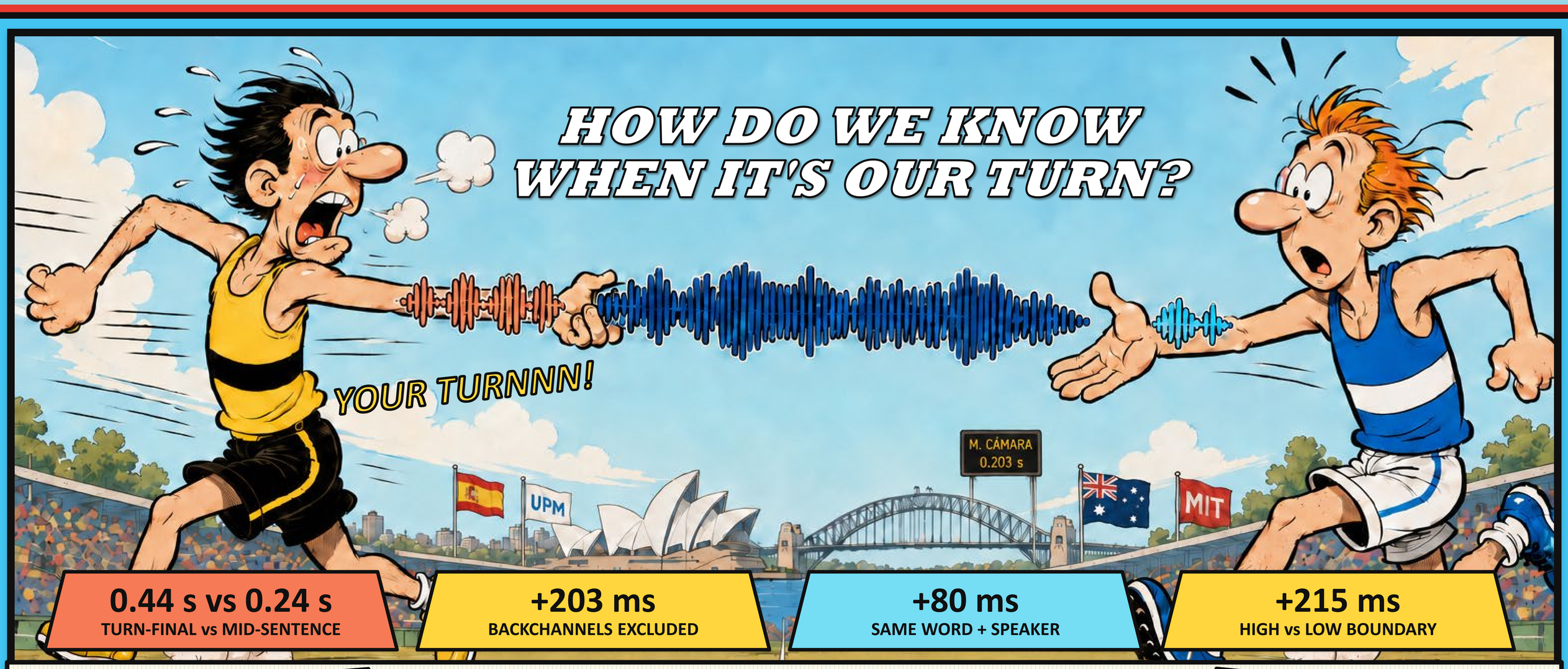
WORD LENGTHENING AS A FUNCTION OF UTTERANCE POSITION: A MULTI-CORPUS STUDY

Mateo Cámara[1] · José Luis Blanco[1] · Juan Ignacio Godino-Llorente[1] · Jeung-Yoon Choi[2] · Stefanie Shattuck-Hufnagel[2]

1 Universidad Politécnica de Madrid · 2 Massachusetts Institute of Technology



0.44 s vs 0.24 s
TURN-FINAL vs MID-SENTENCE

+203 ms
BACKCHANNELS EXCLUDED

+80 ms
SAME WORD + SPEAKER

+215 ms
HIGH vs LOW BOUNDARY

TURN-FINAL = last word before a speaker change or a long completion silence

4 CORPORA · 2 LANGUAGES · ~39,500 TURNS

MID-SENTENCE = Sentence endings inside the same turn, based on break index

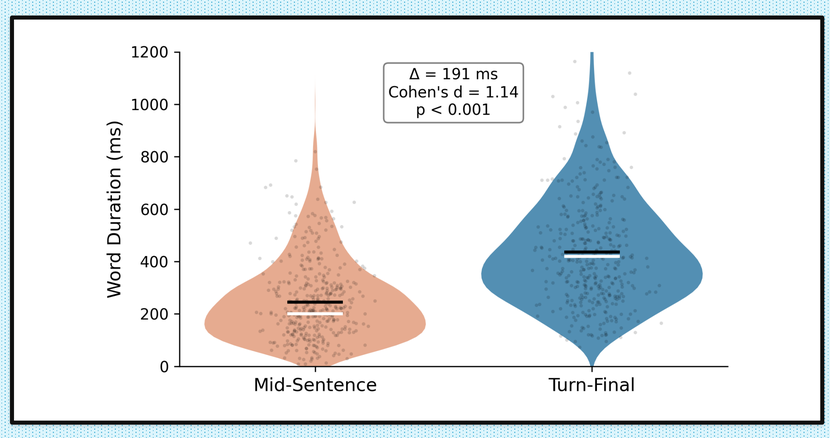
SWITCHBOARD
Telephone dialogue
17,151 TF / 79,736 MS

COLUMBIA GAMES
Task-oriented dialogue
7,437 TF / 66,417 MS

BU RADIO
Read broadcast speech
7,821 TF / 23,954 MS

GLISSANDO
Spanish read speech
7,061 TF / 36,161 MS

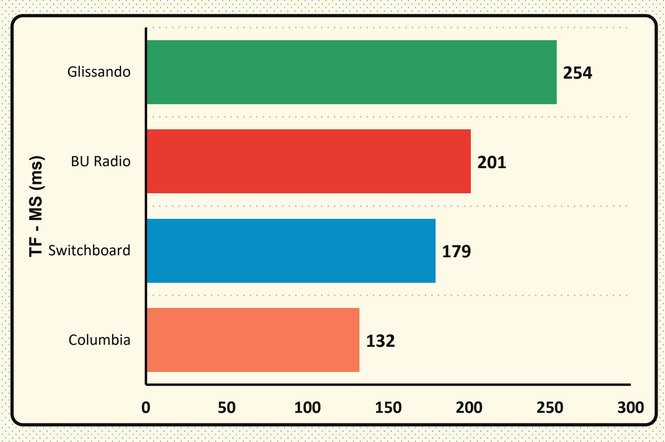
1 POOLED EFFECT: +203 ms



TF 0.44 s
MS 0.24 s
Δ +203.1 ms
RATIO 1.85×
EFFECT d = 1.22
TEST p < 0.001

Including backchannels: +191 ms, d = 1.14. Excluding them strengthens the effect.

2 4 / 4 CORPORA AGREE

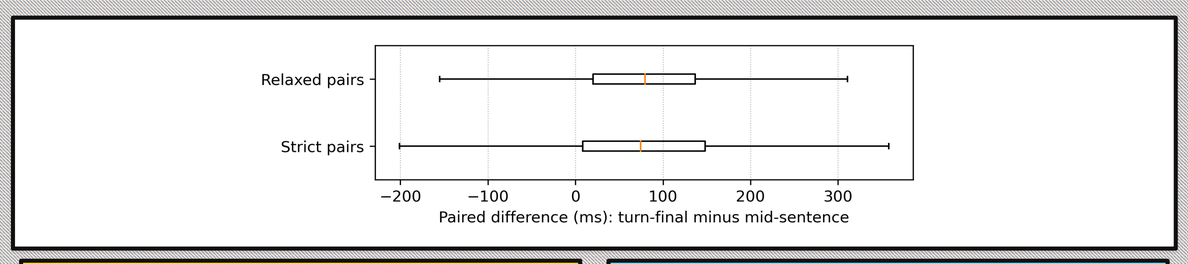


CORPUS	TF	MS	Δ	d
Columbia	.38	.25	+132	.78
Switchboard	.40	.22	+179	1.19
BU Radio	.50	.30	+201	1.08
Glissando	.51	.26	+254	1.47

TF/MS in seconds · Δ in milliseconds

Effect sizes span d = 0.78-1.47 across spontaneous, task-oriented and read speech.

3 SAME WORD. SAME SPEAKER. STILL LONGER.

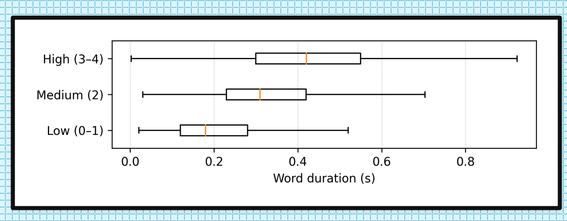


STRICT PAIRS N = 9,018
+79.9 ms · median 74.5 · d = .59 · 77.1% positive

RELAXED PAIRS N = 6,547
+81.1 ms · median 79.5 · d = .69 · 81.6% positive

92.8% of word types are positive (452 / 484). The 32 negative items are mostly backchannel-like words.

4A BOUNDARY STRENGTH LENGTHENS WORDS

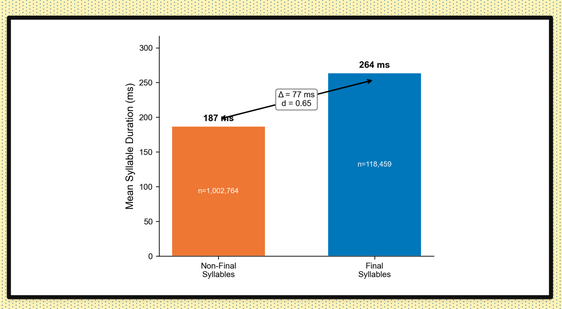


LOW 0.22 s → MEDIUM 0.34 s → HIGH 0.44 s

HIGH vs LOW +215 ms · 1.97× · d = 1.35

ToBI-style labels from Switchboard + BU Radio reveal a graded timing cue.

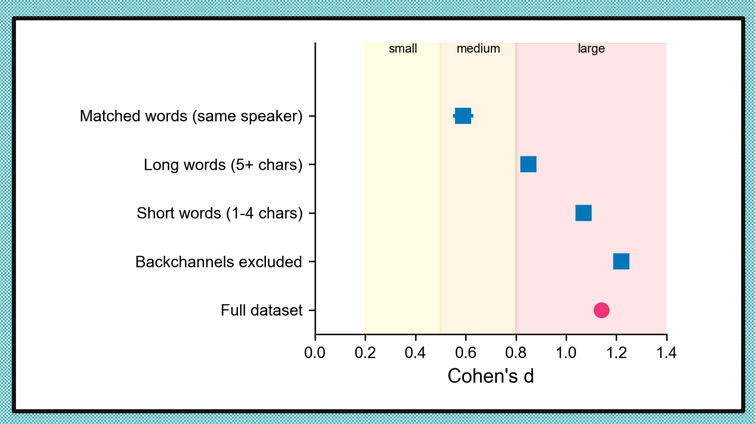
4B THE EXTRA TIME SITS AT THE EDGE



Within turn-final words: 0.29 s vs 0.18 s · d = 0.89

Across turn positions: final-syllable d = .09; non-final d = .01. Boundary-adjacent, not global slowing.

5 THE EFFECT SURVIVES THE CONTROLS



BACKCHANNELS ONLY +22 ms · d = .13

EXCLUDED +203 ms · d = 1.22

SHORT 1-4 CHARS +143 ms · d = 1.07

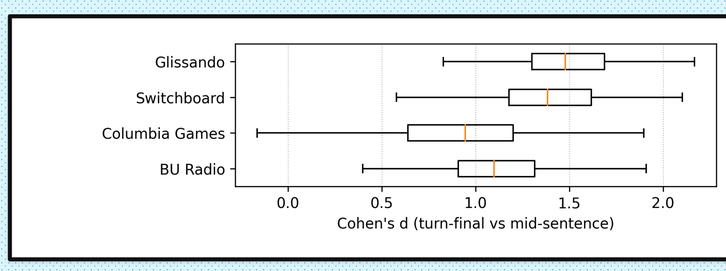
LONG 5+ CHARS +141 ms · d = .85

GLISSANDO BY SYLLABLE COUNT

MONO d = 3.08 · DI d = 1.42 · POLY d = .75

Not driven by backchannels, orthographic length or syllable count.

6 POSITIVE ACROSS RECORDINGS - AND USEFUL FOR MODELS



4 / 4
CORPORA WITH POSITIVE MEDIAN d
Non-positive units appear only in a small fraction of Columbia sides.

- **ROBUST:** 4 corpora, 2 languages, spontaneous + read speech.
- **CONTROLLED:** same word + speaker + session: +79.9 ms.
- **LOCALIZED:** boundary d = 1.35; final syllable d = 0.89.

TAKEAWAY
Turn-final lengthening is a robust, localized cue to floor transfer.
One acoustic cue - not a complete turn-end predictor.

HEAR THE TURN ENDING

Audio examples + supplementary materials

DEMO

FULL PAPER

arXiv:2606.23232

PAPER

mateo.camara@upm.es

ACKNOWLEDGMENTS

Supported by PID2021-128469OB-I00, Fundación Santander, and MISTI MIT Global Experiences.

Interspeech 2026