A Speech Resource for the Pontic Greek Dialect: Transcription Choices and Baseline ASR Evaluation

Rodanna Konstantinidou[1], **Chara Tsoukala**[2], **Vivian Stamou**[2]
Voula Giouli[1,2], **Stella Markantonatou**[2]
[1]Aristotle University of Thessaloniki, [2]ILSP, Athena R.C.
{rkonstaa,pgiouli}@del.auth.gr, {chara.tsoukala,vivian.stamou,marks}@athenarc.gr

Abstract

Pontic Greek is a living but endangered Modern Greek dialect that lacks publicly available AI-oriented speech resources and ASR benchmarks. This work reports on the first systematic inference-only (zero-shot) ASR evaluation on authentic Pontic speech. Progress on Pontic ASR is hindered by two coupled challenges: the scarcity of transcribed speech data and the absence of a standardized orthography, which makes it difficult to create consistent reference transcriptions for evaluation. We address these challenges by releasing a new speech corpus of contemporary Pontic as spoken in Northern Greece, derived from natural conversations and provided with manual, utterance-level, time-aligned transcriptions. To reduce annotator bias and increase practical usability, we collect community evidence on written-form preferences via a small questionnaire and use the observed patterns to guide a consistent Greek-script transcription scheme. We use this corpus to perform inference-only (zero-shot) ASR evaluation, benchmarking four state-of-the-art pretrained speech recognition models under a unified evaluation protocol. Results show that zero-shot recognition remains challenging, establishing baseline figures and underscoring the need for dialect-specific data and adaptation.

Keywords: non-standard languages, low-resource ASR, Pontic dialect

1. Introduction

We initiate an effort to provide Pontic Greek, a living, endangered dialect of Modern Greek, with AI-oriented tools and resources to contribute to its vitality. In this work, the first attempt to evaluate Automatic Speech Recognition (ASR) tools against authentic Pontic spoken data is presented. The enterprise raises questions about the orthographic representation of the dialect; this study makes a first attempt to discuss this issue by consulting a small part of the community and it appears to reveal interesting requirements.

Our evaluation relies on ASR systems trained on the related well-resourced language variety, namely Standard Modern Greek (SMG). Using technology developed for a standard variety in order to work with related dialectal data is almost a rule-of-thumb since standard varieties tend to have more resources than the related dialects. However, the fact is that the more distant the dialect from the standard, the less accurate results are obtained. Pontic turns out to be an excellent example of this situation.

Our contributions are threefold: (i) a new Pontic Greek speech corpus[1] (Northern Greece) with utterance segmentation and time-aligned manual transcriptions (Praat/TextGrid), enabling reproducible speech-text pairing for ASR evaluation; (ii) community-informed, internally consistent transcription conventions in Greek script (favoring broadly usable forms and avoiding diacritics) to support reliable ASR references and other downstream work; and (iii) the first systematic inference-only (zero-shot) ASR baseline on authentic Pontic speech, benchmarking state-of-the-art pretrained speech recognition models using WER and CER.

Figure 1: Geographic extent of the Pontus region until 1922 (olive) and the main areas of Pontian settlement in Greece following migration (blue).

The paper is structured as follows: An introduction to the history of Pontic is given along with a brief description of its phonological features that differentiate it from SMG. Next, we discuss relevant work with a focus on work on other Modern Greek dialects. The methodology we used to collect, transcribe and prepare the data for the ASR experiments is presented next followed by a description of the experimental work. We conclude with a brief discussion of the results.

[1]https://huggingface.co/datasets/ilsp/pontic-speech-corpus

Proceedings of Workshop on Dialects in NLP — A Resource Perspective (DialRes) , pages 300–307
16 May 2026.

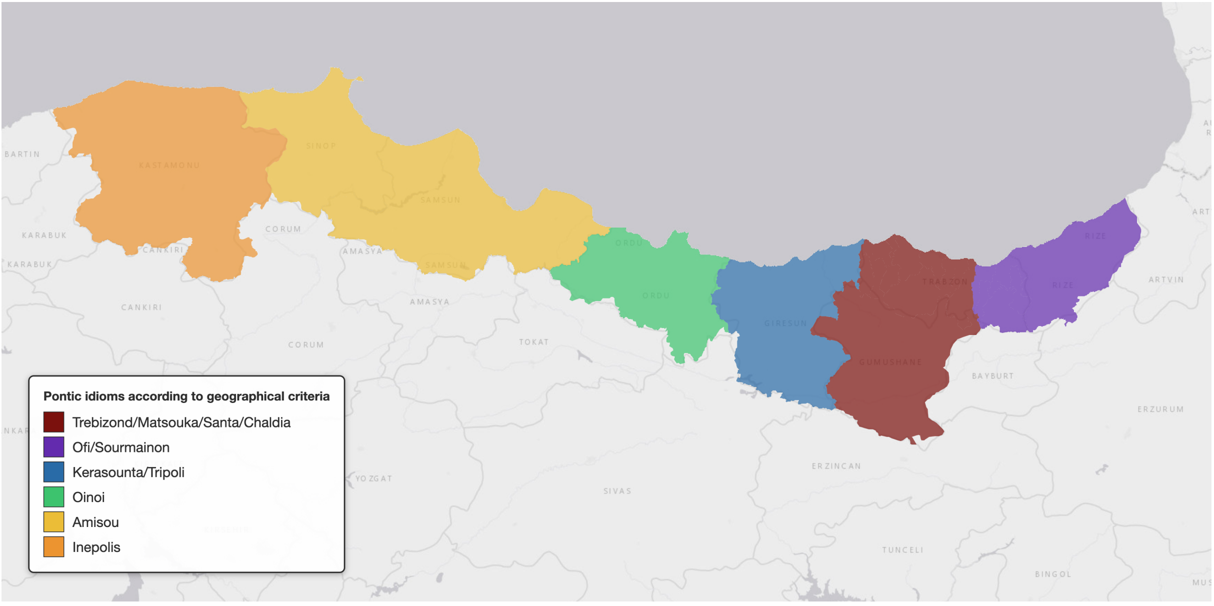

Figure 2: Pontic idioms according to geographical criteria (Papadopoulos, 1953)

2. About Pontic Greek

Pontic Greek (ISO 639-3: pnt; Glottolog: pont1253) is a Modern Greek dialect that belongs to the Indo-European language family. The dialect was spoken until 1922 along the southern coast of the Black Sea and in the interior of Asia Minor, extending from Inepolis (in the west) to Rizounta and Colchis (in the east) as seen in Figure 1. Beyond the coast, Pontic was spoken south of Trabzon, particularly in the regions of Gemoura, Matsouka, Santa, Kromni, Chaldia, and Cherianon, as well as in the mining region, the most prominent of which being Gümüş-Maden and Akdağmadeni (Tombaidis, 1988). In several areas, Pontic speakers (Pontians) coexisted with Turkish-speaking populations, a factor that significantly influenced the development of the dialect (Drettas, 1997; Dimela, 2013).

According to Papadopoulos (1953), the Pontic dialect was divided into six different idioms based on geographical criteria: (i) the idiom of Trebizond, Matsouka, Santa, and Chaldia (the most populous), (ii) Ofi and Sourmainon, (iii) Kerasounta and Tripoli, (iv) Oinoi, (v) Amisou, and (vi) Inepolis, as seen in Figure 2.

After the population exchange in 1923, following the signing of the Treaty of Lausanne, Pontian refugees settled in various areas of Greece, mainly in Northern Greece (Macedonia and Thrace) and, to a lesser extent, in Attica (Figure 1); Pontians also settled out of Greece, namely in the Caucasus region and Ukraine. Currently, the dialect is estimated to be spoken by approximately 400.000 speakers in Greece according to Ethnologue (Eberhard et al., 2026). UNESCO (UNESCO, 2010) classifies the Pontic dialect as "definitely endangered" as its use is declining and younger generations tend to adopt Standard Modern Greek (SMG).

In this study, we focus on the variety of Pontic as spoken today in Northern Greece, which is the most widespread contemporary form of the dialect.

A primary challenge for the development of ASR tools for the Pontic dialect is the total lack of publicly available datasets of recorded and transcribed speech. In addition to the lack of data, the internal variation of the dialect is significant.

Pontic Greek exhibits characteristic phonological features that complicate the development of ASR models since they differentiate it from SMG:
1. The vowels as /æ/ and /ø/, often represented orthographically with letters of the Greek alphabet plus diacritics (e.g. α̈, ο̈); they originate from vowel coalescence phenomena (e.g. ια, εα > α̈).
2. The consonants /ʃ/, /kʃ/, /pʃ/, typically rendered as σ̆, ξ̆, ψ̆, which are absent from SMG phonology.

Like many other dialects of Modern Greek, Pontic includes spoken features like the ones listed above, that are not represented with the orthographic conventions of SMG. Several proposals have been made regarding the orthography of Pontic: Papadopoulos (1955, 1958), Symeonidis and Tombaidis (1999), Oikonomidis (1958), Valavanis (1892), and Parcharidis (1885), in a comprehensive study (Katsouda, 2012). However, these proposals do not agree among themselves, and, most importantly, they have not been adopted by the Pontic community widely, see Section 4.2.1: Pontic has not been standardised so far.

Overall, Pontic Greek constitutes a representative case of a low-resource dialect, where the combination of limited data availability, phonological and orthographic variability, and lack of standardisation poses significant challenges for the development of reliable ASR systems.

In this paper, where we present the first systematic evaluation of ASR models for the Pontic Greek language, we pay particular attention to transcription choices by consulting the community in an effort to better capture the phonetic characteristics of the dialect and improve model compatibility.

3. Related Work

Recent progress in ASR technologies has been driven by large-scale multilingual models capable of handling multiple languages with minimal task-specific adaptation. Prominent examples include Whisper (Radford et al., 2022) and XLS-R (Babu et al., 2022), which have demonstrated strong performance on a wide range of languages, including low-resource settings. These models benefit from training on large and diverse speech corpora, enabling robust generalisation and competitive zero-shot capabilities.

Dialectal ASR turns out to be rather challenging. Dialects are typically under-resourced. While turning to ASR systems of the related well-resourced varieties -typically of the standard variety- is the obvious choice, recent research suggests that dialects behave as distinct acoustic-linguistic domains rather than minor variants of a canonical form (Torgbi et al., 2025; Dhasmana et al., 2026). Likewise, the MADASR initiative highlights persistent performance gaps across closely related South Asian varieties (Singh et al., 2023).

In the case of SMG, for which sufficient data is available, several studies report relatively strong ASR performance. Fine-tuned variants of Whisper achieve WER in the range of 13-16%, while XLS-R-based models achieve comparable or slightly better results on benchmark datasets such as Common Voice (Ardila et al., 2020). Recent work leveraging weakly supervised data, such as large-scale podcast corpora, demonstrates that increasing data volume and domain diversity can significantly improve ASR performance for SMG (Paraskevopoulos et al., 2024).

However, performance degrades substantially with dialectal distance from the Standard variety. Vakirtzian et al. (2024) provide the first systematic evaluation of ASR models on multiple Greek dialects, including Aivaliot, Cretan, Griko, and Messenian. Their results show that state-of-the-art models that achieve low WER on SMG (11-14%) experience severe degradation with dialectal data: WER often exceeds 50% and even surpasses 100% in zero-shot settings. In the same line, a recently published expanded benchmark for low-resource Modern Greek dialects (Tsoukala et al., 2026) covers Aperathiot, Cretan, Lesbian, and Cappadocian, which are varieties with varying degrees of divergence from SMG. The benchmark provides transcriptions following SMG-based orthographic conventions, while preserving dialectal lexical and morphophonological forms. Zero-shot results with several ASR models reveal a clear performance gradient with dialectal distance from SMG, with best WERs ranging from about 60-70% for southern dialects, which are close to SMG, to over 80% for Lesbian and nearly 97% for Cappadocian. Fine-tuning substantially reduces error rates (up to 47% relative WER improvement), with Cappadocian remaining the most challenging variety (best WER 68.17%).

Despite the growing body of work on Greek dialects, Pontic Greek remains largely underexplored in the context of ASR, and there is currently no systematic evaluation of speech recognition models specifically targeting this variety. Nevertheless, evidence from closely related NLP tasks suggests that Pontic Greek poses significant challenges for modern language technologies. In particular, (Chatzikyriakidis et al., 2025) demonstrates that large language models (Llama-3-8B (Grattafiori et al., 2024), Llama-3.1-8B, Krikri-8B (Roussis et al., 2025)) exhibit substantially degraded performance on dialectal Greek, with Pontic Greek being among the most difficult cases. The authors further show that targeted fine-tuning on dialectal data leads to notable improvements, highlighting both the limitations of pretrained models and the importance of dialect-specific adaptation. Given the shared challenges across speech and text modalities, such as lexical variation, non-standard orthography, and scarcity of annotated speech data, these findings provide strong indirect evidence that ASR systems are also likely to struggle with Pontic Greek. This gap in the literature underscores the need for dedicated evaluation and modeling efforts for this dialect.

4. Methodology

The recordings were carefully collected, transcribed, and annotated to create a consistent and reliable dataset for ASR evaluation. Manual transcription ensured orthographic standardisation and precise time-aligned segmentation of the audio. The resulting dataset was then used to evaluate ASR performance using the Whisper and XLS-R models, with WER and CER as evaluation metrics. The following subsections describe the decisions and procedures employed to ensure accuracy, consistency, and fidelity to the Pontic Greek dialect.

4.1. Data Collection

Three native speakers of the Pontic dialect participated in the audio recordings, namely two females and one male, aged between 70 and 80. Originally from Neokaisareia (currently known as Niksar), the speakers were raised in Kallithea and Lofos, two villages in the Elassona region, and share a common variety of the Pontic dialect. This variety reflects the spoken form currently spoken by Pontic communities in different regions of Greece, mainly in Northern Greece, and does not correspond to historical or regional variants such as Ofitic.

The recordings were designed to capture spontaneous, natural speech through informal discussion. Participants were prompted to speak about topics related to everyday life, cultural memories, traditional food, local customs, games, and schooling. The aim was to elicit authentic dialectal usage in a conversational context, producing speech that reflects the natural patterns, vocabulary, and pronunciation of the Pontic dialect as it is spoken today.

The total recording duration is 1 hour, 26 minutes, and 38 seconds. It was conducted in Katerini, Greece, where the speakers currently reside, in a quiet indoor environment with minimal background noise to ensure high audio quality. This controlled setting allowed for clear capture of speech, facilitating subsequent transcription, segmentation, and annotation for the creation of a reliable and representative dataset for ASR evaluation.

4.2. Transcription & Annotation

This section describes the methods used to transcribe, segment, and annotate the recordings, as well as how the evaluation of models on Pontic Greek was conducted. It provides details on the orthographic decisions, manual transcription process, and time-aligned segmentation that together produced a consistent and reliable dataset for ASR evaluation.

4.2.1. Orthographic Decisions & Community Validation

Due to the lack of a standardized orthography, written representation practices vary considerably among speakers of Pontic, rendering the establishment of consistent and reliable reference transcriptions for ASR evaluation challenging. Our methodology prioritised developing a consistent set of transcriptions while minimising annotator bias in orthographic choices.

To inform the transcription strategy, we designed a written form questionnaire that targeted native speakers of the Pontic dialect. Ten participants aged between 40 and 80, with varying levels of daily dialect use (3/10 speakers speak the dialect daily, 4/10 occasionally, and 3/10 rarely), contributed written responses. The questionnaire included isolated lexical items, focusing on key phonological phenomena that would pose a problem in transcription. Participants were instructed to write words as they would naturally represent them, without looking them up in dictionaries or using copy and paste; the aim was to obtain evidence that captures authentic variation in orthographic practices.

The analysis of the responses identified recurring spelling patterns and informed the orthographic strategy for the transcription of the spoken data. The consonants "/ʃ/", "/pʃ/", "/kʃ/", and "/tʃ/", and the vowel "/æ/" at the ends of words, where the phenomenon of synizesis is observed are not represented in the Greek alphabet. Their representation in the proposed Pontic orthographies is done with Greek letters plus diacritics. Table 1 shows these representations and the responses of the speakers of Pontic. Regarding the phoneme /ʃ/ (e.g. χέρι=hand), the responses of speakers of Pontic were divided between renditions with "σ" and "χ", with variations such as "σσιέρ", "χεριν", and "σshέρι". Although there was no clear majority among the speakers, most researchers agree on the use of "χ" (Katsouda, 2012). For this reason, the spelling "χ" was chosen to ensure consistency with theoretical guidance and maintain scientific validity. /pʃ/ (e.g. ψυχή=soul) and /tʃ/ (e.g. καλατσεύω=speak) were represented by the majority of speakers with "ψ" and "τσ", with minor variations such as "ψι", "ψύ", "ψυη", or "καλατζευω" and "καλατshεύω". Regarding /kʃ/ (e.g. ρίχνω=throw), speaker responses varied, partly due to influence from SMG which uses the verb forms ρίχναμε, ερίχναμε; as a result, only five interesting spelling choices were obtained. Among these, the majority favored "ξ" over variants such as "έκχ̆υναμε" and "εκσσσιναμε". Finally, although the vowel /æ/ (e.g. ποτήρια=glasses) cannot be represented with the Greek alphabet, speaker responses showed strong agreement in rendering it as "ε".

These orthographic decisions favored the most frequently attested forms, adopted simplified representations, which are all characters of the Greek alphabet, and eliminated the use of diacritics. This approach provided an internally consistent reference set suitable for evaluating ASR models while reflecting the authentic linguistic patterns of Pontic Greek.

4.2.2. Transcription, Segmentation and Annotation

The transcription of the Pontic Greek recordings was performed semi-automatically, combining automatic transcription with manual post-editing to ensure accuracy and consistent application of the orthographic conventions. Whisper large-v2 was

IPA	Prompt	Responses (10)	Decision	Papadopoulos 1955/58	Parcharidis 1885	Valavanis 1892	Oikonomidis 1908	Oikonomidis 1958	Symeonidis–Tombaidis 1999
/ʃ/	e.g. χέρι (hand)	σιέρ (2), σεερ (1), χ̆έρ′ (1), χερα (1), σσσερ (1), σshέρι (1), χεριν (1), χέρι (1), σσερ (1)	χ	$\check{\chi}$	$\dot{\chi}$	χ	$\dot{\chi}$	$\hat{\chi}$	$\check{\chi}$
/pʃ/	e.g. ψυχή (soul)	ψυ (2), πσι (1), ψή (1), πσσσσυ (1), ψυη (1), ψύ (1), ψshιν (1), ψυμ (1), ψυν (1)	ψ	$\check{\psi}$	$\dot{\psi}$	ψ'	$\dot{\psi}$, $\psi'\dot{\chi}$	$\hat{\psi}$, $\psi'\hat{\chi}$	$\check{\psi}$
/tʃ/	e.g. καλατσεύω (speak)	καλατσευω (3), καλατσέυω (2), λαλώ (1), καλατσσσεύω (1), καλατσέβω (1), καλατσshεύω (1), καλατζευω (1)	τσ	τσ̆	$\dot{\tau}$, $\dot{\kappa}$	τσ̇, τσ́	τσ̇	τσ̂	τσ̆
/kʃ/	e.g. ρίχνω (throw)	έκχ̆υναμε (1), εκσσσιναμε (1), έξιναμ (1), έξυναμε (1), εξύγαμεν (1)	ξ	$\check{\xi}$	ξ̣	ξ'	$\dot{\xi}$, $\xi'\dot{\chi}$	$\hat{\xi}$	$\check{\xi}$
/ae/	e.g. ποτήρια (glasses)	ποτήρια (2), ποτήρε (2), κενό (2), ποτηρε (1), ποτυρε (1), ποτήρεα (1), ποτίρε (1)	ε	α̤	$^{\varepsilon}\alpha$	$i\alpha$	α̈	α̈	α̤

Table 1: Comparison of consonant and vowel representations across Pontic Greek orthographic systems. (Katsouda, 2012)

used to generate rough transcriptions as an initialisation step. Automatically-generated transcriptions were carefully reviewed and corrected by a linguist who is a speaker of the Pontic dialect, ensuring adherence to the spoken language while following the orthographic guidelines established through the community validation step.

Audio segmentation divided the recordings into individual utterances with the Praat tool (Boersma and Weenink, 2022). Each utterance was aligned with its corresponding audio segment to provide accurate timing information. The transcriptions and their time-aligned segments were stored in `TextGrid` files, which labeled each utterance with start and end times. From the original 1 hour, 26 minutes, and 38 seconds of recording time 39 minutes and 58 seconds of speech free from significant background noise were obtained.

The combination of manual verification, Whisper-assisted initialisation, and precise segmentation resulted in a high-quality dataset suitable for evaluating ASR models while maintaining linguistic and orthographic accuracy.

4.3. Corpus Construction and ASR Evaluation

The manual transcription and time-alignment process described above resulted in a set of Praat `TextGrid` files, where each utterance is associated with start/end timestamps and an orthographic transcription in the Greek script. To enable ASR benchmarking, we converted these annotations into a standard speech-corpus format consisting of audio–text pairs. The resulting corpus is thus an utterance-segmented version of the original recording, aligned to the dialectal transcriptions.

Preprocessing. To ensure consistency with common ASR benchmarking practices for Greek dialectal data, we applied lightweight preprocessing similar to prior work on Modern Greek dialect ASR evaluation resources (Vakirtzian et al., 2024)[2]. Audio was converted to mono wav at 16 kHz. On the text side, we lowercased and removed punctuation. Note that we evaluate dialectal transcription rather than dialect-to-standard normalisation.

Segmentation and filtering. Utterance boundaries were defined in Praat and exported from the TextGrid annotations. During this step, non-speech material (e.g., long pauses and other non-linguistic intervals) was excluded. The resulting segments constitute the evaluation units used for all ASR runs.

Models and inference-only setup. Given the small size of the corpus (single recording session, limited speaker and domain diversity), we report inference-only (zero-shot) performance without dialect-specific adaptation, to avoid overfitting and to quantify the out-of-the-box robustness of current ASR models on Pontic Greek. We evaluated two families of pretrained ASR models without dialect-specific adaptation: (i) Whisper-based (Radford et al., 2022) multilingual sequence-to-sequence models (Large-v2, Large-v3, Medium), and (ii) a Greek-adapted XLS-R (Babu et al., 2022) model[3] (wav2vec2/CTC-style).

All models were run on the utterance-segmented corpus created from the `TextGrid` annotations,

[2] https://github.com/athena-ilsp/greek-dialects-asr

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

producing a hypothesis transcription per segment. For Whisper, we used standard transcription inference with the language set to Greek (language="el") and no user-supplied initial prompt. No external language model or additional decoding constraints were used. For evaluation, both references and hypotheses were lowercased and stripped of punctuation; no further text normalization was applied. We computed Word Error Rate (WER) and Character Error Rate (CER) on the normalised text (lowercased, punctuation removed). Reporting both WER and CER is particularly useful for dialects, where orthographic variability and dialect-specific phonological realisations may disproportionately impact word-level errors, while character-level scoring can better reflect partial matches and subword robustness.

Model	WER (%)	CER (%)
Whisper large-v2	91.24	52.77
Whisper large-v3	**86.59**	**47.72**
Whisper medium	105.47	65.41
XLS-R-53-greek	111.70	98.88

Table 2: Inference-only ASR results on the Pontic Greek corpus.

5. Experimental Results

Table 2 reports inference-only ASR performance on the Pontic Greek corpus. Overall, zero-shot recognition is challenging: WERs are high across all models, and in some cases exceed 100%, indicating that the number of edit operations (substitutions, insertions, deletions) surpasses the number of reference words. These findings are consistent with prior work on Greek dialectal ASR evaluations (Vakirtzian et al., 2024; Tsoukala et al., 2026), where zero-shot evaluations on Cappadocian and Griko similarly yielded WER values above 100%, underscoring the broader difficulty of recognizing non-standard varieties without dialect-specific adaptation.

The inference-only results indicate that the Whisper-based models outperform the Greek-adapted XLS-R model on the Pontic Greek corpus. Among the Whisper models, large-v3 achieves the best performance, with a WER of 86.59% and a CER of 47.72%, followed by large-v2. The medium-sized Whisper model exhibits substantially higher error rates, suggesting that model capacity strongly impacts zero-shot recognition quality for this low-resource dialect.

The Greek-adapted XLS-R model performs poorly in this zero-shot setting (WER 111.70%, CER 98.88%), indicating limited generalization to Pontic speech despite adaptation to SMG. The very high CER suggests that the model frequently fails to recover even approximate character sequences under the combined acoustic, lexical, and orthographic mismatch induced by the Pontic dialect.

Taken together, these results establish strong zero-shot baselines and indicate that Pontic Greek is difficult for current general-purpose ASR. By releasing an utterance-segmented and time-aligned Pontic speech corpus with consistent transcriptions, we aim to support reproducible benchmarking and to enable future work on dialect-specific adaptation and evaluation protocols.

6. Discussion

It was already known that Pontic Greek is dialectologically distant from SMG as compared to other Modern Greek dialects. This fact is demonstrated in the ASR inference-only results, for instance, Pontic results match those of Cappadocian, another Modern Greek dialect of Asia Minor that is quite distant from SMG dialectologically (Tsoukala et al., 2026). These results underline the need for more Pontic data as well as the application of a variety of techniques to develop ASR models for the Pontic dialect.

Another interesting point touched upon in this work is the orthography issue. Despite the small number of native speakers who responded to a first simple questionnaire about the orthography of the dialect, the tendency that emerged is clear: native speakers of unstandardised varieties do not adhere to detailed, more precise orthographies and tend to cling to the orthography of the Standard variety which is taught in school and has all the necessary technical support (of course, dialectal speakers speak both varieties, Pontic and the Standard). This is a tendency that should be investigated in more depth and, probably, taken seriously into account in the development of AI-oriented technology for Pontic.

7. Conclusion

We have presented the first evaluation of ASR technology on data from Pontic Greek, which is a living but endangered dialect of Modern Greek. Although Pontic Greek is spoken by a community of about 400K people, it is considered endangered as transmission to newer generations has largely ceased. Our results show that substantial data requirements and the adoption of techniques developed for low-resource languages are necessary to achieve high-quality ASR performance for Pontic.

Nevertheless, this study is the first of its kind and establishes an initial benchmark for the development and evaluation of ASR technologies that may contribute to the vitality of this Modern Greek dialect.

Future work is planned towards expanding the dataset by incorporating the speech of a larger number of native speakers of the dialect. In parallel, we plan to conduct a large-scale survey on orthographic preferences and to experimentally evaluate the effects of different orthographic conventions on transcription consistency and downstream tasks.

8. Limitations

We are aware that the body of the data is small, which has dictated inference-only experiments. In addition, the number of speakers contacted for othography issues is small as well as the range of phenomena discussed.

9. Ethical considerations

Speech data were collected with the speakers' written permission for recording and public release. The corpus is openly available under a Creative Commons license.

10. Acknowledgements

The work presented in this paper was partially supported by the Aristotle University of Thessaloniki, Greece (Grant ELKE-AUTh-13337). We would like to thank the anonymous reviewers for their valuable comments and constructive feedback.

References

Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4218–4222, Marseille, France. European Language Resources Association.

Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. 2022. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale. In *Proc. Interspeech*.

Paul Boersma and David Weenink. 2022. *Praat: doing phonetics by computer*. Version 6.2.06, retrieved 23 January 2022 from http://www.praat.org/.

Stergios Chatzikyriakidis, Dimitris Papadakis, Sevasti-Ioanna Papaioannou, and Erofili Psaltaki. 2025. Grdd+: An extended greek dialectal dataset with cross-architecture fine-tuning evaluation.

Akriti Dhasmana, Aarohi Srivastava, and David Chiang. 2026. Dialect matters: Cross-lingual asr transfer for low-resource indic language varieties.

Eleonora Dimela. 2013. Ta tourkika epithetika daneia stis dialektous tis mikras asias: mia proti prosegnisi. *Patras Working Papers in Linguistics*, 3:92–109.

Georges Drettas. 1997. *Aspects pontiques*. Association de recherches pluridisciplinaires, Paris.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2026. Ethnologue: Languages of the world.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, et al. 2024. The llama 3 herd of models.

Georgia Katsouda. 2012. Dialectal lexicography: review and issues. *Leksikografikon Deltion*, 26:133–159.

Dimosthenis Oikonomidis. 1958. *Grammatiki tis Ellinikis dialektou tou Pontou*. Academy of Athens, Athens.

Anthimos Papadopoulos. 1953. *Characteristika tis Pontiakis dialektou*, volume 18. Epitropi Pontiakokn Meleton.

Anthimos Papadopoulos. 1955. *Historiki grammatiki tis Pontikis dialektou*. Committee of Pontic Studies [Archeion Pontou, Supplement 1], Athens.

Anthimos Papadopoulos. 1958. *Historikon Lexikon tis Pontikis dialektou*. Committee of Pontic Studies [Archeion Pontou, Supplement 3], Athens.

Georgios Paraskevopoulos, Chara Tsoukala, Athanasios Katsamanis, and Vassilis Katsouros. 2024. The greek podcast corpus: Competitive speech models for low-resourced languages with weakly supervised data.

Ioannis Parcharidis. 1885. Glossologika. ta en chrisei stoicheia. *Astir tou Pontou*, 1:27–28.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision. arXiv:2212.04356.

Dimitris Roussis, Leon Voukoutis, Georgios Paraskevopoulos, Sokratis Sofianopoulos, Prokopis Prokopidis, Vassilis Papavasileiou, Athanasios Katsamanis, Stelios Piperidis, and Vassilis Katsouros. 2025. Krikri: Advancing open large language models for greek.

Abhayjeet Singh, Arjun Singh Mehta, Ashish Khuraishi K S, Deekshitha G, Gauri Date, Jai Nanavati, Jesuraja Bandekar, Karnalius Basumatary, Karthika P, Sandhya Badiger, Sathvik Udupa, Saurabh Kumar, Savitha, Prasanta Kumar Ghosh, Prashanthi V, Priyanka Pai, Raoul Nanavati, Rohan Saxena, Sai Praneeth Reddy Mora, and Srinivasa Raghavan. 2023. Model adaptation for asr in low-resource indian languages. *arXiv preprint arXiv:2307.07948*.

Charalambos Symeonidis and Dimitris Tombaidis. 1999. *I simerini dialektos tis Oukranias (periochis Marioupolis)*. Committee of Pontic Studies [Archeion Pontou, Supplement 20], Athens.

Dimitris Tombaidis. 1988. *I pontiaki dialektos*. Parartima periodikou "Archeion Pontou". Epitropi Pontiakon Meleton.

Melissa Torgbi, Andrew Clayman, Jordan J. Speight, and Harish Tayyar Madabushi. 2025. Adapting whisper for regional dialects: Enhancing public services for vulnerable populations in the United Kingdom. In *Proceedings of the 12th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 29–38, Abu Dhabi, UAE. Association for Computational Linguistics.

Chara Tsoukala, Stavros Bompolas, Antigoni Margariti, Konstantina Panagiotou, Maria Elisavet Plaiti, Nefeli Tzanakaki, Petros Karatsareas, Angela Ralli, Antonios Anastasopoulos, and Stella Markantonatou. 2026. Extending asr evaluation resources for modern greek dialects. In *Proceedings of the 13th Workshop on NLP for Similar Languages, Varieties and Dialects, Rabat, Morocco. Association for Computational Linguistics.*

UNESCO. 2010. *Atlas of the World's Languages in Danger*, 3rd edition. UNESCO Publishing, Paris.

Socrates Vakirtzian, Chara Tsoukala, Stavros Bompolas, Katerina Mouzou, Vivian Stamou, Georgios Paraskevopoulos, Antonios Dimakis, Stella Markantonatou, Angela Ralli, and Antonios Anastasopoulos. 2024. Speech recognition for greek dialects: A challenging benchmark. In *Proc. Interspeech 2024*.

Ioannis Valavanis. 1892. *Zonta mnimeia tis ana ton Ponton idiotikis. Tmema A', Lexilogion*. Typographeion Perri, Athens.