Towards Cross-Language Prosody Transfer for Dialog

Jonathan E. Avila, Nigel G. Ward • University of Texas at El Paso

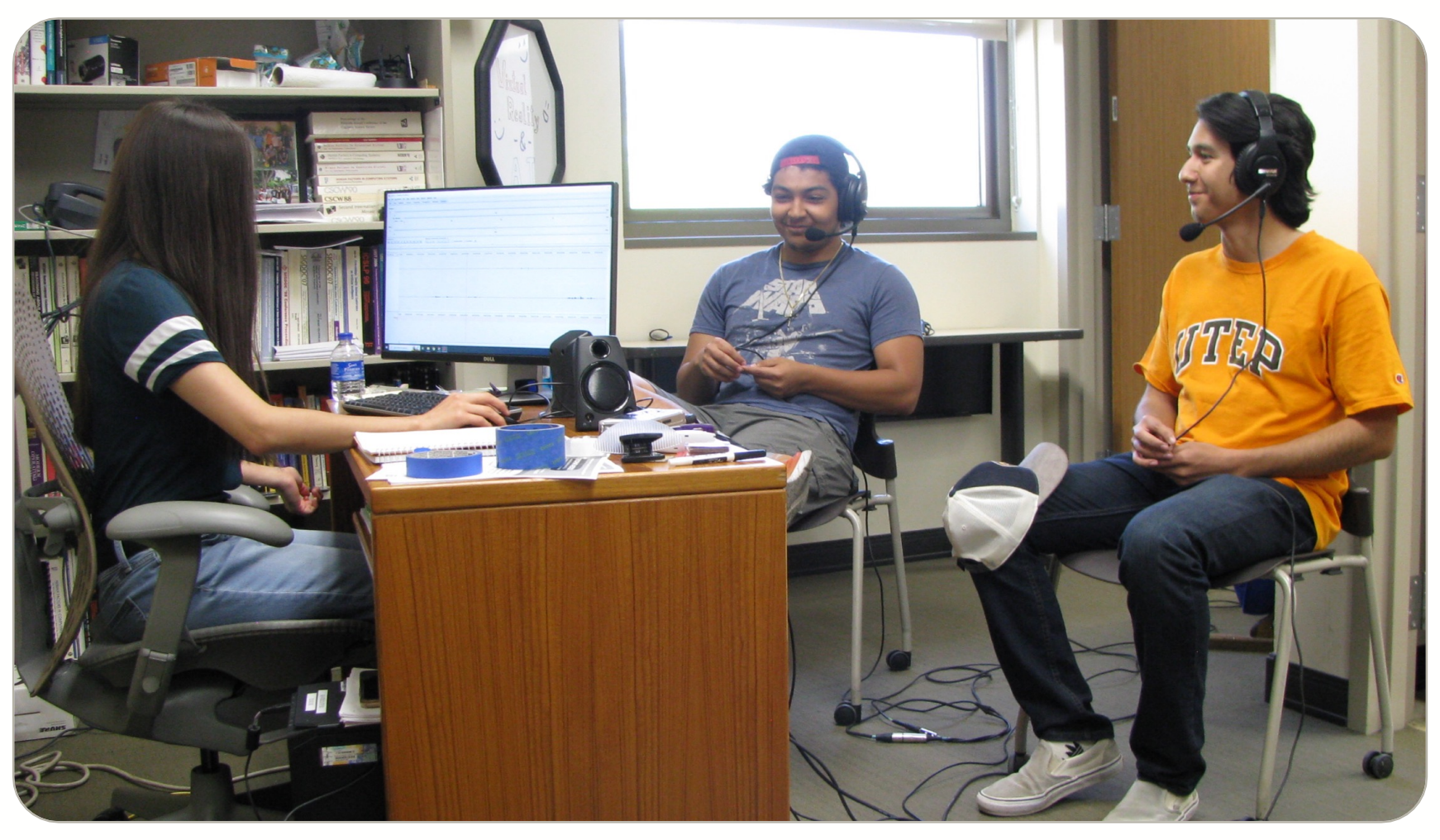
Motivation

- **Speech-to-speech translation systems** are useful
- but their output prosody is generic
- which is unsatisfactory for dialog beyond simple factual or transactional interactions

⇒ **We need to map prosody across languages**

Contribution: A Bilingual Corpus of Matched Utterances

- From dialog
- Pragmatically diverse
- Natural, faithful, and freely available
- Sourced from bilingual speaker pairs who converse in one language, then re-enact some utterances in the other language
- English (EN) and Spanish (ES)
- 3816 utterance pairs, average duration 2.7 s



X: *Vas a tener tu propio,*	X: *You're going to have your own,*
Y: *Ai, si cierto.*	Y: *Ah, that's right.*
X: *departamento.*	X: *apartment.*
Y: *Ya el jueves.*	Y: *Already on Thursday.*
X: *¿El jueves?*	X: *On Thursday?*
Y: *El jueves me lo van a dar, el jueves a las tres de la tarde.*	Y: *On Thursday they're going to give it to me, on Thursday at three in the afternoon.*
X: ***¿Van a venir, venir tus papás para?***	X: ***Are you parents gonna come, or?***

Findings about the difficulty of mapping prosody

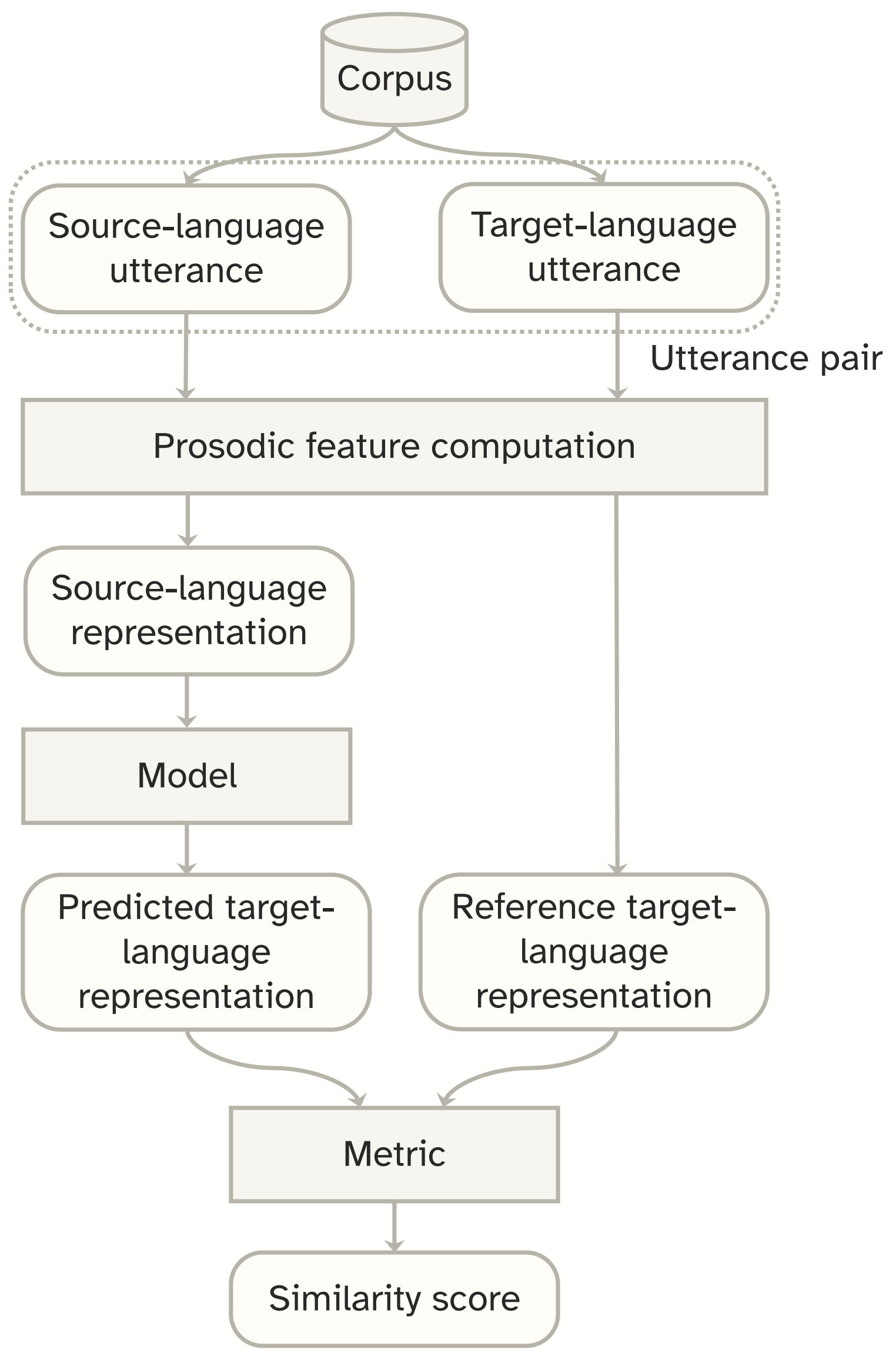
1. Ignoring the source-language prosody does not work well: synthesizer-output prosody is quite unlike the human reference
2. Copying the source-language prosody is only slightly better
3. Using the source-language prosody helps, even with a simple linear regression model

⇒ **Doing better should be easy**

Average Prediction Error of Models in Prosody Translation Tasks

Model	EN→ES	ES→EN
Source-ignoring	12.6	12.3
Direct-copy	11.4	11.4
Linear regression	9.2	9.4

- Task: predict the target-language prosodic representation
- Representation: 100 diverse time-spanning features that are robust to speaker differences, all z-normalized
- Metric: Euclidean distance from predicted representation to the reference


Corpus
Source-language utterance
Target-language utterance
Utterance pair
Prosodic feature computation
Source-language representation
Model
Predicted target-language representation
Reference target-language representation
Metric
Similarity score

Overview of Prosody Translation Task

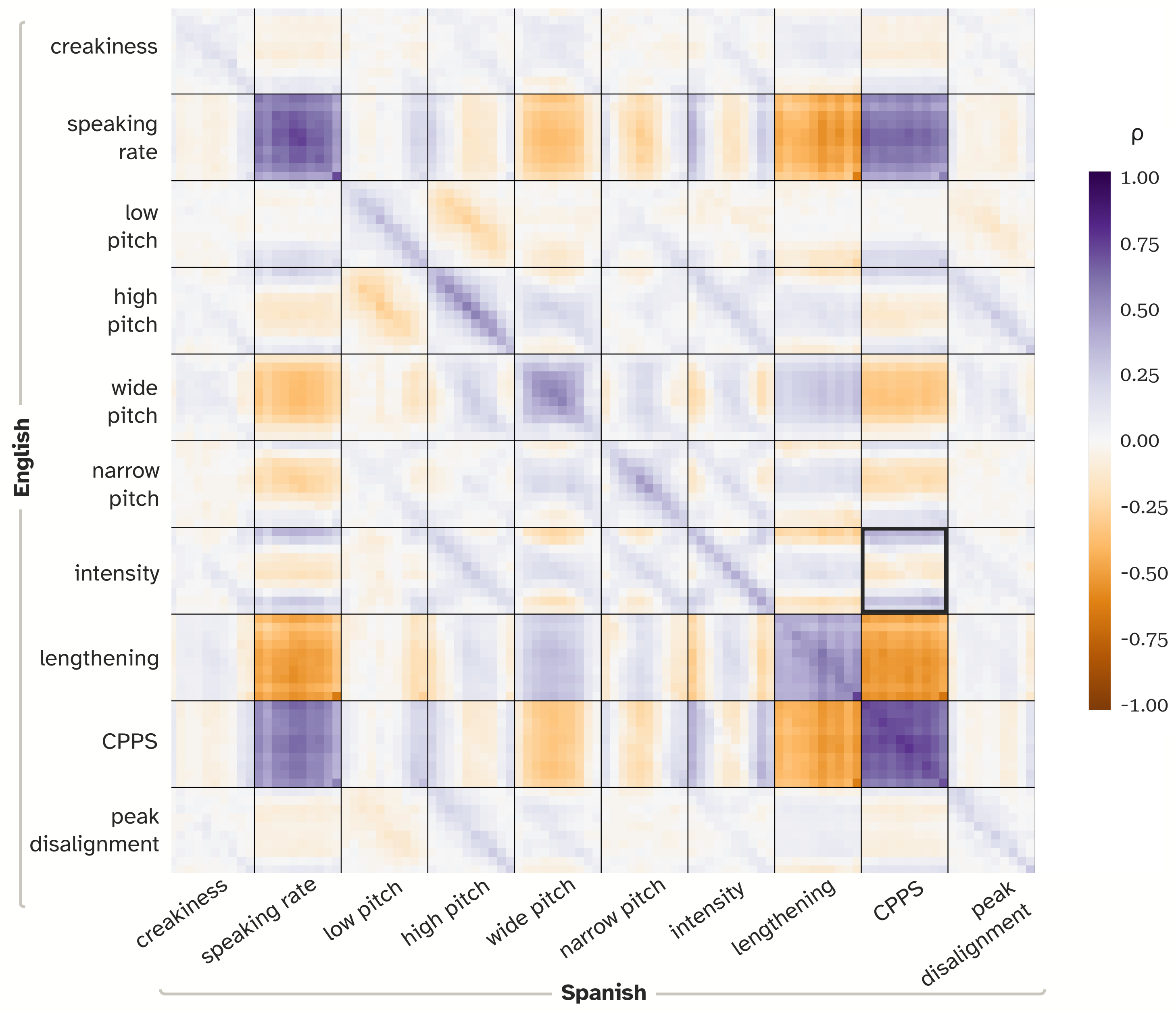
Findings about English and Spanish prosody differences

1. Overall, many similarities between English and Spanish
   - Shown by the strong diagonal
2. English near-final intensity and Spanish high CPPS correlate
   - Often seems to convey an upcoming continuation

     Example EN: *If you have an undergrad in anything, you can just, skip to a Master's in anything else*
     ES: *Si tienes carrera en cualquier cosa, puedes brincar a la maestría en lo que sea*
3. English breathiness and final pitch rise, not seen in the Spanish translations*
   - Grounding can be realized by uptalk in English

     Example EN: *I was in the varsity team*
     ES: *Estaba en el varsity team*

*Based on failure analysis of the direct-copy model



Correlation Matrix of Spanish vs. English Prosodic Features

Future Work

- Bilingual corpus: larger, additional language pairs
- Prosodic similarity metric: will be trained to match human perceptions of pragmatic similarity
- Prosody mapping models: possibly using pre-trained models

References

- Jonathan E. Avila. Forthcoming. Towards a Model of the Mapping Between English and Spanish Prosody. Dissertation, University of Texas at El Paso.
- Wen-Chin Huang, Benjamin Peloquin, Justine Kao, Changhan Wang, Hongyu Gong, Elizabeth Salesky, Yossi Adi, Ann Lee, and Peng-Jen Chen. 2023. A Holistic Cascade System, Benchmark, and Human Evaluation Protocol for Expressive Speech-to-Speech Translation.
- Daniel J. Liebling, Michal Lahav, Abigail Evans, Aaron Donsbach, Jess Holbrook, Boris Smus, and Lindsey Boran. 2020. Unmet Needs and Opportunities for Mobile Translation AI. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, ACM, 1–13.
- Nigel G. Ward, Jonathan E. Avila, Emilia Rivas, and Divette Marco. 2023. Dialogs Re-enacted Across Languages, Version 2. University of Texas at El Paso.