebaro, fmuller}@ufpa.br.

engine and starts the recognition process. Each pronounced answer will be recognized and store at the database. Some questions may have predefined answers, such as yes or no, which are recognized on-the-fly. If after three attempts there is no recognized answer, the audio is recorded for further manual processing at a later time. Other, the answer may be free form and the audio is saved for later use of the health professionals.

After the survey ends, all the generated data is gathered. Paje's interface shows plots for the gathered data, allowing for easy visualization, as illustrated in Figure 2.

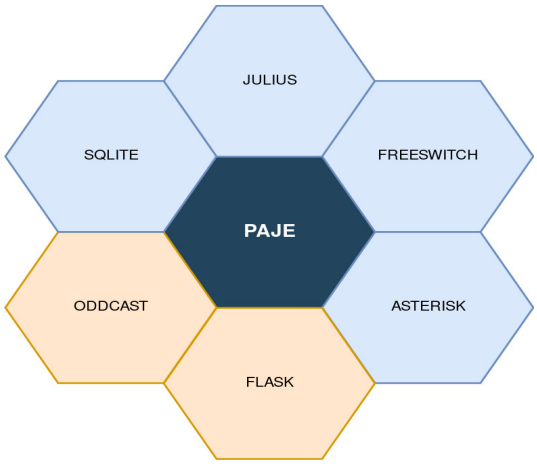

Fig. 1. A simple scheme displaying Paje's software components.

III. Preliminary Results

In order to assess the software's behavior, a test was carried out in the laboratory. The test simulates practical usage conditions and was performed with 15 voluntary students, using trivial questions. The students have subscribed to a private test network with their personal mobile phones. No one was informed previously about the questions or how the to interact with the system. An instance of Paje was running connected to that network, using a module developed for Freeswitch. A form composed by five questions was created, three with predefined answers and two with free-form answers. Every participant was able to answer the calls and answer the questions. In this preliminary test, the software worked as expected. Figure 2 shows the results for one of the questions with predefined answers obtained using ASR.

One of the detected issues during the test is the quality of the audio during calls being too low. This is mostly due to the way Freeswitch handles audio, as it converts incoming audio before sending it through the VoIP network. As the system was developed to repeat an given question three times in case of no recognizable answer, the issue was circumvented in this test and it needs to be solved before actual use in the communities.

Another issue was that the Oddcast synthesizer sometimes mistakenly pronounces some tonic syllables for Portuguese. Besides some minor intelligibility issues, the engine worked well during the test.

IV. Conclusion

This paper describes Paje, a framework that helps gathering information from community users using automatic voice calls

Fig. 2. A chart showing the results after speech recognition results for one of the survey questions in the laboratory test.

inside the community network, resorting to open software. The data is processed locally, avoiding the need of continuous Internet access and the risk of processing sensitive data using third-party services. Preliminary tests inside our laboratory showed good recognition results and demonstrate that the software works well in a practical situation, with the same hardware and software used in the CELCOM pilots. Future work includes improvements in the audio quality and usage tests in the communities supported by the CELCOM project.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), process number 2015/24508-9.

References

[1] M. Dias, P. Batista, and A. Klautau, "Estudo tecno-econômico parconectar amazônia usando telefonia comunitária," *WTICp/D*, p. 6, May 2017.

[2] E. Junior, M. Dias, B. Boas, A. Klautau, "CELCOM: Tecnologias e Experiências Adquiridas em Redes de Telefonia Comunitárias no Brasil" *Anais XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. p. 929, September 2018.

[3] M. Rossi, "Saúde pública no Brasil ainda sofre com recursos insuficientes", *Agência Câmara de Notícias - Câmara dos Deputados - Palácio do Congresso Nacional - Praça dos Três Poderes - Brasília*, Report, January 2015.

[4] "E-Health Project - Universal Service Provision Fund (USPF)", http://www.uspf.gov.ng/index.php?option=com_content&view=article&id=14:e-health-project&catid=14&Itemid=128, accessed in May 2019.

[5] "MyCyFapp - A Project for patients", https://www.mycyfapp.eu/index.php/en/what-is-mycyfapp/objectives, accessed in April 2019.

[6] BATISTA, Pedro dos Santos, "Avanços em Reconhecimento de Fala para Português Brasileiro e Aplicações: Ditado no LibreOffice e Unidade de Resposta Audível com Asterisk" *Dissertações em Engenharia Elétrica (Mestrado) - PPGEE/ITEC*, March 2013.

[7] Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org

[8] Oddcast Inc. "TTS Demo" https://www.oddcast.com/technologies/tts/, accessed in April 2019.

[9] Y. Zhu, "Introducing Google Chart Tools and Google Maps API in Data Visualization Courses", *IEEE Computer Graphics and Applications*, Vol. 32, Issue 6, p. 6-9, November 2012.