ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447345

SELF-SUPERVISED MODELS OF SPEECH INFER UNIVERSAL ARTICULATORY KINEMATICS

Cheol Jun Cho[1] *Abdelrahman Mohamed*[2] *Alan W Black* [3] *Gopala K. Anumanchipalli*[1]

[1]UC Berkeley [2] Rembrand [3] Carnegie Mellon University

ABSTRACT

Self-Supervised Learning (SSL) based models of speech have shown remarkable performance on a range of downstream tasks. These state-of-the-art models have remained blackboxes, but many recent studies have begun "probing" models like HuBERT, to correlate their internal representations to different aspects of speech. In this paper, we show "inference of articulatory kinematics" as fundamental property of SSL models, i.e., the ability of these models to transform acoustics into the causal articulatory dynamics underlying the speech signal. We also show that this abstraction is largely overlapping across the language of the data used to train the model, with preference to the language with similar phonological system. Furthermore, we show that with simple affine transformations, Acoustic-to-Articulatory inversion (AAI) is transferrable across speakers, even across genders, languages, and dialects, showing the generalizability of this property. Together, these results shed new light on the internals of SSL models that are critical to their superior performance, and open up new avenues into language-agnostic universal models for speech engineering, that are interpretable and grounded in speech science.

Index Terms— Self-Supervised Learning; Articulatory Kinematics; Cross-lingual and Multilingual Speech Processing;

1. INTRODUCTION

Self-supervised learning (SSL) has revolutionized every field of machine learning, by providing rich features of natural data without human-annotated labels. Likewise, speech SSL models have been proven to be successful in various speech downstream tasks [1]. To understand such utility, the internal representation of speech SSL models has been scrutinized by probing analyses for known speech and linguistic features, such as low-level acoustics, phonetics, and lexical semantics [2, 3, 4, 5]. A comparative analysis by Cho et al. [5] demonstrates that the state-of-the-art SSL models are highly correlated with articulatory kinematics and the correlation score can indicate the success of the SSL model in downstream tasks. This finding is extended to developing a high-performance Acoustic-to-Articulatory inversion (AAI) model [6].

Here, we test an intriguing hypothesis – speech SSL models infer the causal articulatory processes that generate the speech acoustic signal. If this hypothesis is true, such inference should be agnostic to language or dialect, as speech acoustics are a result of the resonance associated with vocal tract shapes that add spectral detail to the vocal source, i.e., vibration of the vocal folds. Besides, the human species has a common canonical structure of vocal tract anatomy and orofacial muscles, regardless of ethnic group, language, and dialect.

We validate this hypothesis using two approaches: cross-lingual articulatory probing and cross-speaker transferability analysis. First, we probe articulatory representation in speech SSL models trained on different languages, using a comprehensive set of public electromagnetic articulography (EMA) datasets that cover up to 62 speakers from different language and dialect groups. Then, we test how well each individual articulatory system can be transferred to another system, by fitting an affine transformation. Combining these two analyses elucidates the universality of the articulatory representation in speech SSL models.

While EMA has been a major interface of studying articulatory phonology [7, 8, 9, 10], the data collection procedure is highly complicated with many physical restrictions that have posed significant barriers in collecting a large scale dataset [11]. Moreover, a majority of the publicly available datasets are concentrated in a few common languages like English.[1] Therefore, the development of a universal AAI model can reduce the burden of collecting new EMA data for investigating diverse languages. This is especially valuable for languages with low resources or near extinction where only a short amount of audio is accessible. Our results suggest that the articulatory inference in speech SSL models is transferable across languages, making a critical step toward the universal AAI model.

Our major findings are:

- Articulatory kinematics of different languages and dialects can be recovered from speech SSL models with a simple linear projection, regardless of the language for which the SSL models are trained.
- Individual articulatory systems can be aligned by affine transformations, implying the existence of a canonical basis of articulatory kinematics.
- While the overall scores are surprisingly high, we observe a preference for language or dialect, providing evidence of language-specific articulatory phonology reflected in speech SSL models.

2. RELATED WORK

Speech SSL. Wav2vec-2.0 (W2V2) [12] and HuBERT (HB) [13] have been the most successful SSL models of speech. Both models are trained by masked prediction objectives, where the Transformer encoder predicts randomly masked parts of input features. As raw audio waves are inadequate for target reference, instead, the models predict self-driven units, either from vector quantization (W2V2) or online clustering (H2B) of internal features.

Cross-lingual speech model. XLS-R [14] and MMS [15] are direct extensions of W2V2, which aim to learn cross-lingual speech representation by training on multiple languages. XLS-R is trained on 50K hours of speech from 53 languages, and MMS is trained on 55K hours of speech from more than 1,000 languages. As the models are trained across languages, they may learn articulatory representation

[1]This is a significant bottleneck given there are more than 7,000 languages around the world.

that is shared across languages.

Probing SSL representation. Several efforts have been made to explain which features are encoded by SSL. Pasad et al. [2, 4] compare lexical semantics with speech SSL models across layers, revealing that lexical semantics rise in the later layer of SSL models. Shah et al. [3] train non-linear models to probe audio, fluency, pronunciation, and text features, providing a detailed description of information processing in SSL models. Furthermore, clustering analysis suggests that the primary information encoded in SSL models is fine-grained subphonemic units [16, 17]. Our work is a direct extension of Cho et al. [5] that demonstrates evidence of high-fidelity articulatory kinematics in SSL models.

AAI. Inversion models often suffer from individual differences in EMA data. A major source of such variability is anatomical differences across speakers. Furthermore, though key articulators are standardized, there is some level of arbitrariness in the locations of the sensors. Together, such misalignment across subjects has challenged developing a speaker independent AAI model. The recent work by Wu et al. [6] mitigates the problem by utilizing relative distances between articulators, which are proven to be more consistent across speakers than EMA. This is well aligned with our finding that affine transformations can align individual articulatory systems.

3. METHODS

3.1. Speech SSL models trained for different languages

We use HB and W2V2 with large sizes (300M parameters) to extract speech representations. We retrieved either of those models from Hugginface, trained with a specific language: English, Mandarin, Korean, French, or Dutch. Furthermore, two cross-lingual W2V2 models are included: XLS-R [14] with 300M parameters trained on 53 languages, and MSS [15] with 1B parameters trained on more than 1000 languages. For baseline reference, a set of acoustic features (filter bank, mel spectrogram, and MFCC) are included as well.

3.2. Comprehensive set of multi-lingual, multi-dialect EMA

Our comprehensive analyses include five public EMA datasets that cover 62 speakers from three languages: English, Mandarin, and Italian (Table 1). For English speakers, there are three regional dialect groups where speakers are from the UK, USA, or China (Beijing or Shanghai). As the EMA data can be highly noisy due to the known instability in the data collection procedure, we filter out three Mandarin speakers from the original corpus. Note that the speakers validated in the original dataset papers are included in the final dataset. Each dataset has a frame-wise annotation of kinematic traces of six articulators: low incisor (LI), upper lip (UL), lower lip (LL), tongue tip (TT), tongue blade (TB), and tongue dorsum (TD)), on the midsagittal plane: X (front-back) and Y (up-down). Each trace is normalized to be zero mean and unit variance.[2] All of these data were collected while naturally speaking full sentences.

3.3. Probing SSL models with articulatory kinematics

We follow the previous articulatory probing approach [5] to measure the correlation between SSL representation and articulatory kinematics. Features from 20 ms audio frames are extracted from each layer of each SSL model and projected to the EMA space by fitting a linear inversion model. One difference from the previous study is that a 6 Hz low-pass filter is applied to remove high-frequency noise.

[2]The normalization is done per channel and within the clip.

Table 1. Summary of the public EMA datasets used in the analyses. Different language-dialect groups are denoted as EN.{UK, US, BJ, SH}: English speakers from the UK; the US; Beijing, or Shanghai, China, respectively, MAN: Mandarin speakers, and IT: Italian speakers. The number of speakers used for the transferability analysis is denoted in $(\cdot)$ if different. The gender distribution is balanced in MOCHA-TIMIT, HPRC, and EMA-MAE.

Corpus	Lang.	Spk #	m/spk
MNGU0 [18]	EN.UK	1	75
MOCHA-TIMIT [19]	EN.UK	7	27
HPRC [20]	EN.US	8	59
EMA-MAE [21]	EN.US	20 (18)	12
	EN.BJ	10 (9)	17
	EN.SH	9 (5)	16
DKU-JNU-EMA [22]	MAN	4 (2)	20
MSPKA [23]	IT	3 (2)	47
Total	–	62 (52)	–

The prediction performance is assessed by calculating the correlation between predictions and the ground truth reference, and the results are averaged over a 5-fold cross-validation.

3.4. Evaluating transferability between speakers

We hypothesize that the linear inversion models obtained from the probing analysis are leveraging the same basis of the articulatory subspace that resides in SSL representation. To test this hypothesis, we devise a transferability metric, which measures the correlation between two articulatory systems. Suppose we have two AAI models, f_A and f_B, for different speakers, A and B. As there is a variability in the vocal tract anatomy across individuals, we trained another linear model, $g_{A\to B}$, to align the articulatory systems, which is trained with L1 regularization weighted with α = 0.01 to impose sparsity in coefficients. The transferability from speaker A to speaker B is then measured as the correlation of the prediction by these two aligned inversion systems, corr$(g_{A\to B} \circ f_A, f_B)$. For this analysis, linear inversion models built upon the 17th layer of XLS-R are used and the data from the target speaker (B) is used for training and testing. We only include speakers with high inversion performance (corr $\geq$ 0.8), totaling 52 speakers. We denote the A and B speaker pair as the source and target pair.

4. RESULTS

4.1. Articulatory probing shows high correlations agnostic to training language of SSL models

Fig.1 shows the prediction performance of probing SSL models in §3.1 that are trained in different languages. For specific SSL models, the performance of the layer with the highest correlation is reported.[3] For some feature sets with multiple models, the models with higher scores are selected.[4] Despite the substantial discrepancy between languages, the average correlations are surprisingly high, reaching over 0.8 regardless of the training language (Fig.1 *left*), which are far beyond the baseline acoustic features (corr = 0.66$\pm$0.040). This indicates that the SSL models naturally learn

[3]The layer-wise patterns are omitted due to the limited space, but the patterns are consistent with [5].

[4]English and Mandarin have both HB and W2V2.

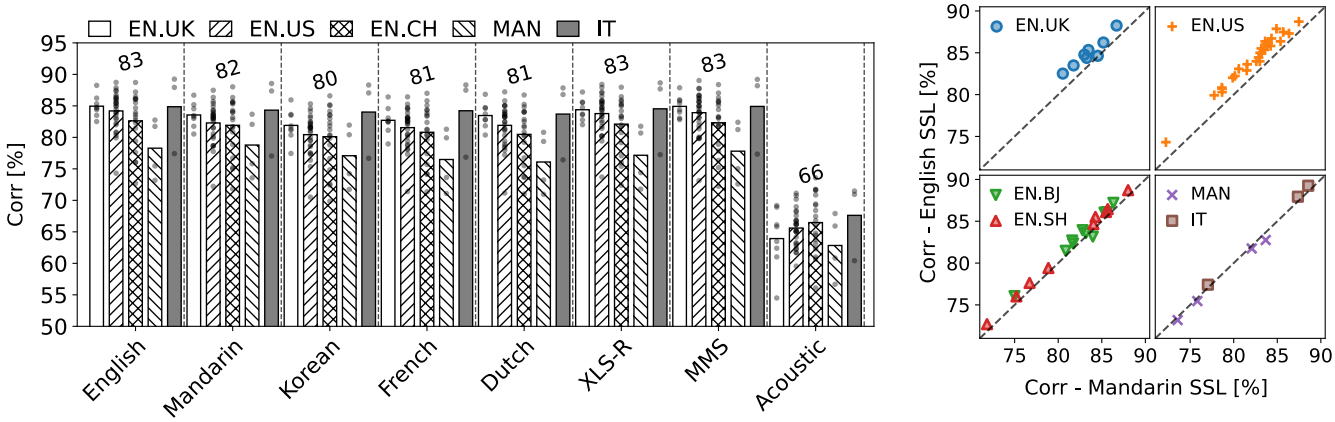

Fig. 1. (*left*) EMA prediction performance of probing SSL models from different languages. The correlations are averaged across 12 EMA channels. Each language-dialect group is denoted by a hatch pattern and each dot denotes an individual speaker. Regardless of the language, the average performance reaches over 0.8. (*right*) Performance comparison of English SSL versus Mandarin SSL, each panel shows specific language-dialect groups. The diagonal dashed lines denote identity lines. For English, native speakers (EN.UK/US) prefer English models over Mandarin models, scattered slightly above the diagonals, but speakers from China (EN.BJ/SH) show almost identical scores.

physical relationships between raw speech signals and vocal tract articulations, which are not bound to specific language, but rather universal properties of the human vocal tract system. Amongst languages, the highest performance is achievable by English SSL models (corr = 0.835±0.039) and cross-lingual models (XLS-R: 0.830±0.039, MMS: 0.832±0.040). The performance difference between MMS and the other two best models is not significant. Given the scale of MMS in both training data and model size, this finding suggests that scaling up has a minimal effect, and learning articulatory representation is readily saturated in the SSL regime.

Probing models for individual speakers show variable performances due to the noisy nature of the data collection, which is mostly induced by the unstable attachment of sensors. Still, 44% of speakers show high performance above 0.85 correlations, and the best subject achieves 0.89 correlation.[5] This is remarkable given that we are applying a very simple linear regression on the frozen SSL models that are never exposed to EMA data while training.

4.2. Preference for language in articulatory probing

Despite the overall high correlations, some noteworthy variability across languages is observable. English models and Mandarin models show higher correlations than models of other languages. As most of the speakers are from either the US, UK, or China, this may indicate some language specificity. However, this difference can not be fully attributed to the language difference because the probing performance is also correlated with in general representational quality of SSL model [5].

To further test language specificity in SSL models, we compare English SSL models and Mandarin SSL models for each dialect group in English speakers (Fig.1 *right*). The native English speakers (EN.UK/US) prefer English models over Mandarin models (Fig.1 *right*; top two panels), to a small extent yet statistically significant (mean diff = 0.019; $p < 1e-5$). However, the English speakers from China (EN.BJ/SH) show almost identical probing performances of these models (Fig.1 *right*; bottom left panel), showing non-significant difference (mean diff = -0.002; p = 0.29). As the language is controlled to be English, the observed preference is attributed to differences in articulatory phonology rather than other linguistic components (e.g., semantics or syntax). This indicates that there exist some residual articulatory habits originating from speaking Mandarin as a mother tongue language. The other language groups, MAN and IT, show relatively minor differences and the number of speakers is too few to evaluate the significance.

4.3. Individual articulatory systems are mutually transferable with affine transformations

For each pair of speakers, we measure the transferability scores (§3.4) between the articulatory systems obtained from the probing analyses. The far-left matrix of Fig.2 shows the average correlations between source and target group pairs. The scores within groups are strikingly high, showing near or above 0.85 average correlation (diagonal brown boxes). This suggests that the articulatory systems of individual speakers are affine transformations of each other. Such high correlations are also observable in the transformations across language-dialect groups, especially between the Italian (IT) and the native English speakers (EN.UK/US), which show even higher scores than within-group scores. The IT has in general high transferability to others, indicating that the AAI models built upon one of these particular speakers can serve as the closest universal AAI model. These high-fidelity affine transformations between different speakers demonstrate a significant level of isometry in the human articulatory system. This is strong evidence that speech SSL models infer the canonical basis of the articulatory kinematics.

The coefficients in the affine transformations are mostly assigned to the same articulators (Fig.2 *far-right*; red boxes). This suggests that the transformations are geometric alignments that correct the variability induced by the difference in vocal tract anatomy or the sensor locations. While most of the six articulators show definite self-assignments, the tongue blade (TB) is affected by other tongue parts, which is natural since the tongue is highly deformable and TB bridges the tongue tip and dorsum. Fig.3 shows transferability scores for each articulator. Some channels, LIX, LIX, ULY, and LLX, are relatively difficult to align, which reflects their innate subtlety in shaping phonetic information.

[5] One of the Italian speakers using the 10th layer of HB trained on English (among 24 layers).

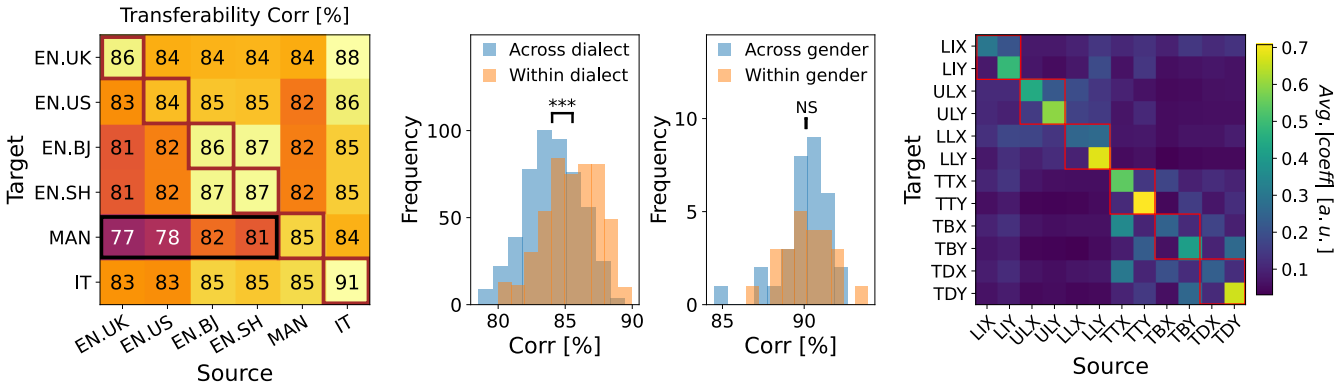

Fig. 2. (*far-left*) Correlation matrix denoting transferabilities between language-dialect groups. The scores are averaged over possible pairs between groups. (*mid-left*) Distributions of correlations across dialects (blue) and within dialects (orange). The English speakers from China (EN.SH+EN.BJ) and the native English speakers from the US (EN.US) groups in the EMA-MAE dataset are used. (*mid-right*) Distributions of correlations across gender (blue) and within gender (orange). Four male and four female speakers from the HPRC dataset are used. (*far-right*) The average absolute coefficients in the affine transformations between the articulatory systems.

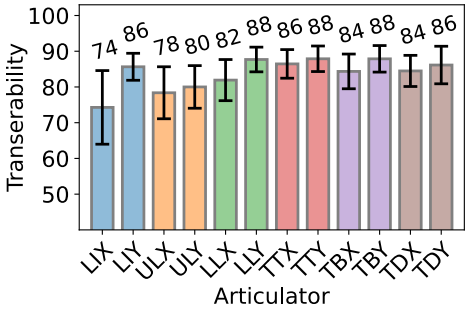

Fig. 3. Transferability scores of each articulator averaged over all source-target pairs.

4.4. The variance in transferability reflects language-specific articulatory phonology

When we compare the within and across dialects transferability, the across dialect transformations show significantly lower scores than those of the within dialect cases (Fig.2 *mid-left*). However, when groups are divided by gender, there is no significance between the within and across gender transformations (Fig.2 *mid-right*). As gender is the most significant factor of anatomical differences, this result suggests that the transferability between articulatory systems is not affected by the vocal tract anatomy. Therefore, the observed patterns across languages and dialects are more likely due to differences in articulatory phonology. This is consistent with the language preference in cross-lingual probing §4.2. To avoid variance induced by the data collection site, we confined analyses on dialect groups and gender groups to the EMA-MAE corpus and HPRC corpus, respectively.

This language-specific pattern is also reflected in the transferability matrix (Fig.2 *far-left*; black box). While the transferability scores from the English groups (EN.*) to the Mandarin group (MAN) are generally lower than in other cases, the Chinese English groups (EN.BJ/SH) are more transferable than the native English speakers (EN.UK/US). Such a pattern is not observed in the opposite cases (MAN $\rightarrow$ EN.*).

5. DISCUSSION

We demonstrate that the recent speech SSL models can recover articulatory kinematics with simple linear mapping, achieving high performance inversion regardless of speakers, languages, and dialects. This is further verified by finding affine transformations from one articulatory system to another. Our findings provide strong evidence that there is a canonical basis of articulatory phonology which is naturally emerging in self-supervised learning of speech.

Our findings evoke another interesting hypothesis that articulatory kinematics are not only the physical interface of speech but also the continuous embedding representations of phonetics. As the SSL models are trained by the masked prediction objective without any external labels, the resulting representations are perceptual descriptions of how sounds are shaped in natural speech data, which also shows high correspondence to the human auditory perception [24]. Here, we demonstrate that the natural selection of the perceptual embedding space of speech is a continuous, dynamic articulatory space. Particularly, the group-level analyses in §4.4 reveal the language specificity in articulatory systems but little evidence of gender specificity that entails large anatomical differences. This observation further supports the hypothesis of equivalence between articulatory kinematics and continuous phonetic embeddings.

The articulatory information captured by EMA is not complete as limited to the midsagittal plane, and upper bounded by the inevitable noise ceiling of the sensors. Also, the source information such as vocal fold vibration is missing. However, our findings indicate that a complete set of articulatory kinematics may lie in SSL representations which are missing in EMA. This is an interesting future work toward unsupervised discovery of articulatory kinematics without relying on EMA.

6. ACKNOWLEDGEMENTS

This research is supported by the following grants to PI Anumanchipalli — NSF award 2106928, BAIR Commons-Meta AI Research, the Rose Hills Innovator Program, and UC Noyce Initiative, at UC Berkeley. Special thanks to Shang-Wen (Daniel) Li for discussions and for providing advice.

7. REFERENCES

[1] Shu-wen Yang, Po-Han Chi, Yung-Sung Chuang, Cheng-I Jeff Lai, Kushal Lakhotia, Yist Y Lin, Andy T Liu, Jiatong Shi, Xuankai Chang, Guan-Ting Lin, et al., "Superb: Speech processing universal performance benchmark," *arXiv preprint arXiv:2105.01051*, 2021.

[2] Ankita Pasad, Ju-Chieh Chou, and Karen Livescu, "Layer-wise analysis of a self-supervised speech representation model," in *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021, pp. 914–921.

[3] Jui Shah, Yaman Kumar Singla, Changyou Chen, and Rajiv Ratn Shah, "What all do audio transformer models hear? probing acoustic representations for language delivery and its structure," *arXiv preprint arXiv:2101.00387*, 2021.

[4] Ankita Pasad, Bowen Shi, and Karen Livescu, "Comparative layer-wise analysis of self-supervised speech models," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[5] Cheol Jun Cho, Peter Wu, Abdelrahman Mohamed, and Gopala K Anumanchipalli, "Evidence of vocal tract articulation in self-supervised learning of speech," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[6] Perter Wu, Li-wei Chen, Cheol Jun Cho, Shinji Watanabe, Louis Goldstein, Alan W Black, and Gopala K Anumanchipalli, "Speaker-independent acoustic-to-articulatory speech inversion," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[7] Catherine P Browman and Louis Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[8] Vikram Ramanarayanan, Louis Goldstein, and Shrikanth S Narayanan, "Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1378–1394, 2013.

[9] Jiachen Lian, Alan W Black, Louis Goldstein, and Gopala Krishna Anumanchipalli, "Deep neural convolutive matrix factorization for articulatory representation decomposition," *arXiv preprint arXiv:2204.00465*, 2022.

[10] Peter Wu, Shinji Watanabe, Louis Goldstein, Alan W Black, and Gopala Krishna Anumanchipalli, "Deep Speech Synthesis from Articulatory Representations," in *Proc. Interspeech 2022*, 2022, pp. 779–783.

[11] Teja Rebernik, Jidde Jacobi, Roel Jonkers, Aude Noiray, and Martijn Wieling, "A review of data collection practices using electromagnetic articulography," *Laboratory Phonology*, vol. 12, no. 1, pp. 6, 2021.

[12] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems*, vol. 33, pp. 12449–12460, 2020.

[13] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed, "Hubert: Self-supervised speech representation learning by masked prediction of hidden units," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3451–3460, 2021.

[14] Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, et al., "Xls-r: Self-supervised cross-lingual speech representation learning at scale," *arXiv preprint arXiv:2111.09296*, 2021.

[15] Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, et al., "Scaling speech technology to 1,000+ languages," *arXiv preprint arXiv:2305.13516*, 2023.

[16] Amitay Sicherman and Yossi Adi, "Analysing discrete self supervised speech representation for spoken language modeling," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[17] Badr M Abdullah, Mohammed Maqsood Shaik, Bernd Möbius, and Dietrich Klakow, "An information-theoretic analysis of self-supervised discrete representations of speech," *arXiv preprint arXiv:2306.02405*, 2023.

[18] Korin Richmond, Phil Hoole, and Simon King, "Announcing the electromagnetic articulography (day 1) subset of the mngu0 articulatory corpus," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[19] Alan Wrench, "Mocha: multichannel articulatory database," http://www.cstr.ed.ac.uk/research/project/artic/mocha.html, 1999.

[20] Mark Tiede, Carol Y Espy-Wilson, Dolly Goldenberg, Vikramjit Mitra, Hosung Nam, and Ganesh Sivaraman, "Quantifying kinematic aspects of reduction in a contrasting rate production task," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3580–3580, 2017.

[21] An Ji, Jeffrey J Berry, and Michael T Johnson, "The electromagnetic articulography mandarin accented english (ema-mae) corpus of acoustic and 3d articulatory kinematic data," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 7719–7723.

[22] Zexin Cai, Xiaoyi Qin, Danwei Cai, Ming Li, Xinzhong Liu, and Haibin Zhong, "The dku-jnu-ema electromagnetic articulography database on mandarin and chinese dialects with tandem feature based acoustic-to-articulatory inversion," in *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. IEEE, 2018, pp. 235–239.

[23] Claudia Canevari, Leonardo Badino, and Luciano Fadiga, "A new italian dataset of parallel acoustic and articulatory data," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[24] Yuanning Li, Gopala K Anumanchipalli, Abdelrahman Mohamed, Junfeng Lu, Jinsong Wu, and Edward F Chang, "Dissecting neural computations of the human auditory pathway using deep neural networks for speech," *bioRxiv*, pp. 2022–03, 2022.