Int. j. inf. tecnol. (October 2023) 15(7):3533–3540
https://doi.org/10.1007/s41870-023-01354-z

ORIGINAL RESEARCH

Amazigh speech recognition based on the Kaldi ASR toolkit

Fatima Barkani[1] · Mohamed Hamidi[2] · Naouar Laaidi[1] · Ouissam Zealouk[1] · Hassan Satori[1] · Khalid Satori[1]

Received: 11 January 2023 / Accepted: 11 June 2023 / Published online: 22 June 2023

Abstract In this work, we offer a new approach to integrating the Amazigh language, which is a less-resourced language, into an isolated speech recognition system by exploiting the Kaldi open-source platform. Our designed system is able to recognize the ten first Amazigh digits and ten daily must-used Amazigh isolated words, which present typical syllabic structure and which are considered a good representative sample of the Amazigh language. The designed speech system was implemented using Hidden Markov Models (HMMs) with different number of Gaussian distributions. In addition, we evaluated our created system performance by varying the feature extraction methods in order to determine the optimal method for maximum performance. The best-obtained result is 93.96% was obtained with Mel Frequency Cepstral Coefficients (MFCCs) technique.

Keywords Kaldi · Speech recognition · HMM · GMM · MFCC · PLP · FBANK · CMU Sphinx4 · Amazigh language

1 Introduction

Automatic speech recognition (ASR) is a technique that allows for transcribing an oral message and extracting linguistic information from an audio signal. ASR is used in different domains such as teaching, interactive services, messaging, machine or robot control, quality control, data entry, remote access, system detection [1–5], etc. Also, several systems have been developed for voice recognition, like Hidden Markov Model Toolbox (HTK) [6], Institute for Signal and Information Processing (ISIP) [7], CMU Sphinx [8–10], and Kaldi [11].

The researchers in [12] have described the development of the Kannada speech recognition system using the Kaldi toolkit. Medennikov et al. [13] talk about the implementation of a Russian speech recognition system using the Kaldi toolkit. The authors [14] have presented a technical overview of the speech recognition systems based on Moroccan dialects. They talk about their recent progress pertaining to feature extraction methods, performance evaluation, and speech classifiers. The authors in [15] have created a Darija speech recognition system based on the CMU Sphinx tools with Hidden Markov Model (HMMs) and Gaussian Mixture Models (GMMs) combination. Their highest accuracy was 96.27%, which was found using 8 GMMs. Ameen et al. [16] have exploited several models for Arabic phonemes recognition. They used different machine learning schemes, neural network, recurrent neural network, artificial neural network (ANN), random forest, extreme gradient boosting (XGBoost), and long short-term memory. The obtained results indicate that the ANN machine learning method outperformed other methods. Table 1 presents some ASR systems studies based on the Amazigh language [17].

In this study, we propose a new method for the integration of the Amazigh language by using the open-source Kaldi based on an isolated variant vocabulary speech recognition system. We propose an open-source platform to evaluate our ASR performance by varying HMMs, Gaussian mixture models (GMMs), and feature extraction techniques, in order to determine the optimal values for maximum performance.

✉ Mohamed Hamidi
mohamed.hamidi.5@gmail.com

[1] Laboratory of Computer Science, Signals, Automation and Cognitivism, Faculty of Sciences Dhar Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco

[2] Team of modeling and scientific computing, Pluridisciplinary Faculty of Nador, Mohammed First University, Oujda, Morocco

Table 1 ASR systems studies based on the Amazigh language

Description	Results
Creating a system for Amazigh speech recognizing based on HMMs-GMMs combination [18]	The best performance of the system was 92.89%
Using HMMs, GMMs, and MFCCs to create a remote recognition system [19, 20]	The best system recognition rate was 92.22%
Use the Raspberry Pi board based on the CMUSphinx4 to construct an intelligent system [21]	90.43%
Implementing smokers detection system based on ASR algorithms [22]	90.13%
Use MFCCs, GMMs, and HMMs to investigate the effect of speech coding on the ASR task [23]	The best performances were found for the G711 codec, 3 HMM, and 16 GMMs
Creating an interactive speaker-independent ASR system [24]	The best system performance was 89.64%
Exploiting CMU Sphinx-4 with HMMs and GMMs to develop an Amazigh automatic speech recognition system [25]	90%
Developing an Amazigh speech recognition system based on CNN and GPU computation using TensorFlow [26]	93.9%
Building an Amazigh automated speech recognition system using the open-source CMU Sphinx-4 [27]	88%

Our paper is organized as follows: an introduction in Secti 1. Section 2 gives a description of the Kaldi toolkit. Section 3 represents the Hidden Markov Models (HMMs). Section 4 gives an overview of Feature extraction methods. The proposed system architecture is detailed in Sect. 5. Experimental results are presented in Sect. 6. Finally, the conclusion is in Sect. 7.

2 Kaldi toolkit

Kaldi is considered an open-source project written in C + + and released under the Apache License v2.0 for speech recognition [11]. Kaldi includes a large set of tools and programs such as HMMs, decision trees, neural networks, and data preprocessing, feature extraction. Their internal structure is shown in Fig. 1 [11]. The modules of the Kaldi library depend on two external libraries, the linear algebra libraries (BLAS / LAPACK) and the library which allows the integration of finite state transducers (OpenFST). The decodable class bridges these two external libraries. This ASR toolkit is still constantly updated and further developed by a pretty large community.

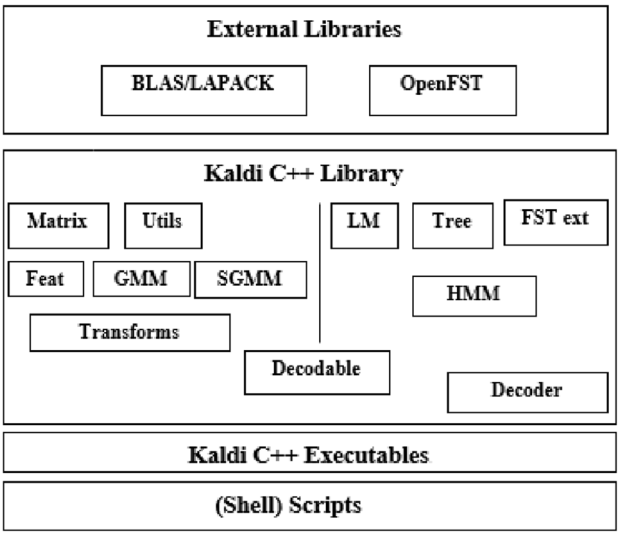

Fig. 1 Kaldi toolkit [5]

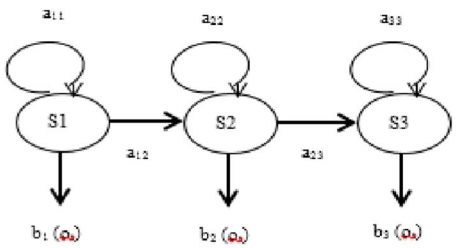

Fig. 2 The 3 states HMM architecture

3 Hidden Markov models (HMMs)

The HMM was introduced in the 1960s [28]. It was considered as one of the most used methods for speech recognition modeling [29]. Also, it used for computational molecular biology [30]. Figure 2 presents a three states Hidden Markov Model topology [31].

4 Feature extraction methods

The feature extraction phase plays a crucial role in the performance of ASR systems. It allows to extract of characteristics that make it possible to discern the components of

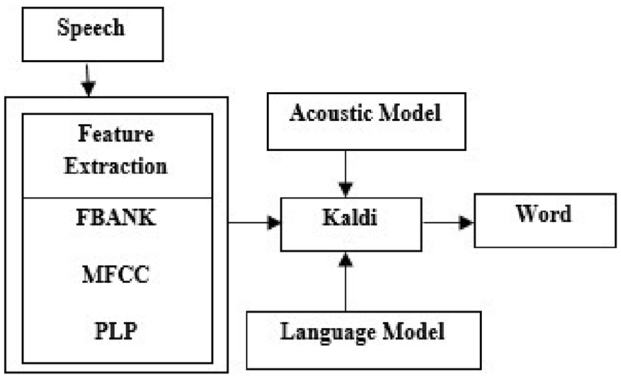

Fig. 3 Proposed system architecture

the audio signal that are relevant for the identification of linguistic content, by rejecting the other information contained in that signal. In this work, the used feature extraction methods are:

- MFCCs are widely employed in speech recognition [32]. The Mel's for a particular frequency is computed according to the formula (1) [33]:

$$Mel(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right) \quad (1)$$

- Perceptual Linear Prediction (PLP) [34].
- Filter Bank Coefficients (FBANK) [35].

5 The proposed system architecture

Repeated In this research, we propose a speech platform for the integration of the Amazigh language into an isolated variant vocabulary speech recognition system. This system is based on the open-source Kaldi using HMMs approach, and a different number of GMMs. In addition, the MFCCs, FBANK, and PLP feature extraction techniques are used. The proposed System Architecture is presented in Fig. 3.

5.1 Corpus

The speech data consists of the 10 first Amazigh digits (0–9) (Table 2 presents the used Amazigh digits) and ten Amazigh words (Table 3 presents the used Amazigh words). This corpus is collected from 30 Moroccan native Tarifit speakers. The speakers are invited to pronounce each Amazigh word ten times. Each digit and word was recorded and visualized back to ensure the inclusion of the entire word in the speech signal where only the corrected words were kept in the database. The speech is recorded with the help of a microphone by recording tool WaveSurfer with wave format and it was saved into one ".wav" file. 16 kHz sampling rate with a resolution of 16 bits was used. Also, two disjoint sets of audio files one for training and the other for testing were created in this work.

Table 2 The Amazigh digits

Amazigh digits	English equivalent	Tifinagh transcription	syllables	Number of syllables
AMYA	Zero	ⵄⵎⵢⴰ	VCCV	2
YEN	One	ⵢⴻⵏ	CVC	1
SIN	Two	ⵙⵉⵏ	CVC	1
KRAD	Three	ⴽⵔⴰⴹ	VCCVC	2
KUZ	Four	ⴽⵓⵣ	CVC	1
SEMUS	Five	ⵙⴻⵎⵓⵙ	CCCVC	2
SEDISS	Six	ⵙⴻⴹⵉⵙ	CCVC	1
SA	Seven	ⵙⴰ	CV	1
TAM	Eight	ⵜⴰⵎ	CVC	1
TZA	Nine	ⵜⵣⴰ	CCCV	1

Table 3 The used Amazigh words

Amazigh Words	English equivalent	Tifinagh transcription	syllables	No. of syllables
AFLLA	Above	ⴰⴼⵍⵍⴰ	VCCCV	2
AFOSI	Right	ⴼⵓⵙⵉ	VCVCV	3
ALNDAD	In front of	ⴰⵍⵏⴷⴰⴷ	VCCCVC	2
AMAGGWAJ	Far	ⴰⵎⴰⴳⴳⵡⴰⵊ	VCVCCVC	3
ANAKMAR	Near	ⴰⵏⴰⴽⵎⴰⵔ	VCVCCVC	3
AWAR	After	ⴰⵡⴰⵔ	VCVC	2
AZLMAD	Left	ⴰⵣⵍⵎⴰⴷ	VCCCVC	2
DAR	Beside	ⴷⴰⵔ	CVC	1
DAT	Before	ⴷⴰⵜ	CVC	1
DDAW	Down	ⴷⴷⴰⵡ	CCVC	1

Table 4 Sample of the used dictionary

AFLLA	AE F L AH
AFOSI	AE F AH S IY
AMYA	A M Y A
AWAR	AH W AO R
AZLMAD	AE Z L M AH D
DAR	D AA R
DAT	D AE T

5.2 Acoustic model

The acoustic model allows the recognition of the phonemes sequences presented in the pronunciation dictionary. 3-State of HMM with a simple monophonic model trained are used for recognizing the speech data. A sample of the used Amazigh dictionary is presented in Table 4.

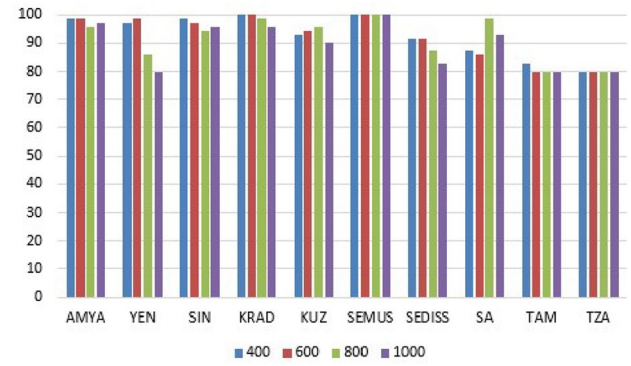

Fig. 4 The recognition rate of Amazigh digits in the function of GMMs by using MFCCs

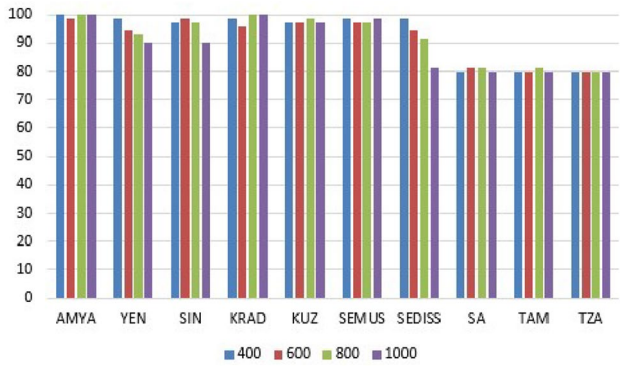

Fig. 5 The recognition rate of Amazigh digits in the function of GMMs number by using PLP

5.3 Decoder

The decoder combines the predictions of the acoustic and linguistic models to propose the most probable transcription in text for a given speech.

In this paper, we focused on the integration of the Amazigh language into an isolated variant vocabulary speech recognition system based on Kaldi with the use of GMM-HMM.

6 Experimental results

In this section, we performed two experiments. In the first, the system was trained and tested with the first ten Amazigh digits (0–9). In the second experiment, the system was trained and tested by the ten daily must-used Amazigh isolated words that present typical syllabic structure and are considered as good representative samples of the Amazigh language. All tests were performed on an Ubuntu 16.04 LTS (64-bit operating system). In our experiments, the data speech is divided into 70% for training and 30% for testing (see Table 4). Different sets of training and testing parameters were used to design an efficient detection system. We have trained and tested the system by using different GMM values, and the MFCC, FBANK, and PLP feature extraction techniques.

Figures 7, 8, and 9 present the results of the first experiment where the system was trained and tested using the digits for MFCC, FBANK, and PLP Coefficients and the GMM values ranging from 400 to 1000.

From Fig. 4 we can read that the most frequently recognized Amazigh digits using MFCCs are KRAD and SEMUS. While in the case of using the PLP coefficient the best frequently recognized Amazigh digits are AMYA and KRAD (see Fig. 5). In addition, in the case of FBANK

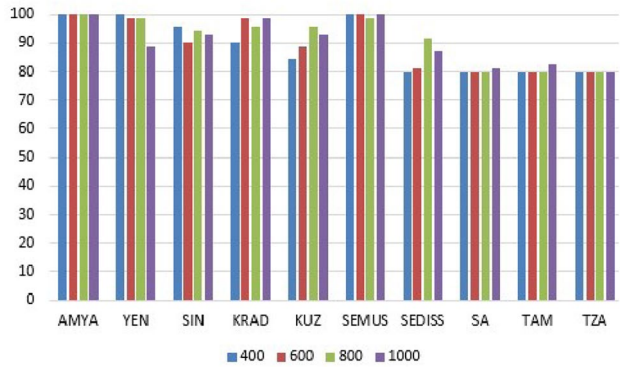

Fig. 6 The recognition rate of Amazigh digits in the function of GMMs by using FBANK

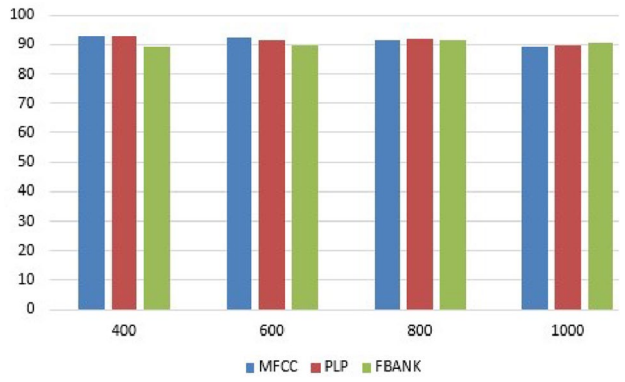

Fig. 7 The recognition rate difference between MFCC, PLP, and FBANK in the function of GMMs for Amazigh digits

coefficients, the most frequently recognized Amazigh digits are AMYA and SEMUS (See Fig. 6).

The system performance of MFCC, PLP, and FBANK extraction feature methods with several GMMS values is

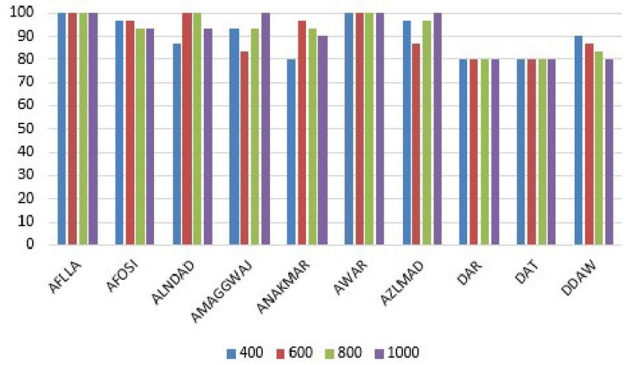

Fig. 8 The recognition rate of Amazigh words in the function of GMMs by using MFCC

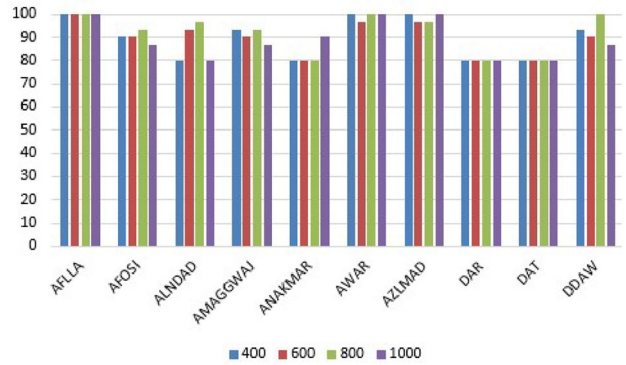

Fig. 9 The recognition rate of Amazigh words in the function of GMMs by using PLP

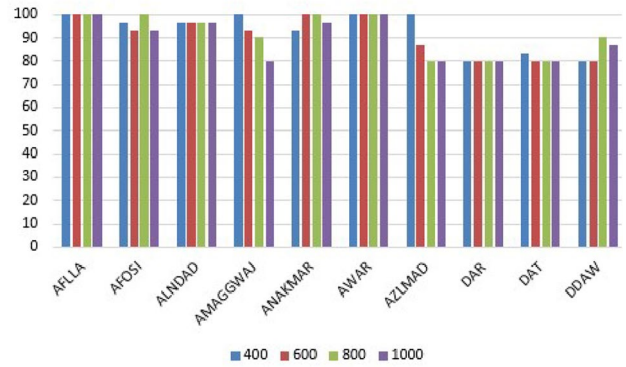

Fig. 10 The recognition rate of Amazigh words in the function of GMMs by using FBANK

shown in Fig. 7. The best results were obtained with 400 GMMs for MFCCs and PLP.

Figures 8, 9, and 10. shown the results of the second experiment. A higher recognition rates was attained particularly especially with the words AFLLA and AWAR for

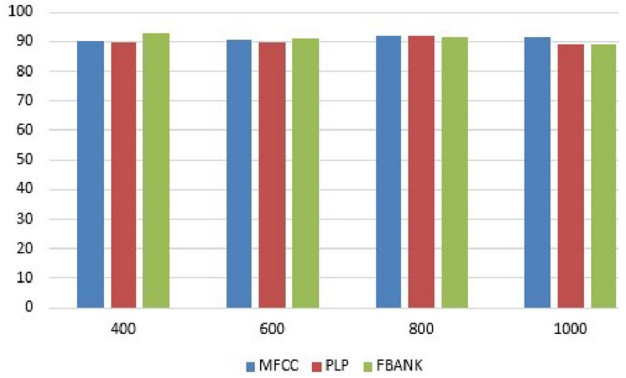

Fig. 11 The recognition rate difference between MFCC, PLP, and FBANK in the function of GMMs for Amazigh words

using MFCC, PLP, and FBANK with the case of 400, 600, 800, and 1000 GMMs.

Figure 11 shows the recognition rate difference between MFCC, PLP, and FBANK in the function of total Gaussian distributions for Amazigh words. The system obtains the best performance when trained by using FBANK with 400 GMMs.

The achieved results in the exeprience1 and exeprience2 show:

- The best result was found with 400 GMMs.
- The FBANK coefficient performance was better for Amazigh words, also, it is noted that the MFCC coefficient performance was better for Amazigh digits.

By considering the Amazigh words digits analysis, all words and digits that consist of two or three syllables achieve a higher rate. As examples:

- The “AFFLA” word its recognition rate is 100%, found for 400, 600, 800, and 1000 GMMs by using MFCC, PLP, and FBANK coefficients, with its number of syllables is 2.
- The “AFOSI” word its number of syllables is 3, and its best recognition rate is 100% found with FBANK coefficient by using 800 GMMs.
- The “ALNDAD” word its number of syllables is 2, the best recognition rate is 100% achieved with 600 and 800 GMMs by using MFCC coefficient.
- The “AMAGGWAJ” word its number of syllables is 3, and its best recognition rate is 100% found with MFCC and FBANK coefficients by using 1000 and 400 GMMs respectively.
- The “ANAKMAR” word it is a number of syllables is 3, and its best recognition rate is 100% found with 600 and 800 GMMs by using FBANK coefficient.

- The "AWAR" word its best recognition rate is 100%, found for 400, 600, 800, and 1000 GMMs by using MFCC and FBANK coefficients, and for PLP coefficient by using 400, 800, and 1000 GMMs, with its number of syllable is 2.
- The "AZLMAD" word its best recognition rate is 100%, found for 1000 GMMs by using MFCC coefficient, and for the PLP coefficient by using 400 and 1000 GMMs, also for 400 GMMs by using FBANK coefficient with its number of syllable is 2.
- The "AMYA" digit its best recognition rate is 100%, found for PLP coefficient by using 400, 800, and 1000 GMMs, also for FBANK coefficient by using 400, 600, 800 and 1000 GMMs with its number of syllables is 2.
- The "KRAD" digit its number of syllables is 2, and its best recognition rate is 100% found with MFCC coefficient by using 600 GMMs, and also with PLP coefficient by using 800 and 1000 GMMs.
- The "SEMUS" digit its recognition rate is 100%, found for 400, 600, and 1000 GMMs by using MFCC and FBANK coefficients, with its number of syllables is two.

The digits and words analysis indicate that the misrecognized Amazigh words are monosyllabic ones like TAM, TZA, DAR and DAT. Our tests and analysis show that the best frequently recognized Amazigh commands are those composed of two or three syllables. While the frequently misrecognized Amazigh commands are monosyllabic. We can say that the number of syllables of Amazigh commands has an effect on the commands recognition rate.

The first objective of this work is to use the Kaldi toolkit to create a speech recognition system for the Amazigh Isolated-Words and Amazigh digits (0–9). As a comparison, we used the HMM-GMM acoustic models with different values of Gaussians (8, 16, and 32 GMMs) and MFCC coefficient trained with Kaldi and CMU Sphinx4 tools in order to establish a comparison in terms of recognition rate. To attain our objective we have performed two tests. The 10 first Amazigh digits (see Table 2) were trained and tested by the system in the first test. The system was trained and tested using the 10 isolated Amazigh words in the second test (see Table 3). The speech audio files used in this work were divided into two disjoint sets, for training and test (see Table 5).

Table 5 Corpus characteristics

Recorder type	Number of recorders used for training	Number of recorders used for testing
Amazigh digits	17	7
Amazigh words	7	3

Figure 12 shows the recognition rate (%) of Amazigh digits in the function of total Gaussian distributions number (GMMs) 8, 16, and 32. In the case of the Kaldi toolkit the result achieved using 8 GMMs is 90.42%, the result obtained using 16 GMMs is 90.28% and the result of 32 GMMs is 90.85%. On the other hand, in the case of CMU Sphinx4 tools, the result gives respectively are 88.7, 88.99, and 86.56 respectively for 8, 16, and 32 GMMs.

Figure 13 shows the recognition rate (%) with both Kaldi and CMU Sphinx4 of Amazigh words in the function of total Gaussian distributions number (GMMs) 8, 16, and 32. Wherein the case of Kaldi, the system correct rates were 87.66, 89.33, and 86.66% for using 8, 16, and 32 GMMs respectively. While in the case of CMU Sphinx4, the system correct rate was 86.66, 86.99, and 85.33% corresponding to 8, 16, and 32 GMMs one by one.

Based on the results of the experience, we can see that Kaldi definitely outperformed CMU Sphinx4 with the use

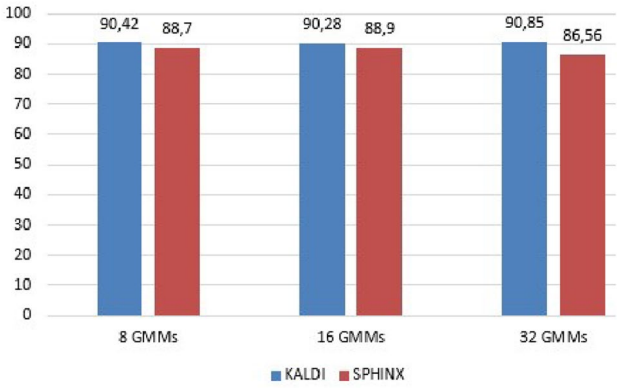

Fig. 12 The recognition rate (%) difference for Amazigh digits between KALDI and SPHINX4

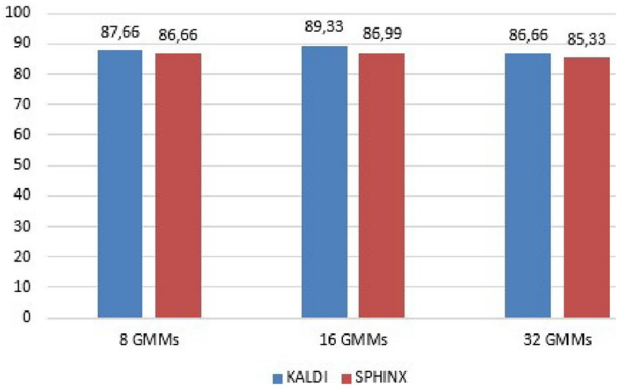

Fig. 13 The recognition rate (%) difference for Amazigh words between KALDI and SPHINX4

Table 6 A summary of some Kaldi ASR systems

Language	Year	Features technique	Acoustic model	Results
Arabic [36]	2021	MFCC	DNN-HMM	97.1%
Bengali [37]	2019	MFCC	GMM-HMM DNN-HMM	WER 2.02% WER 0.92%
Kannada [38]	2019	MFCC	DNN-HMM DNN-SGMM SGMM-MII	DNN-HMM perform better
Bangla [39]	2018	MFCC	GMM-HMM DNN-HMM	WER 3.96% WER 5.30%
German [40]	2015	CMVN MFCC	GMM SGM MDNN	20.5% WER
Proposed work	2022	MFCC PLP FBANK	GMM-HMM	93.96% was obtained by using MFCCs

of GMM-HMM. Table 6 presents the comparison of our obtained results with other works.

7 Conclusion

In this study, we have presented a new approach for the integration of the less-resourced Amazigh language into an isolated variant vocabulary speech recognition system. This system was implemented by using Kaldi toolkit using a 3-State HMM with 400, 600, 800, and 1000 GMMs.. In addition, the MFCCs, FBANK, and PLP feature extraction techniques are used in this work. Our system obtains the best performance of 93.96% when trained by using MFCCs. In another hand, a comparison between Kaldi and CMU Sphinx4 toolkits was presented, and our results showed that Kaldi definitely outperformed CMU Sphinx4 with the use of HMM-GMM.

In our future work, we will be focused on the enhancement of system performances by adopting hybrid and deep learning approaches.

Funding The authors declare they have no financial interests.

Data availability The dataset generated during the current study is not publicly available because it is laboratory-specific data.

Declarations

Conflict of interest All authors declare that there is no conflict of interest.

References

1. Hamidi M, Zealouk O, Satori H, Laaidi N, Salek A (2022) COVID-19 assessment using HMM cough recognition system. Int J Inf Technol 15:193–201
2. Senapati A, Nag A, Mondal A, Maji S (2021) A novel framework for COVID-19 case prediction through piecewise regression in India. Int J Inf Technol 13(1):41–48
3. Hasan I, Dhawan P, Rizvi SAM, Dhir S (2022) Data analytics and knowledge management approach for COVID-19 prediction and control. Int J Inf Technol. https://doi.org/10.1007/s41870-020-00552-3
4. Alafif T, Etaiwi A, Hawsawi Y, Alrefaei A, Albassam A, Althobaiti H (2022) DISCOVID: discovering patterns of COVID-19 infection from recovered patients: a case study in Saudi Arabia. Int J Inf Technol 14(6):2825–2838
5. Barkani F, Satori H, Hamidi M (2020) Cough detection system based on ASR-HMM. In: 2020 Fourth International Conference on Intelligent Computing in Data Sciences (ICDS), IEEE, pp 1–7
6. Young S (2009) The HTK book version 3.4. 1. http://htk.eng.cam.ac.uk
7. Ordowski M, Deshmukh N, Ganapathiraju A, Hamaker J, Picone J (1999) A public domain speech-to-text system. In: Sixth European Conference on Speech Communication and Technology
8. Këpuska V, Bohouta G (2017) Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx). Int J Eng Res Appl 7(03):20–24
9. Ravishankar MK (1996) Efficient Algorithms for Speech Recognition. Carnegie-Mellon University, Pittsburgh PA, Department of Computer Science
10. Zealouk O, Satori H, Laaidi N, Hamidi M, Satori K (2020) Noise effect on Amazigh digits in speech recognition system. Int J Speech Technol 23(4):885–892
11. Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N, Silovsky J (2011) The Kaldi speech recognition toolkit. In: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding (No. CONF), IEEE Signal Processing Society
12. Yadava GT, Jayanna HS (2017) Development and comparison of ASR models using Kaldi for noisy and enhanced kannada speech data. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, pp 1832–1838
13. Medennikov I, Prudnikov A (2016) Advances in STC Russian spontaneous speech recognition system. International conference on speech and computer. Springer, Cham, pp 116–123
14. Zealouk O, Satori H, Hamidi M, Satori K (2019) Speech recognition for Moroccan dialects: feature extraction and classification methods. J Adv Res Dyn Control Syst 11(2):1401–1408
15. Ezzine A, Satori H, Hamidi M, Satori K (2020) Moroccan dialect speech recognition system based on CMU SphinxTools. In: 2020

International Conference on Intelligent Systems and Computer Vision (ISCV), IEEE, pp 1–5
16. Ameen ZJM, Kadhim AA (2023) Machine learning for Arabic phonemes recognition using electrolarynx speech. Int J Electr Comput Eng 13(1):400
17. Saady YE, Rachidi A, Yassa M, Mammass D (2011) Amhcd: a database for amazigh handwritten character recognition research. Int J Comput Appl 27(4):44–48
18. Satori H, Elhaoussi F (2014) Investigation Amazigh speech recognition using CMU tools. Int J Speech Technol 17(3):235–243
19. Hamidi M, Satori H, Zealouk O, Satori K, Laaidi N (2018) Interactive voice response server voice network administration using hidden Markov model speech recognition system. In: 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), IEEE, pp 16–21
20. Hamidi M, Satori H, Zealouk O, Satori K (2020) Amazigh digits through interactive speech recognition system in noisy environment. Int J Speech Technol 23(1):101–109
21. Barkani F, Satori H, Hamidi M, Zealouk O, Laaidi N (2020) Amazigh speech recognition embedded system. In: 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), IEEE, pp 1–5
22. Satori H, Zealouk O, Satori K, Elhaoussi F (2017) Voice comparison between smokers and non-smokers using HMM speech recognition system. Int J Speech Technol 20(4):771–777
23. Hamidi M, Satori H, Zealouk O, Satori K (2019) Speech coding effect on Amazigh alphabet speech recognition performance. J Adv Res Dyn Control Syst 11(2):1392–1400
24. Hamidi M, Satori H, Zealouk O, Satori K (2020) Interactive voice application-based Amazigh speech recognition. Embedded systems and artificial intelligence. Springer, Singapore, pp 271–279
25. Telmem M, Ghanou Y (2018) Estimation of the optimal HMM parameters for Amazigh speech recognition system using CMU-Sphinx. Proced Comput Sci 127:92–101
26. Telmem M, Ghanou Y (2021) The convolutional neural networks for Amazigh speech recognition system. Telecommun Comput Electron Control 19(2):515–522
27. Telmem M, Ghanou Y (2018) Amazigh speech recognition system based on CMUSphinx. Innovations in smart cities and applications: proceedings of the 2nd Mediterranean symposium on smart city applications, 2nd edn. Springer International Publishing, Cham, pp 397–410
28. Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. Ann Math Stat 37(6):1554–1563
29. Rabiner L, Juang B (1986) An introduction to hidden Markov models. IEEE ASSP Mag 3(1):4–16
30. Fine S, Singer Y, Tishby N (1998) The hierarchical hidden Markov model: analysis and applications. Mach Learn 32(1):41–62
31. Beal M, Ghahramani Z, Rasmussen C (2001) The infinite hidden Markov model. Adv Neural Inf Process Syst 14:577–584
32. Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 28(4):357–366
33. Athiramenon G, Anjusha VK (2017) Analysis of feature extraction methods for speech recognition. Int J Innov Sci Eng Technol 4(4)
34. Hermansky H (1990) Perceptual linear predictive (PLP) analysis of speech. J Acoust Soc Am 87(4):1738–1752
35. Meftah A, Alotaibi YA, Selouani SA (2016) A comparative study of different speech features for Arabic phonemes classification. In: 2016 European Modelling Symposium (EMS), IEEE pp 47–52
36. Ouisaadane A, Safi S (2021) A comparative study for Arabic speech recognition system in noisy environments. Int J Speech Technol 24(3):761–770
37. Al Amin MA, Islam MT, Kibria S, Rahman MS (2019) Continuous bengali speech recognition based on deep neural network. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp 1–6
38. Kumar PP, Jayanna HS (2019) Performance analysis of hybrid automatic continuous speech recognition framework for Kannada dialect. In: 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp 1–6
39. Saurav JR, Amin S, Kibria S, Rahman MS (2018) Bangla speech recognition for voice search. In: 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE pp 1–4
40. Radeck-Arneth S, Milde B, Lange A, Gouvêa E, Radomski S, Mühlhäuser M, Biemann C (2015) Open source German distant speech recognition: corpus and acoustic model. International conference on text, speech, and dialogue. Springer, Cham, pp 480–488