Our team's objective is to train an automatic speech recognition model that transcribes speech in the International Phonetic Alphabet (IPA). Since audio transcription takes a large amount of time in the process of language documentation, such a model can be an aid for field linguists. We implemented a speech-to-IPA model by fine-tuning Wav2Vec-Large-XLSR with training dataset from Common Voice with manually added IPA transcription. For evaluating speech-to-IPA models, we also proposed a new metric, Feature-weighted Phoneme Error Rate (FPER), which takes into account (dis)similarities of each phoneme. As the evaluation data, we used a speech clip in Karabakh Armenian, which neither of the models included in the training data. With this metric, our model (16.1% FPER) was twice as accurate as the previous work (Allosaurus, 34.2% FPER).