

Word-Level Error Analysis in Decoding Systems: From Speech Recognition to Brain-Computer Interfaces

Jingya Huang, Aashish N. Patel, Sowmya Manojna Narasimha, Gal Mishne, Vikash Gilja

University of California San Diego

NIH NIH NIDCD R01DC018446

NSF NSF 2223822 EFRI BRAID

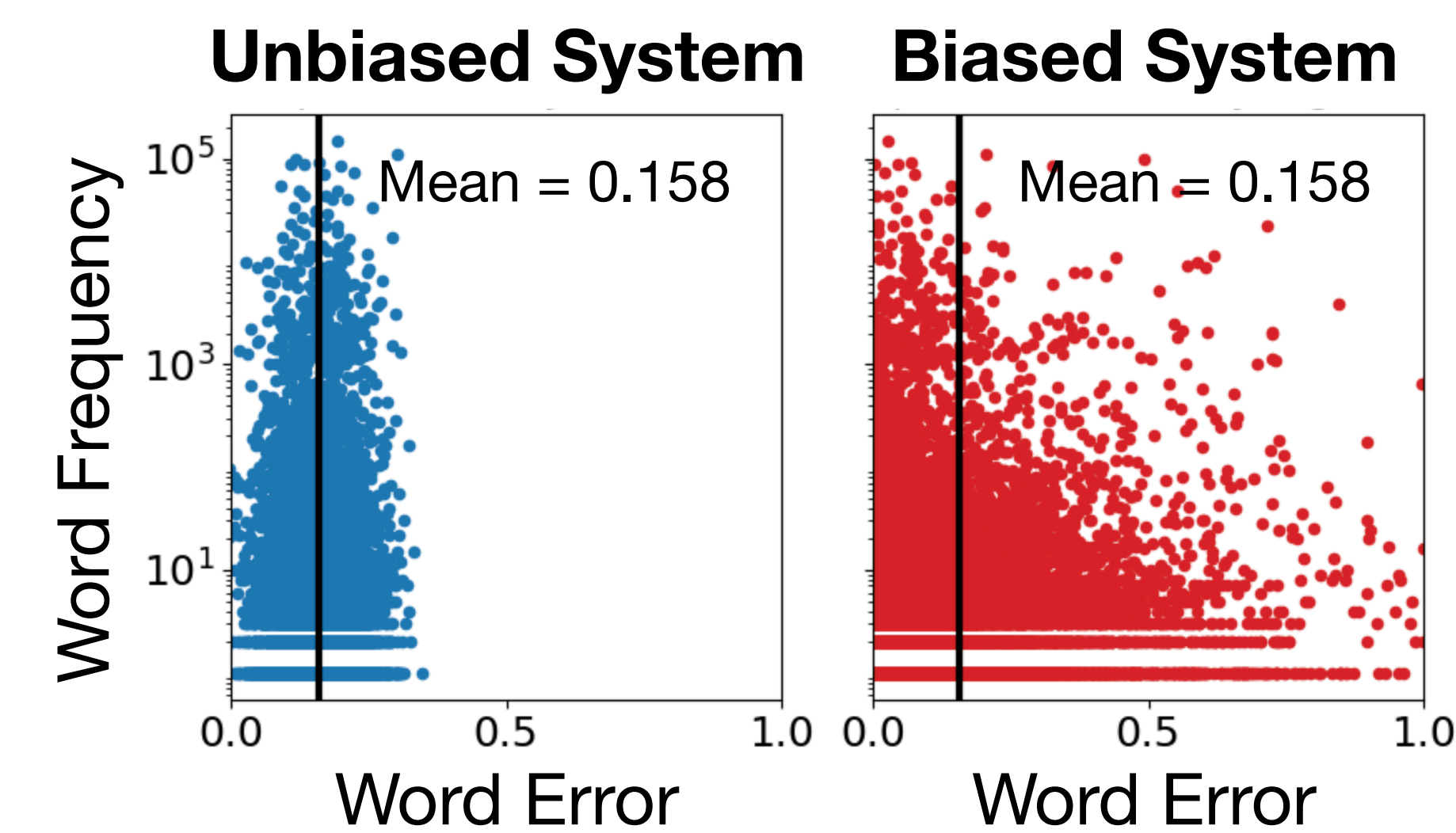
UC San Diego



Low Word Error Rate (WER) \neq Low Error for Every Word

- BCIs restore movement or communication (via Brain-to-Text) for individuals with motor and speech impairments
- SOTA Brain-to-Text achieves 4.2% WER and 32 words per minute, supporting daily conversation use (Willet et al. 2023, Card et al. 2024)
- Brain to text (BTT) and Automatic Speech Recognition (ASR)
 - Shared goal: transcribe intended speech to text
 - Different inputs: data statistics, quality, abundance
 - Shared challenge: learning classification + temporal alignment
- Method:** refined word-level metrics for correctness and semantic cost
- Finding:** SOTA models (1 BTT, 3 ASR) show frequency-driven error gaps
- Impact:** closing the word generalization gap improves ASR & BTT usability

Same WER, Different Error Patterns
“We have **covered** all the chapters **for** the exam.”
“The car **was covered** in snow.”
Decoded as “The car was **come over** in snow.”
“The car **as** covered in snow.”



Word-Level Error Metrics

Different Alignment Strategies Shift How Error are Attributed

Reference: They worked to straighten out the misunderstanding
Decoded: They worked street out the misunderstanding
Naïve Word Edit: = = [S] [D] = = =

Reference: They worked to straighten out the misunderstanding
Decoded: They worked stree t out the misunderstanding
Refined Char Edit: ===== dd ===== dd =====
= = [D] [S] = = =

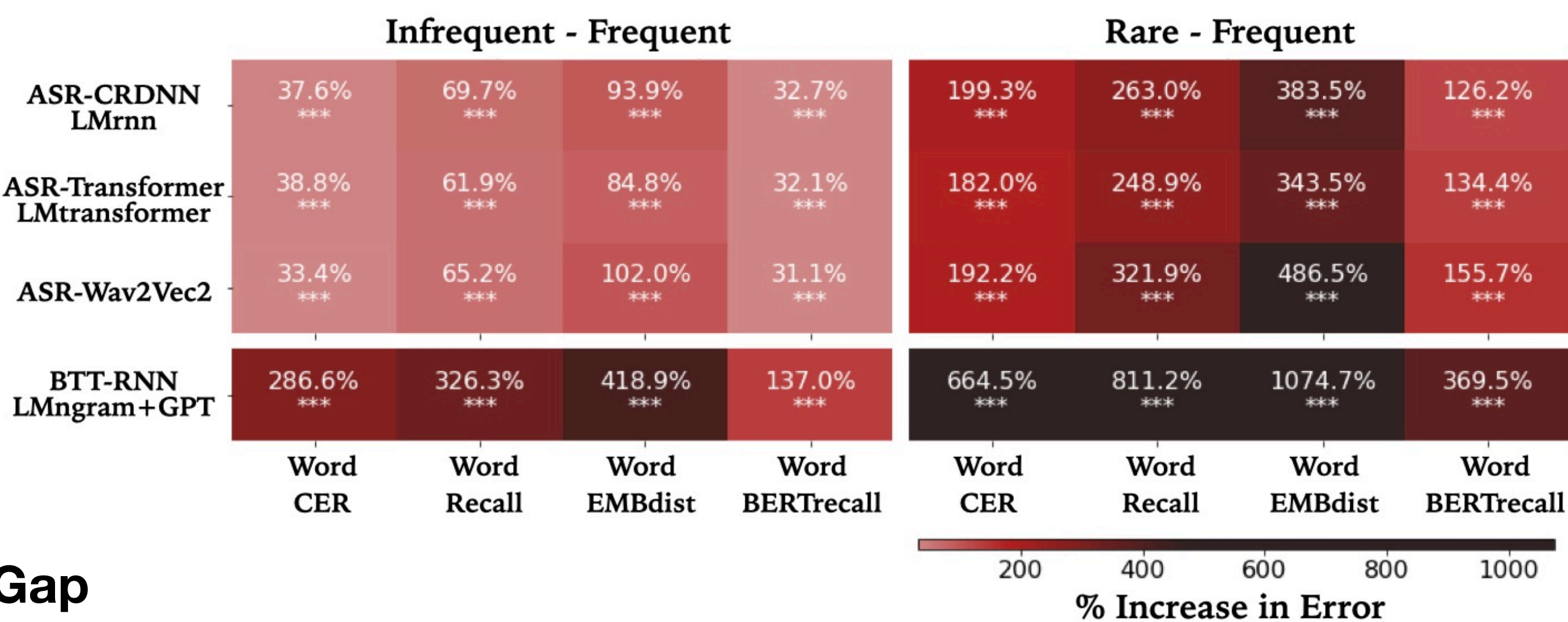
*Additionally, account for compound and split word variations (e.g., “rain-bow” \leftrightarrow “rainbow”)

Word metrics on edit pairs (e.g., “straighten” \rightarrow “street”)

- Word-CER:** degree of textual variation
- Word-Recall:** rate of correct recall
- Word-EMBDist:** semantic distance of the edit
- Word-BERTrecall:** recoverability with context

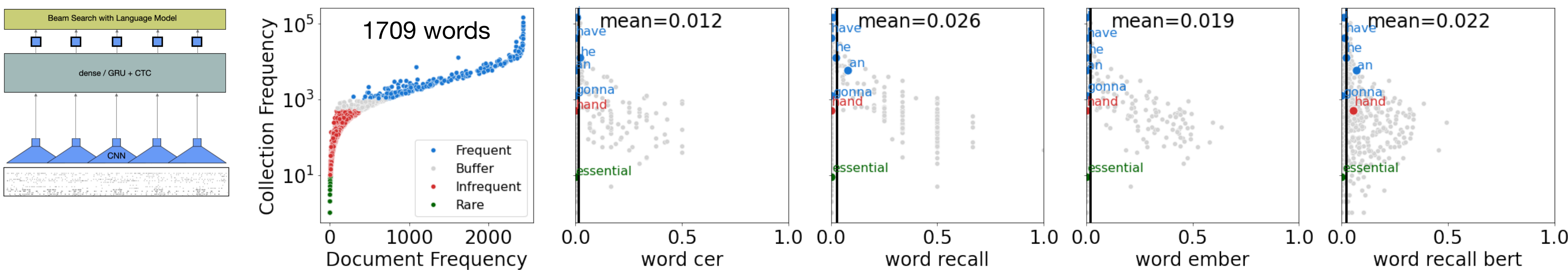
Word-Level Generalization Gap in SOTA BTT and ASR Systems

- Subword decoding still yields uneven word accuracy, with bias likely from imbalanced word occurrence in sentences
- Context variability and alignment ambiguity may amplify bias, mitigable through better representation learning
- Closing the generalization gap for infrequent words can enhance the semantic relevance of decoded output

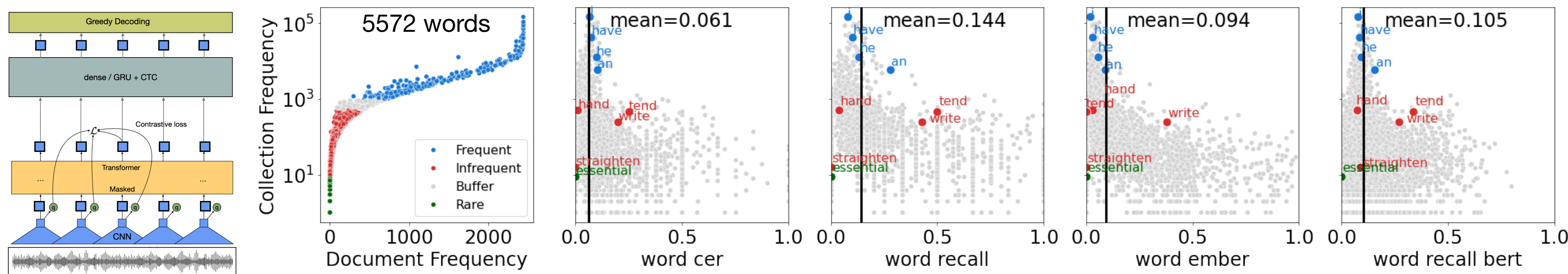


SOTA BTT and ASR Both Show the Word-Level Generalization Gap

BTT-RNN
WER=0.042
Single speaker
Neural spike train
19k Train Sen.



ASR-Wav2Vec2
WER=0.1585
Multiple speaker
Acoustic waveform
500k Train Sen.



Infrequent word edits cause greater semantic distortion

Model	Decoded Sentence	Infrequent Word Edit Only	Frequent Word Edit Only
ASR-Wav2Vec2	but you will still have to have a north ndon straight now your teeth	but you will still have to have an north-ndon to straight out your teeth	but you will still have to have a orthodontist straighten now your teeth
ASR-Transformer	but you will still have to have an orthodontist the street now you are cheap	but you will still have to have an orthodontist to street out your cheap	but you will still have to have an orthodontist the straighten now you are teeth
ASR-CRDNN	but you will still have to have a north that do not it is the street now you keep	but you will still have to have an north to street out your keep	but you will still have a orthodontist that do not it is the straighten now you keep

Comparison of model decoding of sentence “but you will still have to have an orthodontist to straighten out your teeth”.



Code Available
TNEL-UCSD/
word-metrics