Interspeech 2024
1-5 September 2024, Kos, Greece

Speech Recognition for Greek Dialects: A Challenging Benchmark

Socrates Vakirtzian[*1], *Chara Tsoukala*[*2,3], *Stavros Bompolas*[3], *Katerina Mouzou*[1], *Vivian Stamou*[2,3], *Georgios Paraskevopoulos*[2], *Antonios Dimakis*[1,3], *Stella Markantonatou*[2,3], *Angela Ralli*[3,4], *Antonios Anastasopoulos*[3,5]

[1]Department of Informatics, National and Kapodistrian University of Athens [2]ILSP, Athena R.C.
[3]Archimedes AI Unit, Athena R.C. [4]University of Patras, Greece [5]George Mason University, USA

[chara.tsoukala,s.bompolas,vivian.stamou,g.paraskevopoulos,a.dimakis,marks]@athenarc.gr,
socratesvak@hotmail.com, mouzoukaterina25@gmail.com, ralli@upatras.gr, antonis@gmu.edu

Abstract

Language technologies should be judged on their usefulness in real-world use cases. Despite recent impressive progress in automatic speech recognition (ASR), an often overlooked aspect in ASR research and evaluation is language variation in the form of non-standard dialects or language varieties. To this end, this work introduces a challenging benchmark that focuses on four varieties of Greek (Aivaliot, Cretan, Griko, Messenian) encompassing challenges related to data availability, orthographic conventions, and complexities arising from language contact. Initial experiments with state-of-the-art models and established cross-lingual transfer techniques highlight the difficulty of adapting to such low-resource varieties.

Index Terms: low-resource ASR, Greek dialects, benchmarks

1. Introduction

In the progression of language and speech technology development for a particular language, priority tends to be given to varieties and dialects that have a greater abundance of available data. Consequently, this results in an imbalance in the technological support provided to speakers of different dialects within a language. For example, despite significant efforts in English language research, only a small portion of previous studies have focused on dialects or varieties such as African-American Vernacular English compared to Mainstream American English [1]. However, recent advancements in cross-lingual and low-resource Natural Language Processing (NLP) and Automatic Speech Recognition (ASR), coupled with improved accessibility to data for underrepresented language varieties, have enabled researchers to move beyond the traditional notion of a singular language. Likewise, there is a growing call within the NLP community for a more inclusive representation of various dialects and varieties, as highlighted by several initiatives [2, 3].

Most recent work on speech technologies for dialectal resources has focused on varieties of high-resource macrolanguages, like English [4, 5, 6], Arabic [7], Chinese [8, 9], Japanese [10], or German [11]. Other work has focused on endangered languages like North Sami [12] or Irish [13]. Last, some work has focused on mid-resource languages, like Thai [14] (with 40h of audio transcribed for each variety) or Telugu [15] (with more than 100h of audio per variety). For the majority of languages beyond high-resource ones, previous work has largely focused on dialect identification as opposed to transcription.

[*]Equal contribution.

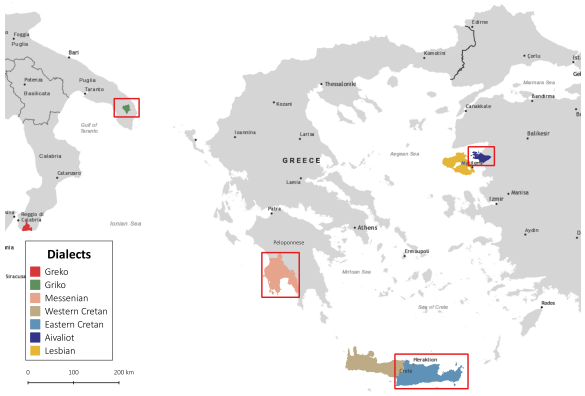


Figure 1: *Geographic distribution of the Greek dialects mentioned in this work. Red boxes mark the dialects under analysis.*

This work focuses on a particularly challenging domain: speech recognition of Greek varieties. One of the primary challenges stems from the low-resource nature of these dialects. The scarcity of available and, at least partially, transcribed datasets for training models significantly impedes the development of robust speech recognition systems. Moreover, the inherent phonological and orthographic variation across these varieties introduces additional complexity. Unlike standardized languages, such as Greek, that have well-documented pronunciation and spelling rules, Greek dialects exhibit a wide range of phonetic and orthographic disparities [16]. This diversity poses significant challenges not only for the human transcription process, typically carried out by field linguists and/or native speakers, but it also complicates the adaptation or refinement of pre-existing speech models. Furthermore, the historical influences from other languages, notably Turkish (in the case of Aivaliot) and Italian (in the case of Griko), infuse these Greek dialects with unique lexical, phonetic, and syntactic elements [17, 18]. Together, these three factors make the task of speech recognition for low-resource Greek varieties not only technically demanding but also an intriguing domain for advancing the state of the art in language processing technologies.

We introduce a benchmark comprised of four low-resource Greek varieties, with a total of 12 hours 33 minutes. We additionally present results with state-of-the-art ASR systems leveraging cross-lingual adaptation techniques. Our experiments highlight how challenging this benchmark is and demonstrate that dialects bearing closer phonetic and orthographic resemblance to Standard Greek exhibit comparatively better performance in speech recognition tasks.

10.21437/Interspeech.2024-2443

2. Greek dialects

The current study investigates four Greek varieties: Aivaliot, (Eastern) Cretan, Griko, and Messenian. Alongside Standard Modern Greek and relevant contact languages/varieties, these varieties form a dialectal landscape (see Figure 1) that offers an ideal testing ground for assessing the performance of speech models when encountering dialectal variation, particularly of those exhibiting varying degrees of influence from other languages, mainly Italo-Romance and Turkish. Further details concerning these varieties are provided below.

Aivaliot is a variety of Greek that was spoken in Aivali (known as Ayvalık in Turkish), located on the Edremit Gulf in Western Turkey, till the beginning of the 20th century. After the end of the war between Greece and Turkey (1919–1922) and the defeat of the Greek army, those Aivaliots who managed to survive flew to Greece, principally to the nearby island of Lesbos, where they settled in various dialectal enclaves. Aivaliot resembles Lesbian in many respects. According to Ralli [17], Aivaliot and Lesbian belong to the group of Northern Greek Dialects, sharing unstressed /i/ and /u/ deletion and unstressed /o/ and /e/ raising. Aivaliot morphology and the lexicon are influenced by Turkish, because of a long domination by the Ottomans, as well as by Italo-Romance, due to the pre-Ottoman Genovese rule and trade with Venice [18]. However, there are no Turkish or Italo-Romance influences on phonology or syntax. In 2002, a handful of first-generation Aivaliot speakers could still be found in Lesbos and elsewhere in Greece and abroad, where they still remembered and practiced their mother tongue [17]. Nowadays, the dialect is on the way to extinction, since second-generation speakers either have a passive knowledge of it, or those living in Lesbos mix their own dialectal variety with the parent Lesbian.

Cretan is a variety of Modern Greek predominantly used by speakers who reside on the island of Crete or belong to the Cretan diaspora. This includes communities of Cretan origin that were relocated to the village of Hamidieh in Syria and to Western Asia Minor, following the population exchange between Greece and Turkey in 1923. The historical and geographical factors that have shaped the development and preservation of the dialect include the long-term isolation of Crete from the mainland, and the successive domination of the island by foreign powers, such as the Arabs, the Venetians, and the Turks, over a period of seven centuries. Cretan has been divided based on its phonological, phonetic, morphological, and lexical characteristics into two major dialect groups: the western and the eastern. The boundary between these groups coincides with the administrative division of the island into the prefectures of Rethymno and Heraklion. Kontosopoulos [19] argues that the eastern dialect group is more homogeneous than the western one, which shows more variation across all levels of linguistic analysis. Contrary to other Modern Greek Dialects, Cretan does not face the threat of extinction, as it remains the sole means of communication for a large number of speakers in various parts of the island.

Griko (or Grico) is part of Italiot, also known as South Italian Greek, which is a dialectal group that contains Griko and Greko. Griko is a Greek variety spoken in Salento (province of Lecce, the so-called Grecía Salentína), and Greko (or Greco) is spoken in South Calabria, in the Bovese area. Today, there are approximately nine Griko villages in Salento, and about five Greko ones in South Calabria. Griko and Greko have been influenced by the local Romance varieties and Italian, the official language of the Italian state [20, 21]. This influence is shown on all linguistic levels: phonology, morphology, syntax and the lexicon. Nevertheless, they preserve several archaic features, some of which can be traced back to Ancient Greek and are not detected in the other Modern Greek dialects [22]. Manolessou [21] suggests that the existence of archaisms in Italiot may be due to the break of contact between South Italy and the rest of the Greek-speaking world, which occurred around the 13th century. While Greko is facing a rapid decline [23], Griko seems to show some resilience, although it is mainly spoken by the older generations.

Messenian is spoken in the Southwest of the Peloponnese and, thus, it belongs to the Peloponnesian varieties. Historically, Peloponnesian has been considered as one of the basic dialects on which Standard Modern Greek was formed [24, 19], mainly because of the pivotal role of the Peloponnese during the Greek Revolution of 1821 and the subsequent formation of the Greek state, as well as of the significant migratory movements from the Peloponnese to Athens. Recent studies of some documented linguistic material from the Peloponnesian varieties reveal substantial deviations from Standard Modern Greek [25], and differences from one variety to the other, particularly on the phonological level [26]. For instance, they display palatalization of the lateral /l/ and the nasal /n/. However, detailed empirical studies and an account of the exact properties of Messenian and how it diverges from other Peloponnesian dialects are lacking.

3. Resources

Aivaliot The present oral corpus is a component of the Asia Minor Archive (AMiGre). It was compiled within the framework of two research projects that ran in the periods 2002-2005 and 2012-2016 (see [27, 28]). We obtained permission to use it from the studies' authors. It consists of narratives elicited from 18 elderly speakers (5 male, 13 female), all refugees from Aivali, who had settled in different villages of the island of Lesbos. The data collection was carried out in 2002-2003, after obtaining a written consent of the informants, as well as the approval of the Ethics committee of the University of Patras. The corpus has a total duration of approximately 14 hours. It has been transcribed and annotated by two native speakers of the dialect, using a transcription system based on the Greek alphabet and orthography, which is adapted according to SAMPA [29]. The annotations include metadata information, such as the source of the data, the identity and background of the informants, and the conditions of the data collection. The corpus is stored on the server of the Laboratory of Modern Greek Dialects of the University of Patras and is freely accessible online.[1]

Cretan For the compilation of this corpus, we gathered 32 tapes containing material from radio broadcasts in digital format, with permission from the Audiovisual Department of the Vikelaia Municipal Library of Heraklion, Crete. These broadcasts were recorded and aired by Radio Mires, in the Messara region of Heraklion, during the period 1998-2001, totaling 958 minutes and 47 seconds. These recordings primarily consist of narratives by one speaker, Ioannis Anagnostakis, who is responsible for their composition. In terms of textual genre, the linguistic content of the broadcasts consists of folklore narratives expressed in the local linguistic variety. Out of the total volume of material collected, we utilized nine tapes. Criteria for material selection included, on the one hand, maximizing digital clarity of speech and, on the other hand, ensuring representative sampling across the entire three-year period of radio recordings. To obtain an initial transcription, we employed the Large-v2 model, which was the largest Whisper model at the time. Subsequently, the transcripts were manually corrected in collaboration with the

[1] http://amigredb.philology.upatras.gr

Table 1: *Speech corpus statistics.*

Corpus	Tokens	Utterances	Audio Duration Original	Audio Duration Processed
Aivaliot	64,821	10,916	13h 50m 31s	10h 14m 44s
Cretan	12,921	2,589	2h 1m 35s	1h 21m 12s
Griko	2,374	330	0h 20m 0s	0h 20m 0s
Messenian	4,721	590	0h 39m 45s	0h 37m 42s

local community. The transcription system that was used was based on the Greek alphabet and orthography.

Griko The corpus was collected in 2013 during a field trip in Puglia, Italy by two linguists, with a particular focus on the use of infinitive and verbal morphosyntax [30, 31]. The corpus[2] contains utterances from nine different speakers (5 male, 4 female) from the four villages (Calimera, Sternatia, Martano, Corigliano) where native speakers could still be found. The digitally collected audio files were manually segmented into utterances, transcribed, glossed in Italian, and annotated with extensive morphosyntactic tags by trained linguists. Here we re-purpose the audio component along with the transcriptions.

Messenian To assemble the corpus, we interviewed residents from the town of Kalamata and five closeby villages (Bounaria, Sotirianika, Petalidi, Filiatra, Altomira), resulting in 39 minutes of narratives obtained from six speakers (2 male, 4 female). This data collection was carried out in 2023-2024, after obtaining written consent from the informants. For the initial transcription of the audio files, Large-v3 was employed. The transcripts were then manually corrected by a native speaker of the dialect.

4. Benchmark Description

To evaluate the effectiveness of cutting-edge speech recognition models in understanding Greek dialects, we employed two leading ASR models: XLS-R [32] and Whisper [33]. XLS-R[3], one of the first large multilingual speech models, was trained on 56k hours of audio across 53 languages. Whisper was trained on a much larger corpus, with the Large-v2 model trained on 680k hours of labeled speech data, and the latest Large-v3[4] trained on 1 million hours of weakly labeled audio and 4 million hours of pseudo-labeled audio collected using Large-v2 [33]. For the XLS-R inference and fine-tuning, we used a model that had been further fine-tuned on Greek (XLS-R-greek)[5].

Given the limited amount of data available for each of the dialects we report on, which range from a few minutes to a few hours, it was not feasible to fine-tune Large-v3, due to the large model parameters and the fact that it quickly overfits in small datasets. However, for the two dialects that have more than an hour of recordings (i.e., Aivaliot, Cretan), we successfully fine-tuned the XLS-R and Whisper-medium models to better accommodate the nuances of the Greek dialects, incorporating these results into our analysis.

Preprocessing Steps[6] To prepare for the analysis, the following steps were uniformly applied across all dialect samples: The texts were normalized[7], and all audio files were converted into a 16 kHz mono format. This preparation was crucial in ensuring that the input data was consistent and optimized for fine-tuning and evaluation by the ASR models.

Data Segmentation Automatic speech recognition systems require short audio segments as training input. Typically, audio segments of up to 30 seconds are used to fine-tune a model. Therefore, to allow the creation of a speech corpus for fine-tuning purposes, the original recordings had to be segmented into smaller parts. Griko was already available in short audio-text segments and the Aivaliot corpus had already been transcribed using Praat[8], which contains timestamp annotations. Using these timestamps, we automatically exported the audio-text segments. For the two newly collected corpora, Cretan and Messenian, we first transcribed the audio using Whisper, and subsequently converted the output and timestamp predictions to a Praat (TextGrid) file. Native speakers corrected the transcriptions and timestamps, and we were able to export the audio-text segments using the corrected TextGrids. Note that the removal of music, long pauses, and non-transcribed segments leads to a reduction of the total audio duration, as can be seen on Table 1 for Aivaliot and Cretan.

Dataset Creation Based on the audio-text pairs created in the previous step, we created a dataset for each language, excluding audio segments that were longer than 30 seconds. Aivaliot and Cretan were split into training, dev, and test sets (80%-10%-10%) in order to allow the fine-tuning of the models, and the results are reported on the test set. For the other two dialects (Griko and Messenian) that had less than an hour of data, the dataset was not split and results are reported on the full set (inference-only). The descriptions of the final speech corpora are provided in Table 1.

Fine-tuning process To obtain the Cretan and Aivaliot models, we fine-tuned XLS-R-greek and Whisper-medium for 35 epochs on the respective datasets on an NVIDIA GeForce RTX 3090. The hyperparams can be seen on the model pages.

Table 2: *Aivaliot model performance comparison.*

Model	Epoch	WER (%)	CER (%)
Large-v3	pretrained	109.60%	80.03%
Large-v2	pretrained	124.90%	98.63%
XLS-R-greek	pretrained	113.67%	104.80%
XLS-R-greek[9]	35	**73.83**%	**39.55**%
Whisper-medium[10]	35	83.75%	60.19%

Table 3: *Model performance comparison on Cretan.*

Model	Epoch	WER (%)	CER (%)
Large-v3	pretrained	58.42%	26.44%
Large-v2	pretrained	74.60%	37.82%
XLS-R-greek	pretrained	104.83%	91.73%
XLS-R-greek[11]	35	**28.27**%	**7.88**%
Whisper-medium[12]	35	47.87%	17.83%

5. Results

ASR performance of the benchmarked models in each variety, measured by WER and CER, is shown in Tables 2–5.

[2] github.com/antonisa/griko-italian-parallel-corpus

[3] huggingface.co/facebook/wav2vec2-large-xlsr-53

[4] huggingface.co/openai/whisper-large-v3

[5] huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-greek

[6] All scripts can be found at: gitlab.com/ilsp-spmd-all/speech/greek_dialects_asr

[7] Griko uses a standardized Latin transcription. The other dialects were normalized using Standard Modern Greek orthography for common phenomena and the Greek alphabet for dialect-specific phenomena, based on legacy texts whenever available.

[8] https://www.fon.hum.uva.nl/praat/

[9] https://huggingface.co/ilsp/xls-r-greek-aivaliot

[10] https://huggingface.co/ilsp/whisper-medium-aivaliot

Table 4: *Whisper zero-shot performance on full Griko dataset.*

Model	Lang	WER (%)	CER (%)
Large-v3	el	108.29%	99.68%
Large-v3	it	113.07%	64.36%
Large-v3	automatic	108.77%	100.31%
Large-v2	el	128.49%	110.13%
Large-v2	it	121.58%	88.85%
Large-v2	automatic	129.27%	111.20%
+romanizing Greek output			
Large-v3	el	**102.00%**	**62.97%**
Large-v2	el	122.19%	75.48%
Evaluation ignoring diacritics			
Large-v3	it	108.14%	**55.98%**
Large-v2	it	118.01%	82.25%
+romanizing Greek outputs			
Large-v3	el	98.62%	**54.29%**
Large-v2	el	119.56%	66.99%

Table 5: *Whisper zero-shot performance on Messenian.*

Pre-trained Model	WER (%)	CER (%)
Large-v3	**32.71**%	**18.28**%
Large-v2	50.40%	37.27%
XLS-R-greek	105.84%	91.86%

The output quality differs substantially for each variety. For Aivaliot, the performance of the pretrained models is unacceptable with word error rates[13] above 100%. However, fine-tuning the models with the limited training data leads to substantial improvements: XLS-R-greek results in a CER around 40% and WER of 74%, while Whisper-medium leads to a WER of 84%. This behavior is perhaps expected, since Aivaliot uses a lot of borrowings from Turkish, leading to a large vocabulary discrepancy between what the pre-trained models were tested on.

For Cretan, the situation is more encouraging. Pre-trained models are somewhat more competitive, with Large-v3 achieving a WER of 58% in zero-shot manner. Again, fine-tuning on the limited training set leads to substantial improvements, with XLS-R-greek achieving a dialect-low WER of 28% and a CER of only 8%. Fine-tuned Whisper-medium results in a higher WER (48%), but it still outperforms the inference-only Large-v3 one.

For Griko, its influences from Italo-Romance and its use of the Latin alphabet (instead of the Greek one, which all other Greek varieties use) complicates things. We remind the reader that due to the small dataset size we can only evaluate using Whisper models in a zero-shot manner. For this evaluation, we specify the audio language as either Greek (el) or Italian (it), or we allow automatic language detection, which typically outputs in the Greek alphabet. In all cases, WER is above 100%, but it seems that using the Italian flag leads to significantly better CER, compared to Greek. This, however, can be misleading, as it is purely due to the alphabet similarity. By employing a simple romanization step, using the `uroman` [34] tool on the Greek output, we find that the Greek flag yields better CER at 63%.

Another confounder on the Griko evaluation is that its orthography uses diacritics to mark stress on almost all polysyllabic words. This is an issue, as it is not common in Italian – and while Modern Greek does use stress marking, the romanization process will remove such diacritics. To quantify the effect of this issue, we perform an additional evaluation, this time removing all diacritics from the references, as well as any diacritics from the models' outputs. We find that around 8-10 CER points are due to this diacritics issue (see last four rows of Table 4).

Last, on Messenian, despite the necessity for zero-shot evaluation due to the lack of training or adaptation data, Whisper Large-v3 achieves a decent performance with WER around 32% and CER around 18%. This good performance even without adaptation is not unexpected, since Messenian is the closest to Standard Modern Greek, out of the varieties we study.

6. Discussion

It is important to note that all the models we use perform well in Standard Greek. In the Greek split of the Common Voice test set [35], XLS-R-greek achieves 11.62% WER and Whisper Large-v3 13.7% WER. In contrast, the lowest WER we obtain is more than double of the Standard Greek one: 28% on Cretan (after fine-tuning on substantial data), 33% on Messenian, and up to 100% for Griko. This increase in WER could be partially attributed to genre differences: Common Voice data consists of read speech, while the dialect corpora are oral narratives, which are more challenging.

Our experiments on ASR for Greek dialects highlight challenges in current methodologies and the inherent difficulties in applying out-of-the-box ASR models, trained on mainstream languages, for the linguistically rich dialects we have explored. This is in line with previous research that has shown that even industry ASR systems struggle with language variations [36]. Therefore, there is a need to develop specialized approaches that are specifically tailored to the phonetic, syntactic, and lexical characteristics of dialects. These approaches must also be capable of functioning effectively in situations with limited data, as required by this task.

To address these challenges, there is significant potential for exploration on the modeling front, particularly through the use of modern unsupervised and weakly supervised adaptation techniques. These approaches allow for leveraging the similarities between dialects and more widely spoken languages (e.g., the relationship between Griko and both Greek and Italian) and adapting to specific accents and vocabularies unique to each dialect (e.g., Aivaliot or Cretan). These strategies are beneficial not only for accent adaptation and vocabulary enhancement but also for facilitating the efficient adaptation to new dialects when resources are scarce. The adaptation techniques proposed in [37, 38, 39] can offer good starting points for this exploration.

Moreover, there is important work to be done on the data frontier, especially for dialects without a standard writing system. The absence of a standardized orthography makes native speakers use various spellings for the same words. Furthermore, in larger communities like the Cretan speaking one, words may be pronounced differently yet written identically to their Standard Greek counterparts. Additionally, without standardized spelling, field linguists often transcribe words as they hear them, resulting in spellings that diverge from Standard Greek, even for phonetically similar words. Consequently, it is essential to consider how to standardize the written form of these dialects when developing speech models, as it plays a critical role in enhancing model accuracy. An alternative research route is to explore developing ASR systems that can robustly train on and handle multiple orthographies and phonological systems for the same language.

[11] https://huggingface.co/ilsp/xls-r-greek-cretan

[12] https://huggingface.co/ilsp/whisper-medium-cretan

[13] Error rates are computed on normalized texts without punctuation.

7. Acknowledgements

This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program. Antonios Anastasopoulos is also partially funded by the US National Science Foundation under award IIS-2125466.

8. References

[1] S. L. Blodgett, J. Wei, and B. T. O'Connor, "Twitter Universal Dependency Parsing for African-American and Mainstream American English," in *Proc. ACL 2018, Melbourne, Australia*, 2018.

[2] B. Plank, "What to do about non-standard (or non-canonical) language in NLP," in *Proc. KONVENS*, 2016.

[3] M. Zampieri, P. Nakov, and Y. Scherrer, "Natural language processing for similar languages, varieties, and dialects: A survey," *Natural Language Engineering*, vol. 26, no. 6, pp. 595–612, 2020.

[4] J. Eisenstein, V. Prabhakaran, C. Rivera, D. Demszky, and D. Sharma, "MD3: The Multi-Dialect Dataset of Dialogues," in *Proc. INTERSPEECH*, 2023, pp. 4059–4063.

[5] J. L. L. Chin, E. Talevska, and M. Antoniou, "Speech-in-Speech Recognition is Modulated by Familiarity to Dialect," in *Proc. INTERSPEECH*, 2023, pp. 3113–3116.

[6] J. Meyer, L. Rauchenstein, J. D. Eisenberg, and N. Howell, "Artie bias corpus: An open dataset for detecting demographic bias in speech applications," in *Proc. PREC*, Marseille, France, 2020.

[7] S. Radhakrishnan, C.-H. H. Yang, S. A. Khan, N. A. Kiani, D. Gomez-Cabrero, and J. N. Tegner, "A Parameter-Efficient Learning Approach to Arabic Dialect Identification with Pre-Trained General-Purpose Speech Model," in *Proc. INTERSPEECH*, 2023.

[8] Y. Hu, H. Feng, Q. Zhao, and A. Li, "Effects of Tonal Coarticulation and Prosodic Positions on Tonal Contours of Low Rising Tones: In the Case of Xiamen Dialect," in *Proc. INTERSPEECH*, 2023, pp. 4748–4752.

[9] Y. Li, Y. Chen, X. Zhang, Y. Chen, and J. Wang, "Effects of Language Contact on Vowel Nasalization in Wenzhou and Rugao Dialects," in *Proc. Interspeech*, 2022.

[10] S. Miwa and A. Kai, "Dialect Speech Recognition Modeling using Corpus of Japanese Dialects and Self-Supervised Learning-based Model XLSR," in *Proc. INTERSPEECH*, 2023.

[11] C. Steiner, D. Studer-Joho, C. Lanthemann, A. Büchler, and A. Leemann, "Sociodemographic and Attitudinal Effects on Dialect Speakers' Articulation of the Standard Language: Evidence from German-Speaking Switzerland," in *Proc. INTERSPEECH*, 2023, pp. 3542–3546.

[12] S. Kakouros and K. Hiovain-Asikainen, "North Sámi Dialect Identification with Self-supervised Speech Models," in *Proc. INTERSPEECH*, 2023.

[13] L. Lonergan, M. Qian, N. N. Chiaráin, C. Gobl, and A. N. Chasaide, "Cross-dialect lexicon optimisation for an endangered language ASR system: the case of Irish," in *Proc. Interspeech*, 2022.

[14] A. Suwanbandit, B. Naowarat, O. Sangpetch, and E. Chuangsuwanich, "Thai Dialect Corpus and Transfer-based Curriculum Learning Investigation for Dialect Automatic Speech Recognition," in *Proc. INTERSPEECH*, 2023, pp. 4069–4073.

[15] A. Yadavalli, G. Mirishkar, and A. K. Vuppala, "Multi-Task End-to-End Model for Telugu Dialect and Speech Recognition," in *Proc. Interspeech*, 2022.

[16] I. Manolessou, S. Beis, and C. Basea-Bezantakou, "The phonetic transcription of the Modern Greek dialects [in Greek]," *Lexikografikon Deltion*, vol. 26, pp. 161–222, 2012.

[17] A. Ralli, "Affixoids and Verb Borrowing in Aivaliot Morphology," in *The Morphology of Asia Minor Greek*. Brill, 2019.

[18] ——, "Greek in Contact with Romance," in *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 2019.

[19] N. Kontosopoulos, *Dialects and idioms of Modern Greek [in Greek]*, 5th ed. Athens: Grigoris, 2008.

[20] P. Martino, "L'isola grecanica dell'Aspromonte. Aspetti sociolinguistici," in *I dialetti e le lingue delle minoranze di fonte all'Italiano*, A. F. Leoni, Ed. Roma: Bulzoni, 1980, pp. 305–341.

[21] I. Manolessou, "The Greek dialects of South Italy: An overview," *Kambos: Cambridge Papers in Modern Greek*, vol. 13, 2005.

[22] G. Rohlfs, *Studi e ricerche su lingua e dialetti d'Italia*. Firenze: Sansoni, 1977.

[23] M. Katsoyannou, "Le parler gréco de Gallicianò (Italie) : description d'une langue en voie de disparition," PhD Thesis, 1995. [Online]. Available: http://www.theses.fr/1995PA070123

[24] G. C. Horrocks, *Greek: a history of the language and its speakers*, 2nd ed. Oxford ; Malden, Mass: Wiley-Blackwell, 2010.

[25] N. Pantelidis, "Peloponnesian dialectal accent and Standard Modern Greek [in Greek]," in *Proc. of the Fourth International Conference of Greek Linguistics*, P. Pavlou and A. Roussou, Eds. Thessaloniki: University Studio Press, 2001, pp. 480–486.

[26] ——, "Phonetic observations in a Messenian dialect [in Greek]," *Studies in Greek Linguistics*, vol. 21, pp. 550–561, 2001.

[27] A. Ralli, Ed., *THALIS Program: "Pontus, Cappadocia, Aivali: in search of Asia Minor Greek" [in Greek]*. Patras: Laboratory of Modern Greek dialects / University of Patras, 2015.

[28] E. Galiotou and A. Ralli, "AMiGre: A unified framework for archiving and processing oral and written dialectal data," in *AIUCD2018 - Book of Abstracts*, D. Spampinato, Ed., 2018, pp. 25–27, alma Mater Studiorum - Università di Bologna.

[29] J. C. Wells, "SAMPA computer readable phonetic alphabet," in *Handbook of standards and resources for spoken language systems*, D. Gibbon, R. Moore, and R. Winski, Eds. Berlin: Mouton de Gruyter, 1997, vol. 4, pp. 684–732.

[30] M. Lekakou, V. Baldissera, and A. Anastasopoulos, "Documentation and analysis of an endangered language: aspects of the grammar of griko," *University of Ioannina*, 2013.

[31] A. Anastasopoulos, M. Lekakou, J. Quer, E. Zimianiti, J. DeBenedetto, and D. Chiang, "Part-of-speech tagging on an endangered language: a parallel Griko-Italian resource," in *Proc. COLING*, Santa Fe, New Mexico, USA, 2018.

[32] A. Babu, C. Wang, A. Tjandra, K. Lakhotia, Q. Xu, N. Goyal, K. Singh, P. von Platen, Y. Saraf, J. Pino, A. Baevski, A. Conneau, and M. Auli, "XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale," in *Proc. Interspeech*, 2022.

[33] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022, arXiv:2212.04356.

[34] U. Hermjakob, J. May, and K. Knight, "Out-of-the-box universal romanization tool uroman," in *Proc. of ACL 2018, system demonstrations*, 2018, pp. 13–18.

[35] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common voice: A massively-multilingual speech corpus," in *Proc. LREC*, 2020, pp. 4211–4215.

[36] A. B. Wassink, C. Gansen, and I. Bartholomew, "Uneven success: automatic speech recognition and ethnicity-related dialects," *Speech Communication*, vol. 140, pp. 50–70, 2022.

[37] T. Asami, R. Masumura, Y. Yamaguchi, H. Masataki, and Y. Aono, "Domain adaptation of dnn acoustic models using knowledge distillation," in *Proc. ICASSP*. IEEE, 2017.

[38] G. Paraskevopoulos, T. Kouzelis, G. Rouvalis, A. Katsamanis, V. Katsouros, and A. Potamianos, "Sample-efficient unsupervised domain adaptation of speech recognition systems: A case study for Modern Greek," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[39] D. Prabhu, P. Jyothi, S. Ganapathy, and V. Unni, "Accented speech recognition with accent-specific codebooks," 2023, arXiv:2310.15970.