Self-supervised Learning for Formosan Speech Representation and Linguistic Phylogeny

Anonymous ACL submission

Abstract

Formosan languages, spoken by the indigenous peoples of Taiwan, have unique roles in reconstructing Proto-Austronesian Languages. This paper presents a real-world Formosan language speech dataset, including 144 hours-news footage of 16 Formosan languages. One merit of the dataset is to look into the relationships among Formosan languages *in vivo*. With the help of deep learning models, we could analyze the speech data without transcription. Specifically, we first train a language classifier based on XLSR-53 to classify the 16 Formosan languages with an accuracy of 88%. Then, we extract the speech vector representations learned from the model and compare them with 152 manually coded linguistic typological features. The comparison suggests that the speech vectors reflect the phonological and morphological aspects of Formosan languages. In addition, these linguistic features are used to construct linguistic phylogeny, and the resulting genealogical grouping corresponds with previous literature. To sum up, the dataset opens up possibilities to investigate the current real-world use of the Formosan language.

1 Introduction

Formosan languages refer to a group of languages spoken by the indigenous peoples of Taiwan regarding their geographic distribution, all of which are Austronesian languages. The 24 Formosan languages respectively belong to 9 subgroups, 8 of which are considered extinct, while the other 12 languages, listed in Table 1, are regarded as national languages of Taiwan.[1] Since most of these currently spoken Formosan languages are extremely fragile or even moribund, the revitalization of these languages must be actively taken into action.

From the perspective of historical linguistics, Formosan languages also stand out in their role in reconstructing Proto-Austronesian Languages (PAn). Blust (1984) proposes the so-called *pulse-pause scenario* of the Pacific settlement, in which the Austronesian speakers originated in Taiwan around 5,200 years ago and rapidly spread through the Pacific in a series of expansion pulses and settlement pauses. Past studies propose rich insights into the linguistic phylogeny of Formosan languages through careful analysis of language innovations. However, due to the difficulties of speech data collection and analysis, it is less clear how to approach the phylogeny problems with real-world data.

We present a real-world dataset of Formosan languages collected from daily news broadcasted over Taiwan's free-to-air channels. The paper is organized as follows: Section 2 introduces the collected speech corpus. This corpus includes news footage covering 16 Formosan languages and aims to provide a valuable source with which researchers study Austronesian. To demonstrate one principal value of the dataset, we investigate the relationships among Formosan languages with speech vectors extracted from a deep learning classifier. Section 3 first describes the language classifier and its implied language phylogeny, and Section 4 analyzes the speech vectors and compares the learned vectors with manually coded linguistic features. Related works are briefly introduced in Section 5, and Section 6 concludes our work.

2 Formosan Speech Corpus

The collected Formosan speech corpus aims to record the real-world usage of the 16 Formosan languages. The primary data source is from daily news broadcasted over Taiwan's free-to-air channels. We use a TV tuner connected to an outdoor antenna to record the news footage to digital files. We capture all 16 Formosan languages news provided by the Taiwan Indigenous Television (TITV)

[1] The Yami language, spoken by Tao people living in Lanyu (lit. Orchid Island) Township, Taitung Country, 46 kilometers southeast of Taiwan, is linguistically Malayo-Polynesian, but geographically Formosan.

Figure 1: Example news broadcast

channel. Newscasts are chosen for the availability of all Formosan languages and to reduce the variability of gathering different languages from different programs. Each program is approximately an hour in duration. The corpus comprises 144 hours of videos with 9 hours for each language's news. [2]

While the news videos serve as an abundant source of information, the interaction among the Formosan languages and Mandarin Chinese in the news provides a unique challenge Figure 1. Specifically, although the news is broadcasted with a given Formosan language, segments still use Mandarin Chinese. These segments are like press conferences or interviews where the most common languages are still Mandarin Chinese. The issue is further complicated because some footage is narrated by the anchor, so there are no consistent visual cues to differentiate the language used in a given video segment. In addition, the Formosan languages are under-resourced, and there are no automatic speech recognition or language identification tools readily available. However, to properly explore the Formosan language in the video, we must at least tag the language used in the segments.

We address the mixed language problem first with automatic preprocessing, with which we gather primitive data to train a language identification classifier. We first assume the anchor always uses (one of the 16) Formosan language, and multiple cues in the video frames indicate that the anchor is speaking. We use two sources of information to determine the frame is an *anchor frame*. The first source is facial recognition, and the second is the headline usually displayed at the lower part of the frame. We first identify the anchor's face from the first 20 seconds of the video. The anchor is introduced and accompanied by a title card showing its name. We use off-the-shelf face recognition and optical text recognition models to pair the faces and the anchor name. After identifying the anchor's face, we detect, in each frame, if the anchor appears along with a headline. From these two cues, we determine, in a five-second interval, whether the anchor is speaking in the specific segment. The automatic anchor detection results, and the number of different anchors appeared in the news of each language are shown in Table 1.

However, while the detected anchor frames are likely the Formosan languages segments, there will be considerable false negatives in this approach. Segments, where the anchor narrates the footage with Formosan, are inevitably missed with the algorithm mentioned above. Therefore, it is still preferable to identify the language with the speech data alone. The trained language classifier not only helps us identify the language, but it also, with the help of computational models, helps us to explore the representations of the underlying speech data.

3 Formosan Language Classification

3.1 Classifier Training

We train a Formosan language classifier based on the Wav2Vec (Baevski et al., 2020) model architecture and the pretrained weights of XLSR (Conneau et al., 2020). We used the XLSR model to take advantage of having already been pretrained on 53 different languages. Although these languages may be significantly different from the Formosan languages, it might be possible for the model to transfer the regularities across languages.

The training data is the anchor segments automatically identified in the preprocessing stage. Among the 144 hours of speech data, 790.58 minutes of audio data are included in the dataset. In addition to the 16 Formosan languages, we add a *other* category, which is randomly sampled from the not-anchor video segments. Finally, we split the dataset so that every language is still equally represented in the test data.

Language classification is fine-tuned on the pretrained XLSR model. The classifier is a fully-connected layer stacked upon the vector ouput of the Wav2Vec2 model. The parameters are optimized with Adam with learning rate warming up to a peak of 10^{-3} in the first 200 steps and decrease to 0 with a half-cycle cosine scheduling. The model

[2] The corpus will be released once the paper is accepted.

Language	Subgroup	Len. (hrs)	Anchor Footage
Amis	Eastern Formosan	9	35.8(1)
Atayal	Atayalic	9	41.0(2)
Bunun	Bunun	9	52.6(1)
Cou (Tsou)	Tsouic	9	69.8(3)
Hla'alua (Saaroa)	Tsouic	9	34.8(1)
Kanakanavu	Tsouic	9	37.1(1)
Kavalan	Eastern Formosan	9	59.3(1)
Paiwan	Paiwan	9	12.3 (1)
Pinuyumayan (Puyuma)	Puyuma	9	40.9(1)
Rukai	Rukai	9	42.8(3)
Sakizaya	Eastern Formosan	9	54.5(2)
Saysiyat	Northwest Formosan	9	43.0(4)
Seediq	Atayalic	9	47.0(2)
Thau (Thao)	Western Plains	9	44.2 (1)
Truku	Atayalic	9	73.5 (1)
Yami	Malayo-Polynesian	9	47.3 (1)

Table 1: Captured video length for each language. Anchor footage denotes the automatically detected anchor segments. The lengths are in minutes. These segments are more likely to only contain the targeted Formosan language. Numbers in the parentheses are the number of different anchors in the news footage. The subgroups of each language follow Blust (2013).

training took approximately two hours on a A5000 GPU.

The language classification model achieved an overall accuracy of 88% across 17 categories (16 languages and the *other* category) languages in the testing set. The classification accuracy shows that the model indeed can identify different Formosan languages. Notably, the anchor's identity is confounded with the language in this dataset. However, the overall classification results show that the languages with only one anchor do not necessarily have better performances than those with multiple anchors. That is, the anchor identities may not directly influence the classifier.

The classifier not only has the practical value in helping identify relevant segments in the dataset. In addition, the self-supervision nature of Wav2Vec2 provides us with a unique opportunity to explore how these languages are related to each other in this formalized vector space. When the model is only trained on the speech signal, it should especially shed light on the phonological or phonetic relationships among these languages.

3.2 Speech Vector Analysis

The fine-tuned language classification model successfully differentiates the language of a specific speech segment. In the speech representation perspective, it is interesting to explore the language

Language	Prec.	Recall	F1	N
Amis	0.97	0.54	0.69	140
Atayal	0.99	0.84	0.91	192
Bunun	0.83	0.99	0.90	115
Cou	0.93	0.99	0.96	154
Hla'alua	0.94	0.92	0.93	89
Kanakanavu	1.00	0.93	0.96	95
Kavalan	0.99	0.98	0.98	145
Paiwan	0.94	0.97	0.96	35
Pinuyumayan	0.65	0.98	0.78	54
Rukai	0.95	0.84	0.89	122
Sakizaya	0.98	0.96	0.97	155
Saysiyat	0.78	0.98	0.87	129
Seediq	0.86	0.78	0.82	102
Thau	0.87	0.99	0.93	110
Truku	0.77	0.99	0.86	205
Yami	1.00	0.93	0.96	215
Other	0.64	0.57	0.60	219

Table 2: Classification results for each of the 16 languages and the "other" category. The overall accuracy is .88, and the weighted average is .87.

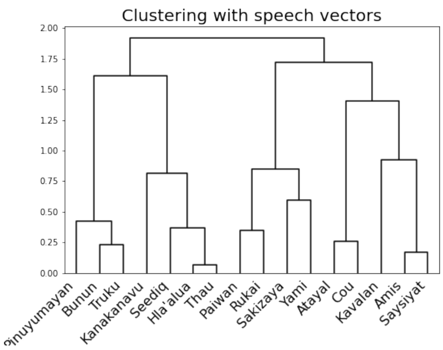

Figure 2: Clustering results with 16 language vectors.

similarities implied by these speech vectors. The idea is consistent with the findings from other domains of deep learning application that the model representation may reflect the intrinsic structure underlying the data. For example, the word analogy relations results naturally from the vector representation learned from skip-gram or CBOW model (Mikolov et al., 2013), and the transformer-based language model also implicitly reflect the syntactic relations in the sentence (Manning et al., 2020).

To compute the language similarities among these 16 languages, we first extract the speech vector representation of each segment from the Wav2Vec2 model; that is, the 1,024-dimension vector before it is fed into the final classifier. These 1,024 dimensional vectors are assumed to carry various information, and only some of which are the ones used in language classification. We thus simplify the vector with a linear dimension reduction model (i.e. PCA) into 5 dimensions. Next, we find the median points, or the medoids, in each language to represent the speech segments of that language. These median point are then clustered to show the structure implied by the model.

The clustering result is shown in Figure 2. One way to interpret the clustering is that it reflects a snapshot of the current linguistic environment. However, there are several possibilities that the model treats two languages as similar in these vectors: such as geographical closeness, phonetic, morphological, and syntactic relations. Therefore, we try to explore the representations of these speech vectors with a correlational similarity study. Next, we manually coded linguistic typological features of the 16 Formosan languages. These features include phonological, morphological, and syntactical ones. We then compare the language similarities implied by these typological features to the ones computed by the speech vectors.

4 Formosan Linguistic Phylogeny

The 'Austronesian homeland' hypothesis was supported by lexical data and other archaeological evidence (Greenhill et al., 2010). However, Dunn et al. (2005) showed that it is also possible to probe the linguistic phylogeny by using non-lexical grammatical traits/features. For instance, the phonetic features of 16 Formosan languages have been manifested in the aforementioned speech corpus data and Wav2Vec model. In the current section, the phonological features will also be adopted in our clustering analysis of Formosan linguistic phylogeny.

4.1 Features Coding

To explore the extent to which a feature system is minimally sufficient to distinguish all the sounds in Formosan languages, we follow Duanmu (2016) that (1) the number of features is small, (2) all features are binary (due to the notion of *contrast*), and (3) features can be compared across languages. Our dataset contains data from 16 taxa (i.e., languages, or tips of the phylogenetic tree) encoded with 152 typological features (manually encoded based on a series of Reference Grammar books by a group of prestigious Formosan linguists) (Wu et.al, 2016-18), including grammatical traits, such as word order (order of noun phrase elements and verb), pronominals, demonstratives, noun formation and verb formation, numerals and the counting system, adjective, syntactic roles of noun phrases, the verb complex, TAM (tense, aspect, mood), core and oblique participants, as well as phonological ones, such as voicing, places and manners of articulation, etc.[3]

4.2 Language Features and Speech Vectors

The coded language features imply language similarities among Formosan languages, which we can compare to those implied by speech vectors. The comparison also sheds light on the nature of representations learned automatically with the deep learning model. Specifically, suppose the language similarities are consistent with a set of language features, e.g., phonological ones. In that case, we could infer that the learned speech vector representations encode phonological aspects of those languages.

[3]The data are attached to the paper.

We first partition 152 features into three categories: phonological, morphological, and syntactical features. Features all coded as ones and zeros are excluded from further analysis. There are 120 features included in this analysis, 56 phonological ones, 43 are morphological, and 21 are syntactical. Among the phonological features, we further distinguish 10 vowel-related features, 46 consonant-related ones, 22 sonorants, and 31 obstruent features. Note that not all phonological features could be classified as sonorant or obstruent, such as syllable-level features (e.g., phonemic stress or consonant clusters). For each feature group, we constructed a correlation matrix from the feature encoding. As a result, eight language correlation matrices are made from eight feature groups, respectively.

We compare the language feature-derived correlation matrices and the speech vector-derived matrices with Spearman's rank correlation coefficients (Spearman's r). Specifically, the lower triangles of each correlation matrix are extracted and flattened as vectors, from which we computed Spearman's r. However, as the data vector comes from a correlation matrix, it is unclear whether the standard inferential statistics apply. Therefore, we bootstrap the speech vectors to infer a confidence interval. Each bootstrap sample comprises 50% of correctly classified sequences in each language. We computed the medoids (following the same procedure in Sec. 3.2) of each language, from which we derived the correlation matrix of this particular bootstrap sample. For each bootstrapped speech vector-derived correlation, we compute one Spearman's r with the language feature-derived correlation. From 100 bootstrapped samples, we calculate the mean, 5% (Q05) and 95% (Q95) quantiles of Spearman's r. The same bootstrapping procedures are repeated for the random feature controls, where values in each feature are randomly permuted. The goal of this permutation is to generate a random baseline where the language features provide no information on the language similarities.

Results are shown in Figure 3. First, the speech vector-derived language similarities are consistent with the one derived from language features, as seen by the non-overlapping confidence intervals computed from the actual samples and the random baseline. This pattern persists into the phonological feature groups. Most notably, the obstruent feature group shows the most significant difference,

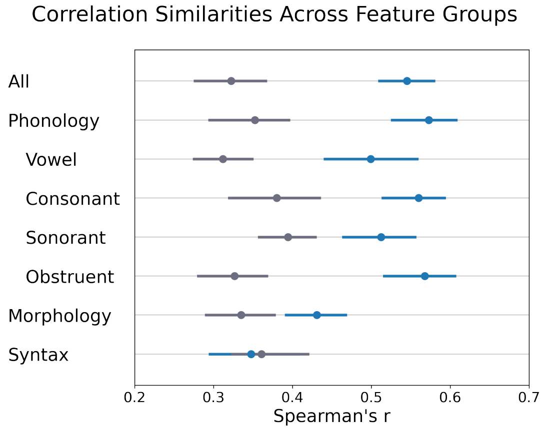

Figure 3: Correlation similarities across different feature groups. The blue segments show the similarities between the correlations of Formosan languages implied by the speech vectors and the one implied by feature groups. The line intervals indicate the bootstrapped confidence interval (Q05-Q95). The gray segments show the similarity scores under a random baseline.

.24, and the sonorant has the smallest ones, .12. Interestingly, the difference is still significant in the morphological group but not in the syntax group. These findings show that the model does capture language-relevant aspects from the audio stream, not just superficial acoustic features (e.g., anchors' voice characteristics). The significant difference in the morphological group also suggests that, while the data is speech only, it doesn't prevent the model from learning morphological information from the audio sequence. In contrast, the syntactic features do not play a role in speech vectors. Possible explanations may include the insufficient number of features in syntactic groups or the nature of language identification tasks that prevent the models from learning such long-ranged features.

The language feature analysis clearly shows that speech vectors encode phonological, even morphological aspects of Formosan languages. However, it is still unclear how these language features relate to the Formosan language similarities, or linguistic phylogenies, in the literature. Therefore, we use our coded linguistic features to proceed with the linguistic phylogenetic inferences.

4.3 Linguistic Phylogenetic Inferences

Comparison between speech vectors and linguistic features reveals significant similarities. It also shows the speech vectors, unsurprisingly, tend to capture the phonological aspects of languages. However, it is not clear whether the coded linguistic features really reflect, or are consistent with the Formosan phylogenies found in literature. Therefore, we construct the Formosan phylogeny from our linguistic features.

We used 61 phonological features in the following phylogenetic inferences. These phonological features account for most of our linguistic features and are the most significant ones in the correlational similarity study. In addition, as they relied more on phonological innovation to infer the subgroupings of Formosan languages, using phonological features provides a better comparison with past studies.

First, we consider *divisive clustering* based on the features, as shown in Figure 4. However, the dendrogram obtained does not fit well with previous reconstruction proposals (Blust, 2013; Li, 2006; Starosta, 1995), and we do not know how accurate and robust the phylogenetic estimates of Austronesian language relationships are. We then turn to a computational phylogenetic method called *neighbor-joining algorithm*) (NJ) (Saitou and Nei, 1987) to create a phylogenetic tree without a defined root. The unrooted tree has its advantage in not presuming information about the temporal sequence of lineage-splitting events. Figure 5 shows the resulting unrooted tree with NJ algorithm that is widely used for phylogeny estimation. The tree presented in the left panel shows that the unrooted phylogenetic tree groups languages according to their geographical region, indicated by different font styles (e.g., bold, italics).

To validate the results of our cluster analysis, the `BOOTSTRAP` method is applied to the present data. The data is sampled with replacement for 200 bootstrap runs. In each sampling run, the distance matrix is calculated to further yield the unrooted tree with the NJ algorithm. With the resulting dendrograms from the bootstrap samples, we compare them to the original one, and calculate the proportions of bootstrapped dendrograms that support the subtrees in the original tree. The proportion of support for different subtrees is shown in the middle panel with the sign of thermometers, of which the higher degree indicates greater support. The consensus tree, where the subgroups that are not observed in all bootstrap trees are collapsed, is shown in the right panel.[4]

[4]We use the `ape` package developed by (Paradis et al., 2004) to implement the calculations

5 Related Works

Studying language families has long been of high interest in historical linguistics. Among language families around the world, Austronesian, which contains more than 1,250 languages and spans across the Indian Ocean into the western Pacific, has been one family tracts significant research interest. The expansion origin of Austronesian is inevitably controversial. Nevertheless, past studies combine data both from linguistics and archaeologists and suggest the Formosan language has played a significant role in Austronesian expansion (Blust, 2019, 1999; Gray et al., 2009; Bellwood, 1984).

Being the origin of expansion, Formosan languages show great diversity. Studies of Formosan phylogeny follow the cladistic principles, where each tree node is supported by a language innovation, such as phonological, morphological, or basic numeral vocabulary (Blust, 1999; Ho, 1998; Sagart, 2004; Ross, 2012). Another approach to study the relationships among the languages is from structural similarities. These structural features are abstract and were selected to reflect the known linguistic topology in the region. Genealogical groupings are then constructed by computational algorithms, such as maximal parsimony, from their shared structural features (Dunn et al., 2005). However, the structural features are abstract, and not all structural features are equally prominent in actual usage. Therefore, the similarities implied by the structural features may not directly reflect the similarities in real-world use.

The recent speech recognition model allows us to work with natural speech data without directly transcribing it. This approach opens up the possibility of looking into the real-world usage of Formosan language and studying them with a systematic methodology. Hartmann (2019) uses deep neural networks to reconstruct the phonetic features of historical sounds based on a language's synchronic phonological features, such as co-articulatory and phonological constraints. Korkut et al. (2020) compare several deep learning methods for spoken language identification. The authors use a hybrid CNN-RNN (CRNNs), X-vectors with FFNNs, and Wav2Vec CNNs (Schneider et al., 2019) in a language classification task. They also find that the X-vector-based FFNN classifier outperforms the other two models. They also learn that SpecAugment is suitable for language identification data

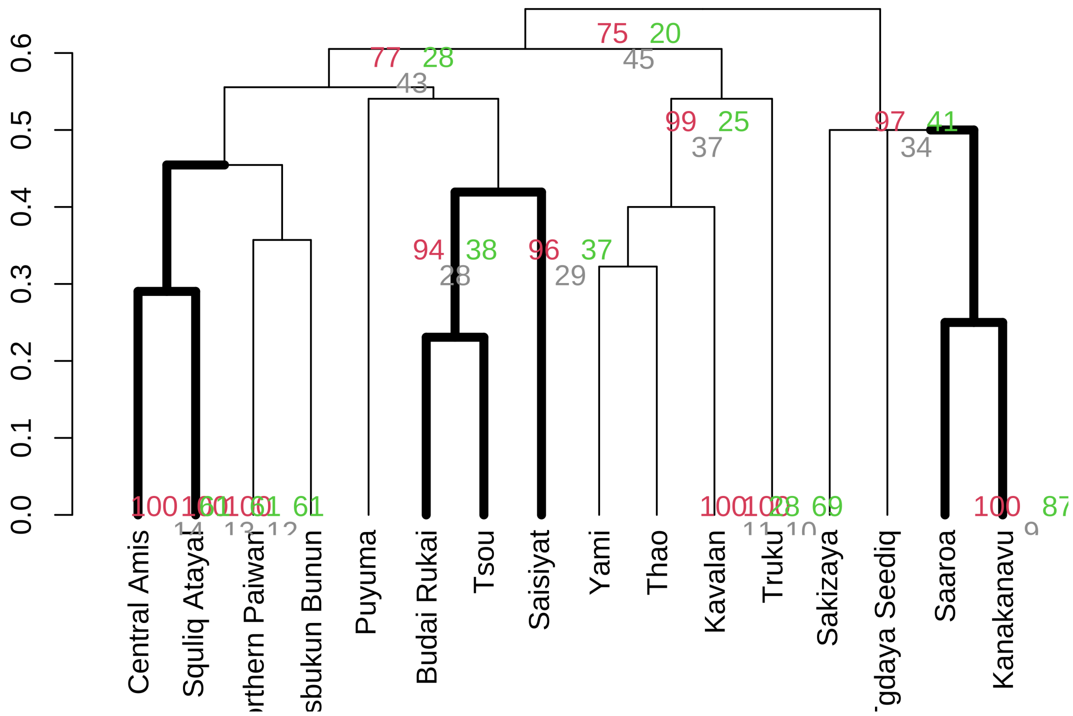

Figure 4: Dendrogram with AU/BP values (%) of divisive hierarchical clustering of 61 phonological features for Austronesian languages in Taiwan. Red: AU (approximately unbiased) p-value; green: BP (bootstrap probability) p-value; gray: SI (Selective inference) p-value.

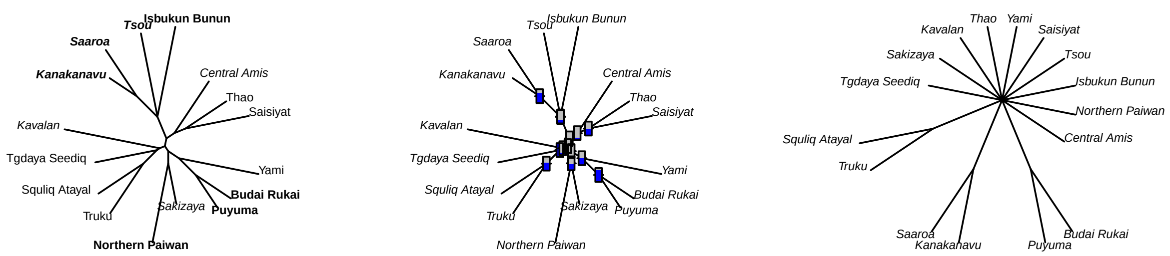

Figure 5: Unrooted phylogenetic trees for Austronesian languages in Taiwan
(Left panel) The geographical information is represented by fonts. Plain: Northern; bold: Southern; italics: Eastern; bold italics: Central/Tsouic. (Middle panel) Validation of clustering using the bootstrap. (Right panel) Consensus tree from 200 bootstrap runs.

augmentation. In this study, we leverage speech vectors, learned by a language identification model, to study the relationships among the Formosan languages.

6 Conclusion

In this paper, we present a Formosan Speech Corpus. We provide two perspectives on Formosan phylogeny studies based on the dataset: a speech vector approach using a Wav2Vec-based deep learning model and linguistic coding with linguistic typological features. The speech vector approach is more data-driven, and more emphasized on the usage aspect of speech data. The speech representation is trained to achieve a language classification task of 16 Formosan languages with 144 hours of speech data collected from the news broadcast. The model achieves overall classification accuracy of 88%. Moreover, correlational similarities analysis shows the speech vector representations reflect the phonological and morphological information. A further look into the typological language features reveals phylogenetic trees correspond well with previous theories.

Overall, this paper tries to approach the Formosan language similarities guided by model-learned representation from real-world data and linguistic typological features. Future works include how to interpret the language similarities implied by the speech vectors and further explore the multimodal nature of the dataset. This paper, along with its dataset, is expected to help investigate the linguistic phylogeny simultaneously with the actual usage patterns in the current language environment.

References

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems*, volume 33, pages 12449–12460. Curran Associates, Inc.

Peter Bellwood. 1984. A hypothesis for austronesian origins. *Asian Perspectives*, 26(1):107–117.

Robert Blust. 1984. The austronesian homeland: a linguistic perspective. *Asian Perspectives*, 26(1):45–67.

Robert Blust. 1999. Subgrouping, circularity and extinction: some issues in austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, volume 3, pages 1–94.

Robert Blust. 2013. *The Austronesian languages (Revised Edition)*. Australian National University.

Robert Blust. 2019. The austronesian homeland and dispersal. *Annual Review of Linguistics*, 5(1):417–434.

Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2020. Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.

San Duanmu. 2016. *A theory of phonological features*. Oxford University Press.

Michael Dunn, Angela Terrill, Ger Reesink, Robert A Foley, and Stephen C Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science*, 309(5743):2072–2075.

R. D. Gray, A. J. Drummond, and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science*, 323(5913):479–483.

Simon J Greenhill, Alexei J Drummond, and Russell D Gray. 2010. How accurate and robust are the phylogenetic estimates of austronesian language relationships? *PloS one*, 5(3):e9573.

Frederik Hartmann. 2019. Predicting historical phonetic features using deep neural networks: A case study of the phonetic system of proto-indo-european. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 98–108.

D.A. Ho. 1998. Taiwan nandaoyu de yuyan guanxi [genetic relationships among the formosan languages]. *Chinese Studies*, 16(2):141–171.

Can Korkut, Ali Haznedaroglu, and Levent Arslan. 2020. Comparison of deep learning methods for spoken language identification. In *Speech and Computer*, pages 223–231. Springer International Publishing.

Paul Jen-kuei Li. 2006. The internal relationships of formosan languages. In *Tenth International Conference on Austronesian linguistics (10-ICAL), Puerto Princesa, Palawan, Philippines*, pages 17–20. Citeseer.

Christopher D. Manning, Kevin Clark, John Hewitt, Urvashi Khandelwal, and Omer Levy. 2020. Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 117(48):30046–30054.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. 2004. Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290.

Malcolm Ross. 2012. In defense of nuclear austronesian (and against tsouic). *Language and Linguistics*, 13(6):1253–1330.

Laurent Sagart. 2004. The higher phylogeny of austronesian and the position of tai-kadai. *Oceanic Linguistics*, 43(2):411–444.

Naruya Saitou and Masatoshi Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425.

Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. 2019. wav2vec: Unsupervised pre-training for speech recognition. *arXiv preprint arXiv:1904.05862*.

Stanley Starosta. 1995. A grammatical subgrouping of formosan languages. *Austronesian studies relating to Taiwan*, pages 683–726.

Joy Jing-Lan Wu et.al. 2016-18. *Táiwān nándǎo yǔyán cóngshū 1-16 [Formosan languages Reference Grammar book series 1-16].* Council of Indigenous Peoples, Executive Yuan. Taiwan.