Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1061–1067
Marseille, 20-25 June 2022

Phone Inventories and Recognition for Every Language

Xinjian Li, Florian Metze, David R. Mortensen, Alan W. Black, Shinji Watanabe
Carnegie Mellon University
Language Technologies Institute
5000 Forbes Ave, Pittsburgh, PA 15213 US
{xinjianl, fmetze, dmortens, awb, swatanab}@andrew.cmu.edu

Abstract

Identifying phone inventories is a crucial component in language documentation and the preservation of endangered languages. However, even the largest collection of phone inventory only covers about 2000 languages, which is only 1/4 of the total number of languages in the world. A majority of the remaining languages are endangered. In this work, we attempt to solve this problem by estimating the phone inventory for any language listed in Glottolog, which contains phylogenetic information regarding 8000 languages. In particular, we propose one probabilistic model and one non-probabilistic model, both using phylogenetic trees ("language family trees") to measure the distance between languages. We show that our best model outperforms baseline models by 6.5 F1. Furthermore, we demonstrate that, with the proposed inventories, the phone recognition model can be customized for every language in the set, which improved the PER (phone error rate) in phone recognition by 25%.

Keywords: Phone Inventory, Multilingual Speech Recognition, Endangered Languages, Language Documentation

1. Introduction

A fundamental aspect of the description or documentation of any language is establishing its phone inventory (Bird and Simons, 2003; Michaud et al., 2018). This is a necessary prerequisite to further phonetic and phonological analysis (including transcribing text, discovering allophonic patterns, and developing an orthography), these are foundations upon which other facets of linguistic description can be built. Traditionally, phone inventories have been discovered by field linguists using a mixture of audio, visual, and lexical tools to arrive at a set of sounds sufficient to characterize the phonetics of the language. The largest collection of phone inventory aggregated so far is the PHOIBLE dataset (Moran, Steven and McCloy, Daniel and Wright, Richard, 2014), which is a collection of phone inventories from over 2000 languages. However, there are around 8000 languages in the world, for most of which no documented phone inventory exists. Unfortunately, those languages are typically endangered (Nettle et al., 2000). Language preservation projects typically target languages in this category. Field linguists starting work on a new language will benefit from knowing, in approximate terms, what the phone inventory of that language is like.

In this work, we attempt to solve this problem by estimating the phone inventory for any language listed in Glottolog (Nordhoff and Hammarström, 2011), which contains around 8000 languages. In particular, we take advantage of the phylogenetic trees from Glottolog (as this information is available for almost every language in the world)[1]. We propose two approaches that exploit this tree structure: First, we impose a probabilistic structure on the phylogenetic tree ("language family tree"), where each child node is expected to have a similar phone distribution to its parent. Next, we introduce *nearest language ensemble* approach, in which we compute the nearest neighbor languages for any unseen target language and we ensemble the phone inventory from those nearest languages as the inventory for the target language. Note there are other features such as geographical coordinates to derive closeness between languages. These features, however, are not easy to model for non-leaf nodes in our approach (e.g:

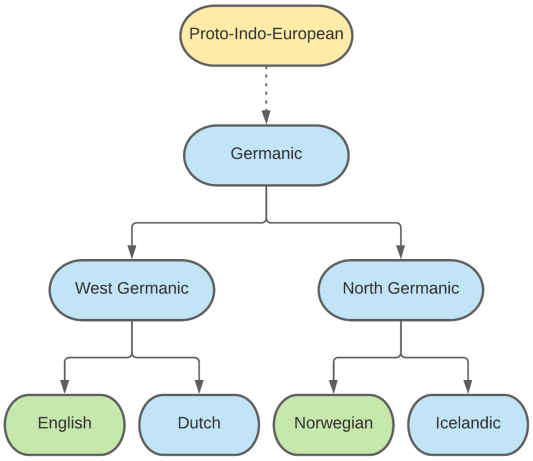

Figure 1: Illustration of a branch sample from the Germanic branch a phylogenetic tree. We derive the testing inventory for Dutch and Icelandic using the training inventory from English and Norwegian

[1]Since linguists do not always agree upon the phylogenetic groupings of languages—especially of poorly-studied languages—the trees from Glottolog are necessarily imperfect. However, they usually represent state-of-the-art classifications and are thus useful for our experiments here.

it might not make sense to assign a specific coordinate to the Indo-European family node). As a result, we only consider the simple tree structure in this work. We apply our approach to 77 languages, whose inventories are excluded from our training set. This experiment shows that our approach achieves an F1 score of 65.9, which is 6.5 points better than the best baseline model. Finally, we demonstrate that, using the proposed phone inventories, we enable a recently proposed phone recognizer to recognize all 8000 languages (Li et al., 2021a). Our results show that with the hypothesized phone inventories, we achieve 64.2% PER (phone error rate), which is 25% better than the original model (Li et al., 2020). To the best of our knowledge, this is the first speech recognition system that has been successfully customized for almost every known language known to comparative linguistics.

2. Related Work

Compiling the phonemic/phonetic inventory for a single target language is typically an important task in phonetic and phonological analysis (Hayes, 2011). However, not all languages in the world are equally well-researched. For example, much phonological research has focused on richly resourced languages (therefore they usually have well-defined phone inventories (International Phonetic Association et al., 1999)), while other, low-resource languages have historically received less attention. Recently, there have been several unsupervised models proposed that are meant to discover linguistic units for unwritten languages (Varadarajan et al., 2008; Müller et al., 2017; Dunbar et al., 2019; Dunbar et al., 2020), those models typically require the raw speech recordings for discovery, whose resources are limited for most languages (Black, 2019).

While most traditional phonetic research has been focused on a single language or a few languages, there have been several attempts to compile large databases to collect many phone inventories of a diversity of languages. PHOIBLE (PHOnetics Information Base and Lexicon) is a phonological inventory database which contains inventory information of more than 2000 distinct languages (Moran and McCloy, 2019), each phone also has been assigned distinctive phonological features (Jakobson et al., 1951; Chomsky and Halle, 1968). Another large database compiled by Merritt Ruhlen is the Ruhlen Database (Creanza et al., 2015). It contains not only the phonological information for each language, but also a wealth of extra-linguistic information (e.g: number of speakers and the geographical location of each language). While both projects have successfully collected many sound inventories, the majority of the inventories of the world's languages remain undocumented. To address this problem, this work attempts to give a reasonable approximation of each phone inventory for every language registered in Glottolog.

3. Approach

In this section, we introduce our two proposed approaches: Bayesian Network Estimation and Nearest Languages Ensemble. Before that, though, we propose two baselines and setup notations used in this work.

3.1. Baseline

Assume a set of training languages is L. For every training language $l \in L$, we have access to its phone inventory Σ_l. The simplest inventory estimation model uses the inventory $\hat{\Sigma}_{\text{fixed}}$ from a fixed language, for example Tagalog: $\Sigma_{\text{<tgl>}}$. This is because Tagalog's inventory has a reputation for typicality. Note that not every well-known language can be a good baseline. The English inventory $\Sigma_{\text{<eng>}}$, for example, is atypical: it includes some very rare phones like [θ] and [ð] but lacks (depending on analysis and dialect) some very common phones like [a], [e], and [o]. This *Fixed Inventory*, however, only covers phones from a single language; therefore it fails to include common phones in other languages and has low recall. Another possible baseline would be to the use the entire phone inventory available from all training languages:

$$\Sigma = \bigcup_{l \in L} \Sigma_l \tag{1}$$

This is a default inventory used in some phone recognition works (Li et al., 2021a; Li et al., 2020). This naïve approach should improve recall but it includes far more phones that any individual language and most of them are, invariably, false positives. To improve the precision, we sort all phones by the number of times they appear in our training languages and only keep the top-n most frequent phones based on the following statistics. This inventory baseline is the *Global Inventory* $\hat{\Sigma}_{\text{global}}$

$$\sum_{l \in L} \mathbb{1}([p] \in \Sigma_l) \tag{2}$$

3.2. Bayesian Network Estimation

The global inventory reflects the overall trend of phones across all languages, but it does not capture the local similarity between languages. We propose to exploit a phylogenetic tree to capture the local relations between languages (based on the insight that languages that are phylogenetically close also have similar phone inventories). Our first model is to impose a probabilistic structure to the tree. In particular, we consider the tree to be a *Bayesian Network* (i.e: a directed probabilistic graphical model). For each node in the tree, a multinomial distribution over the entire inventory Σ is assigned. We assume that the inventory of the child node is drawn from its parent's multinomial distribution. Formally, suppose we have a parent node r and its child l where the child l is one of our training languages. We can model the probability of drawing the child inventory using r's multinomial distribution:

$$\text{Prob}(\Sigma_l|\Theta_r) = \frac{|\Sigma_l|!}{\prod_i(x_i!)} \prod_{i \in \Sigma_l} \theta_i^{x_i} \quad (3)$$

where $\Theta_r = \{\theta_1^r, ..., \theta_\Sigma^r\}$ is the parameter of parent node r, and each parameter θ_i^r is the probability to draw the i-th phone from all available phones Σ, and x_i is the indicator function whether the i-th phone is contained in the child l's inventory. The parameter Θ_r can be inferred using *Maxmimum Likelihood Estimation* (MLE). After obtaining the parameter Θ_r of the parent node r, we could construct the phone inventory $\hat{\Sigma}_{\text{bayes}}$ for the parent node by selecting phones with the top-n highest probability. This is equivalent to selecting the top-n phones which have the highest counts in children of r. The counting can be computed as follows:

$$\sum_{l \in \text{Children(r)}} \mathbb{1}([p] \in \Sigma_l) \quad (4)$$

3.3. Nearest Language Ensemble

The Bayesian Network model can infer parent's inventory using its children information, however, it cannot take advantage of information from other close nodes (e.g: sibling nodes). To fix this issue, our second model is to use the nearest languages to approximate the inventory of the target language. The metric to define distance between languages is the length of the shortest path between any two languages in the phylogenetic tree. The shortest path between any two language nodes can be efficiently computed with *Lowest Common Ancestor* (LCA) whose time complexity is $O(log(H))$ where H is the height of the phylogenetic tree (Cormen et al., 2009). For a target language, suppose we find the top-k nearest languages L_k, then we first count the appearance of each phone $[p]$:

$$\sum_{l \in L_k} \mathbb{1}([p] \in \Sigma_l) \quad (5)$$

Then we could select the top-n phones $\hat{\Sigma}_{\text{nearest}}$ using these counts. For example in Figure 1, suppose that our training languages are English and Norwegian, and we would like to estimate the inventory for Dutch, when we use the top-1 nearest language, only English would be selected and we could simply copy the English inventory to the Dutch inventory, when we use $k = 2$, we would identify English and Norwegian as the nearest languages, and average them using counts.

4. Universal Phone Recognition

The hypothesized phone inventories pave the way for many new applications. Notably, they allow us to create phone recognition systems for (almost) every language in the world. In this section, we first introduce the acoustic model we use in this work, then we explain how to apply the estimated inventory for the recognition task.

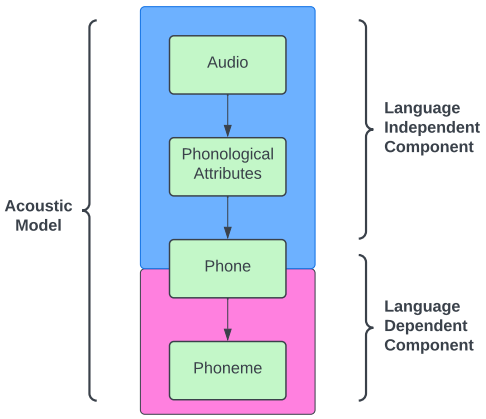

Figure 2: The architecture of the phone recognition model. We first compose the phone representations using their phonological attributes. Then we compute the phone distributions using the hidden vector from the encoder. Next, the language-independent phones are transformed into language-dependent phonemes with the allophone mappings.

4.1. Architecture

We closely follow the architecture described in the previous work (Li et al., 2021a): The architecture has a hierarchical structure which is illustrated in Figure 2. We model three different units explicitly: *phonemes*, *phones* and *phonological attributes*. Phonemes are typically language-dependent units, whereas phones are language-independent units. Phonological attributes or articulatory attributes are a set of discrete properties to characterize each phone. The set of phones corresponding to one phoneme in a particular language is called the *allophones* of the phoneme, which is annotated by phonologists. We use an annotated dataset to map between phone and phonemes (Mortensen et al., 2020). Similarly, each phone can also be decomposed into a set of attributes. The correspondence is also well-studied by linguists and we use tools to extract attributes for each phone (Mortensen et al., 2016). During the training process, the encoder would first encode each frame of the audio into a hidden vector, from which we could obtain the distributions of phones in each frame using their attributes. Each phone distribution is further transformed into phoneme distribution and optimized by the CTC loss function. In this work, the encoder is a 12-layer transformer-based encoder whose hidden size is 640 and multi-head attention size is 4 (Vaswani et al., 2017). The feature is the 40-dimension filter bank. We train the model using eng, cmn, deu, fra, ita, rus, tur, vie languages from the Common Voice corpus (Ardila et al., 2020). Our trained model is available online. [2]

[2]Interspeech21 model at `https://github.com/xinjli/allosaurus`

4.2. Inference

As the lower part in Figure 2 is language-independent, we can apply the trained model to any unseen languages whose inventory is accessible: if both phoneme and phone inventory are available, we can plug those inventories into the model and run the inference. If only the phone inventory is available, we approximate the phoneme set with its phone set, assuming each phone is mapped to the same phoneme.

Even the phone inventory, however, is not always available for every language. For languages whose phone inventory is absent, an approximated phone inventory should be used instead. In previous works, the inventory was chosen to be the global inventory $\hat{\Sigma}_{\text{global}}$: all the available training phones to make their prediction (Li et al., 2020; Li et al., 2021a). This naïve approach, however, has the low precision problem because the set of all training phones is too large. We demonstrate that, employing the hypothesized inventories $\hat{\Sigma}_{\text{bayes}}, \hat{\Sigma}_{\text{nearest}}$ introduced in the previous section, we can improve the phone recognition accuracy.

5. Experiments

In this section, we demonstrate our experimental results for both phone inventory evaluation and phone recognition. As mentioned in the previous section, we first build the phylogenetic tree using Glottolog (Nordhoff and Hammarström, 2011). The tree contains 7915 languages, where there are 43 top-level language families. We further create a root language node which possesses all top-level languages as its children, therefore all the languages are connected and can be reached from a single root node. Most leaf nodes can be identified with ISO 693-3 language ID while most non-leaf nodes have Glottolog IDs attached to them. Next, we use the PHOIBLE as our training phone inventory. PHOIBLE contains 2100 languages, and 2091 of them can be mapped to one of the leaf node in the Glottolog-based tree.

For every unseen node in the tree (leaf or non-leaf), we estimate its phone inventory using our proposed models. For each model, we specify the size of inventory to be $n = 40$, which is a typical size of the phone inventories in our training set. To evaluate the model, we select 77 languages as the unseen testing languages and take them out of our training languages. The languages are selected from a recently proposed multilingual phone dataset (Li et al., 2021c), in which we can identify 77 out of 95 languages in our tree. For every testing language, we evaluate both their inventory coverage (using the F1 score) and the phone recognition accuracy (using phone error rate) as an extrinsic task. The ISO 693-3 id of testing languages are abk, ace, ady, afn, afr, aka, asm, azb, bam, bem, ben, bfd, bfq, bin, brv, bsq, cbv, ces, cha, cpn, dag, dan, deg, dyo, efi, ell, ema, eus, ewe, ffm, fin, fub, gaa, gla, guj, hak, hau, haw, heb, hil, hin, hrv, hun, hye, ibb, ibo, idu, ilo, isl, kan, kea, khm, klu, knn, kri, kub, kye, lad, led, lgq, lit, lkt, lug, mak, mal, mlt, mya, nan, njm, nld, ozm, pam, pes, run, tzm, wuu, yue.

5.1. Phone Inventory Evaluation

Model	F1	Prec	Rec
Fixed Inventory ($\hat{\Sigma}_{\text{fixed}}$)	51.1	48.7	57.5
Global Inventory ($\hat{\Sigma}_{\text{global}}$)	59.4	58.1	64.8
Bayesian Network ($\hat{\Sigma}_{\text{bayes}}$)	61.2	60.0	66.7
Nearest Neighbor ($\hat{\Sigma}_{\text{nearest}}$)	**65.9**	71.3	64.7

Table 1: F1, precision and recall for 77 testing languages and each model. The two models on top are the baseline models and the two on the bottom are the proposed models. We use the Tagalog inventory as the fixed inventory.

Table 1 shows the statistics for the four models: the fixed language inventory has 51.1 F1 with 48.8 precision and 57.5 recall. As mentioned in the previous section, the Tagalog inventory contains many cross-linguistically common phones, which makes the recall much higher than the precision. We found it interesting to investigate which commonly used languages perform better in this regard. We evaluate the top ten languages, ranked by the population of first language speakers (Lewis, 2009). Figure 3 indicates that the Romance branch from the Indo-European language family tends to have relatively high scores, but none of them outperforms the Tagalog inventory. While the fixed language inventory can capture 50% of the inventory, it only consists of the inventory from an individual language and fails to reflect the global properties of all languages. On the contrary, the global inventory baseline is built using statistics from all training languages, which improves the F1 score by 8 points. Our experiment shows that selecting the most frequent phones is essential for the global baseline. We also consider another global inventory baseline which consists of all basic phones available in the IPA table (without diacritics and modifiers). This model only achieves a 27.2 F1 score: it captures most phones in every language (high recall), but it generates many false positives, which significantly decreases the precision.

To further incorporate information from local language branches, we propose the Bayesian Network model and Nearest Neighbor model. Table 1 shows that they further improve the F1 score by 1.8 and 6.5 points respectively. Despite the simplicity of the nearest neighbor model, it outperforms the Bayesian Network model by 4.7 points. This is because the nearest neighbor model can capture more languages than the Bayesian Network Model. Suppose we would like to estimate the inventory of West Germanic (as in Figure 1). The Bayesian model will only rely on the training languages among its children: English alone. On the other hand, the nearest neighbor model can take advantage of training lan-

Figure 3: Comparison of inventory evaluation using different fixed language. Spanish has the highest F1 score among the top-10 languages ranked by the population.

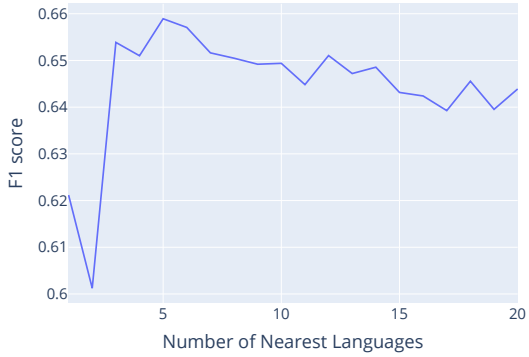

Figure 4: Comparison of performance when using different number of nearest neighbors.

guages in other branches: Norwegian. This gives the nearest neighbor model more information when deciding the inventory, which significantly improve its precision from 60.0 to 71.3. Next, we investigate the effect of using different number of nearest neighbors. Figure 4 is a line plot showing the result of using different number of nearest neighbors (k). We observe a bias-variance trend in our experiment. When $k = 1$, we simply search for the nearest language and use that language to approximate the target language. This suffers from large variance as it only uses one language's inventory. Increasing k reduces the variance by averaging over k nearest languages. However, increasing k too much also hurts the performance as the additional languages are far from the target language and introduce bias into the inventory instead.

5.2. Universal Phone Recognition

Finally, we report the results of the extrinsic task in Table 2. The original phone recognition models propose to use the union of all available phones when the inventory is not available. This approach, again, suffers from the low-precision problem and only achieves 89.2% PER. In contrast, all 4 models introduced in this work (including the two baselines) improve the PER by more than 20%. the nearest neighbor model again achieves the highest performance of 64.2%. The gap between 4 models, however, is smaller than the inventory evaluation. This is because phones are not uniformly distributed in utterances (Li et al., 2021b), and frequent phones typically have already been captured by the global inventory (as we select them based on the sorted order). We observe that adding frequent phones from the global inventory to the proposed models can further improve the results. The major F1 improvements of Bayesian Network and Nearest Neighbor approach comes from the identification of other rare phones, therefore the improvement is reduced in this task. Despite the small gap between the 4 proposed models, we show that using a proper inventory could significantly improve the PER.

Model	PER	Add	Del	Sub
Default Inventory (Σ)	89.2	3.8	16.2	69.1
Fixed Inventory ($\hat{\Sigma}_{\text{fixed}}$)	67.4	3.6	15.4	48.2
Global Inventory ($\hat{\Sigma}_{\text{global}}$)	65.3	3.4	15.2	46.7
Bayesian Network ($\hat{\Sigma}_{\text{bayes}}$)	64.6	2.5	20.1	41.8
Nearest Neighbor ($\hat{\Sigma}_{\text{nearest}}$)	**64.2**	3.2	16.4	44.6

Table 2: Statistics of the universal phone recognition task. Lower PER (phone error rate) indicates better performance. Add, Del, Sub are Addition, Deletion and Substitution errors.

6. Limitations

While we get reasonable performance in our testing languages, we acknowledge that there are several limitations in our approach: first, our approach heavily depends on Glottolog, if the language is not available in the Glottolog database, then our approach cannot be applied to it. Second, if the target language does not have any training languages near it (e.g: it is the only language in its branch), then the approximation might not be accurate.

7. Conclusion

In this work, we propose multiple approaches to estimate phone inventories for unseen languages. By using the knowledge derived from phylogenetic trees, we demonstrate that they significantly improve the inventory quality over competitive baselines and boost performance in a phone recognition task. This work also paves the way for appling speech recognition technology to (almost) every language in the world. All the phone inventories of 7915 languages would be released to enable more researchers to explore them in future research.

8. Bibliographical References

Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., and Weber, G. (2020). Common voice: A massively-multilingual speech corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4218–4222.

Bird, S. and Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, pages 557–582.

Black, A. W. (2019). CMU wilderness multilingual speech dataset. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5971–5975. IEEE.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper& Row, New York.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to algorithms*. MIT press.

Dunbar, E., Algayres, R., Karadayi, J., Bernard, M., Benjumea, J., Cao, X.-N., Miskic, L., Dugrain, C., Ondel, L., Black, A., et al. (2019). The zero resource speech challenge 2019: Tts without t. In *Interspeech 2019-20th Annual Conference of the International Speech Communication Association*.

Dunbar, E., Karadayi, J., Bernard, M., Cao, X.-N., Algayres, R., Ondel, L. B., Sakti, S., and Dupoux, E. (2020). The zero resource speech challenge 2020: Discovering discrete subword and word units.

Hayes, B. (2011). *Introductory phonology*, volume 32. John Wiley & Sons.

International Phonetic Association, International Phonetic Association Staff, et al. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT press.

Lewis, M. P. (2009). *Ethnologue: Languages of the world*. SIL International.

Li, X., Dalmia, S., Li, J., Lee, M., Littell, P., Yao, J., Anastasopoulos, A., Mortensen, D. R., Neubig, G., Black, A. W., and Metze, F. (2020). Universal phone recognition with a multilingual allophone system. In *ICASSP 2020*.

Li, X., Li, J., Metze, F., and Black, A. W. (2021a). Hierarchical Phone Recognition with Compositional Phonetics. In *Proc. Interspeech 2021*, pages 2461–2465.

Li, X., Li, J., Yao, J., Black, A. W., and Metze, F. (2021b). Phone distribution estimation for low resource languages. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7233–7237. IEEE.

Li, X., Mortensen, D. R., Metze, F., and Black, A. W. (2021c). Multilingual phonetic dataset for low resource speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6958–6962. IEEE.

Michaud, A., Adams, O., Cohn, T. A., Neubig, G., and Guillaume, S. (2018). Integrating automatic transcription into the language documentation workflow: Experiments with na data and the persephone toolkit. *Language Documentation and Conservation*.

Steven Moran et al., editors. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., and Levin, L. S. (2016). Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3475–3484. ACL.

Mortensen, D. R., Li, X., Littell, P., Michaud, A., Rijhwani, S., Anastasopoulos, A., Black, A. W., Metze, F., and Neubig, G. (2020). AlloVera: A multilingual allophone database. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*.

Müller, M., Franke, J., Waibel, A., and Stüker, S. (2017). Towards phoneme inventory discovery for documentation of unwritten languages. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5200–5204. IEEE.

Nettle, D., Romaine, S., et al. (2000). *Vanishing voices: The extinction of the world's languages*. Oxford University Press on Demand.

Nordhoff, S. and Hammarström, H. (2011). Glottolog/langdoc: Defining dialects, languages, and language families as collections of resources. In *First International Workshop on Linked Science 2011-In conjunction with the International Semantic Web Conference (ISWC 2011)*.

Varadarajan, B., Khudanpur, S., and Dupoux, E. (2008). Unsupervised learning of acoustic sub-word units. In *Proceedings of ACL-08: HLT, Short Papers*, pages 165–168.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

9. Language Resource References

Creanza, Nicole and Ruhlen, Merritt and Pemberton, Trevor J and Rosenberg, Noah A and Feldman, Marcus W and Ramachandran, Sohini. (2015). *A comparison of worldwide phonemic and genetic variation in human populations*. National Acad Sciences, ISLRN 811-991-003-866-7.

Moran, Steven and McCloy, Daniel and Wright, Richard. (2014). *PHOIBLE online*. Max Planck Institute for Evolutionary Anthropology, ISLRN 596-736-657-858-7.