International Journal of Natural Computing Research
Volume 10 • Issue 1 • January-March 2021

Estimation of ASR Parameterization for Interactive System

Mohamed Hamidi, LISAC, Department of Mathematics and Computer Science FSDM, Sidi Mohamed Ben Abdellah University, Morocco

Hassan Satori, LISAC, Department of Mathematics and Computer Science FSDM, Sidi Mohamed Ben Abdellah University, Morocco

Ouissam Zealouk, LISAC, Department of Mathematics and Computer Science FSDM, Sidi Mohamed Ben Abdellah University, Morocco

Naouar Laaidi, LISAC, Department of Mathematics and Computer Science FSDM, Sidi Mohamed Ben Abdellah University, Morocco

ABSTRACT

In this study, the authors explore the integration of speaker-independent automatic Amazigh speech recognition technology into interactive applications to extract data remotely from a distance database. Based on the combined interactive voice response (IVR) and automatic speech recognition (ASR) technologies, the authors built an interactive speech system to allow users to interact with the interactive system through voice commands. The hidden Markov models (HMMs), Gaussian mixture models (GMMs), and Mel frequency spectral coefficients (MFCCs) are used to develop a speech system based on the ten first Amazigh digits and six Amazigh words. The best-obtained performance is 89.64% by using 3 HMMs and 16 GMMs.

KEYWORDS

Amazigh, Asterisk, Database Server, GMM, HMM, IVR, Speech Recognition, Sphinx, VOIP

INTRODUCTION

Interactive Voice Response (IVR) is a telephony technology that allows the interaction with users, collects information and routes. Also, the IVR system permits callers to extract or enter data from a database using the voice in real-time and it accepts a combination of voice telephone input and keypad and affords the responses in the form of voice, fax, or email, etc. This approach allows an efficient exchange of information with reducing costs (Van Meggelen et al., 2019).

Automatic Speech Recognition (ASR) is a computer technology that allows a user to communicate orally to a machine by extracting the oral message contained in the speech signal. This technology uses a wide variety of tools: signal processing, powerful statistical mathematical models, classification of forms, algorithmics and artificial intelligence (Haton et al., 2006: Mohdiwale, et al., 2020; Taquee et al., 2021; Srivastava et al., 2020). There are many applications that we can imagine such as voice control of machines, booking flights, learning other languages, helping people with disabilities, voice services in mobile phones, automatic transcription, indexing of multimedia documents and machine dialogue. etc. (Huang et al., 1990).

The authors in (H. Aust et al., 1995) have implemented a telephony operator system that allows callers to retrieve train traffic information by speaking fluently with the system. Satori et al. (H. Satori

DOI: 10.4018/IJNCR.2021010103

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

et al., 2014) have designed an ASR system by using CMU Sphinx tools which is based on the Hidden Markov Models (HMMs). They have aimed to create of an Amazigh automatic speech recognition system that includes digits and alphabets where their achieved performance was was 92.89%. T. K. DAS et al. (2016) have implemented a speaker and speech recognition system that permits users to show visual information. Their created system was realized with the help of HMM technique. In their work, the Mel-Frequency Cepstral Coefficients (MFCC) technique was used to produce the speech feature vectors. Their obtained results show that the achieved accuracy is about 90%. The authors in (Alsulaiman, et al., 2017) have examined the effect of phonemes on the performances of the Speaker Recognition (SR) system where the impact of some Arabic phonemes on the recognition rate is observed. As a result, the vowel recognition rates were all above 80%, the consonants recognition rates varied between 14% and 94%.

K. Shah et al. (Shah et al., 2012) have studied the use of the asterisk server in the VoIP network. Their proposed system was created by using VMware and it has been configured with various security settings such as VPN server, firewall IP table rules, intrusion detection and intrusion prevention system. The authors in (Basu et al., 2013) have described the real-time challenges of designing telephonic Automatic Speech Recognition System. In their study, they have used the Asterisk server to design a system that poses some queries and the spoken responses of users are stored and transcribed manually for ASR system training. In this work, the speech data are collected from West Bengal. (C. Bhat et al. 2013) have created a Speech Enabled Railway Enquiry System (SERES) that is a system that permits users to get the railway information considering the Indian scenario, based on the IVR solution. In (Hamidi et al., 2016a; Hamidi et al., 2016b) the authors present their first experiment to integrate the ten first digits of Amazigh language in an Interactive Voice Response (IVR) server where the users use speech (ten first Amazigh digits) to interact with the system. Sehgal et al. (2018) have explained the exploitation of sentiment analysis to identify customers satisfaction based on interactive voice response and automatic speech recognition systems. Also, they have presented their approaches and techniques that are used to recognize the user emotions in the call center via a sentiment analysis system. In (Gravano et al., 2011) researchers they provided new information on the mechanisms in the conversation of human-human for signaling the turn end and situations identification in which a backchannel is appropriate. In their work are determined some situations which are useful for IVR systems designers such as the system output production and the user input recognition. They have examined potential turn-taking signals attached to human-human turn exchanges which are automatically computable. In (Hamidi et al., 2020) authors have presented the performances of an Amazigh ASR system via an IVR server in noisy conditions. Their implementations were conducted for the uncoded speech and decoded speech under the train noisy environment for different signal noise ratios (SNR). Their findings show that the most affected digits are those which include the "S" consonant that drops rapidly at 0% in 30 dB and 27 dB for uncoded and decoded speech respectively.

In this paper, we exploit the Markovian modeling approach and Gaussian Mixture Models to design a speech interactive response system. This created system allows callers to extract data from a remote database using voice commands based on the Moroccan Amazigh language. To the best of our knowledge, this is the first IVR study using less-resourced language such as Amazigh language.

The rest of this paper is organized as follows: Section 2 shows the VoIP technology. Section 3 presents the telephony server implementation. In section 4 we will present the Amazigh automatic speech recognition system. Finally, section 5 gives a view of our system implementation. The authors finished by a conclusion

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/estimation-of-asr-parameterization-for-interactive-system/274936

Related Content

Data Highways: An Activator–Inhibitor–Based Approach for Autonomic Data Dissemination in Ad Hoc Wireless Networks1
Karina Mabell Gomez, Daniele Miorandiand David Lowe (2012). *Biologically Inspired Networking and Sensing: Algorithms and Architectures (pp. 223-241).*
www.irma-international.org/chapter/data-highways-activator-inhibitor-based/58309

A Comprehensive Analysis on Biodiesel Blend Model
Yarrapragada K.S.S. Raoand Bala Krishna B. (2018). *International Journal of Artificial Life Research (pp. 25-46).*
www.irma-international.org/article/a-comprehensive-analysis-on-biodiesel-blend-model/227072

A Study of Predicting Ability in State-Action Pair Prediction: Adaptability to an Almost-Periodic Disturbance
Masashi Sugimoto, Naoya Iwamoto, Robert W. Johnston, Keizo Kanazawa, Yukinori Misakiand Kentarou Kurashige (2017). *International Journal of Artificial Life Research (pp. 52-66).*
www.irma-international.org/article/a-study-of-predicting-ability-in-state-action-pair-prediction/182578

Lifelike Self-Replicators
Eleonora Bilottaand Pietro Pantano (2010). *Cellular Automata and Complex Systems: Methods for Modeling Biological Phenomena (pp. 210-247).*
www.irma-international.org/chapter/lifelike-self-replicators/43222

Challenges and Future Prospects of Intelligent Decision Making
Heena Irfan Shaikhand Imran Baig Mirza (2024). *Bio-Inspired Intelligence for Smart Decision-Making (pp. 179-188).*
www.irma-international.org/chapter/challenges-and-future-prospects-of-intelligent-decision-making/347321