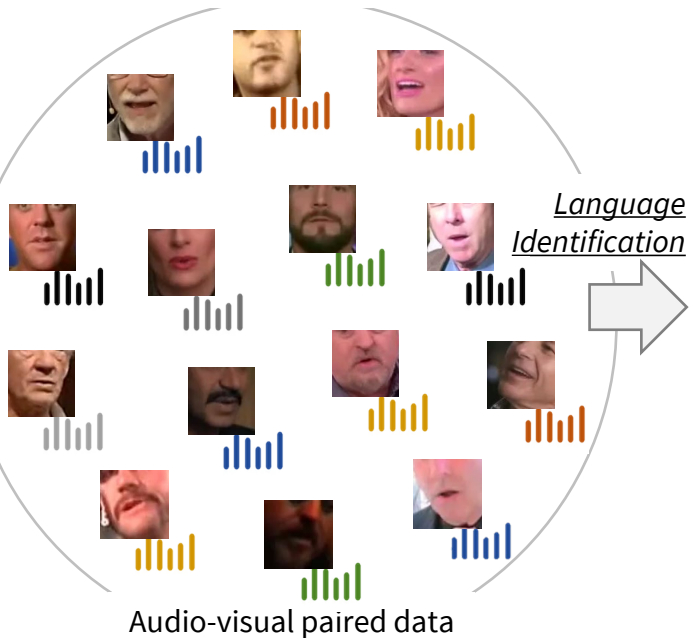
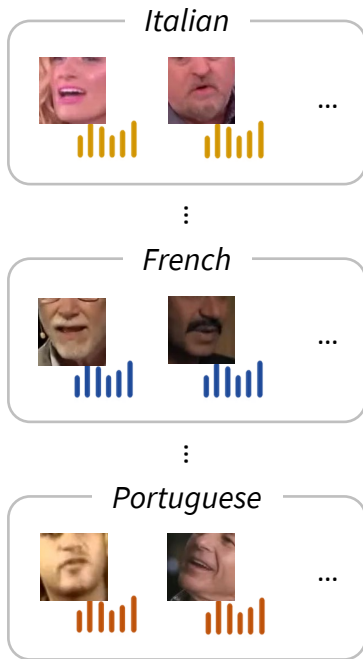


Large-scale Multilingual  
Unlabeled Audio-Visual Data Pool



Target Language Sets



Automatic Labels

