Université Paris Cité

Nicolas Ballier

Whisper for L2 scoring from segmental to suprasegmental features

Grenoble, 11 April 2025 LIDILEM / Maison des Langues

Joint research with Maelle Amand, Taylor Arnold, Maelle Bourbon, Léa Burin, Tori Fullerton, Adrien Méli, Behnoosh Namdarzadeh, Sara Ng, Erin Pacquetet, Chloé Scholent, Guillaume Wisniewski, Richard Wright & Jean-Baptiste Yunès (also made possible through a CNRS deputation at LLF)

Université Paris Cité

ALTAE

Langues et cultures connectées

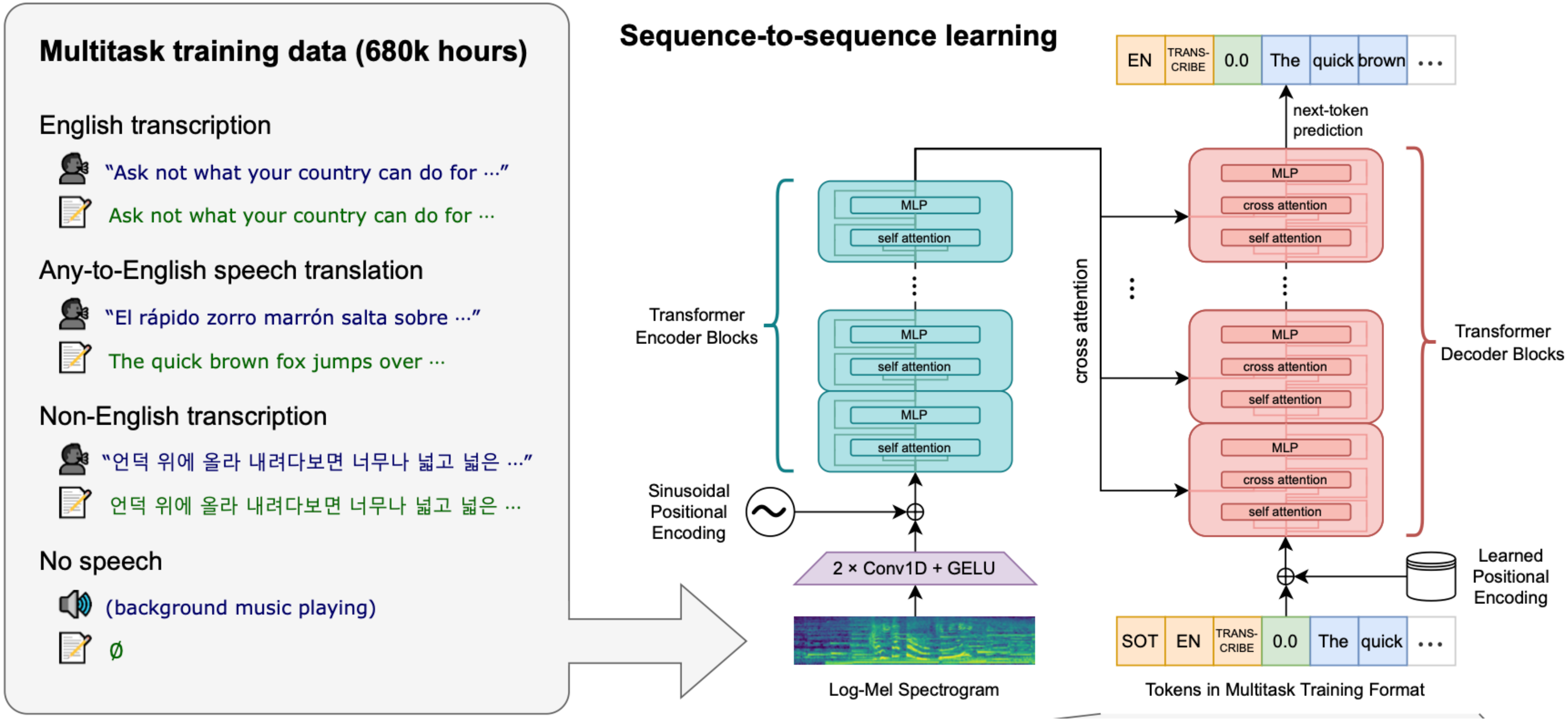
Outline of the Presentation

- Quick introduction to Whisper
- Main Results for the investigation of segmental features with Whisper (IJSNLP2023, IJST 2024, NLP4CALL2024) : global scoring
- Segmental analysis : A roadmap (Janus WP2.1) for subtoken level scoring and potentially mispronunciation detection and diagnosis module
- Suprasegmental analysis (Janus WP2.2-4)

Lexical stress: the reanalysis hypothesis?

- Intonation : the 3 Ts and fine-tuning strategies
- Next Plans (Method : reverse engineering)
- Discussion

Whisper training (Radford 2023)



Multitask training data (680k hours)

English transcription

"Ask not what your country can do for ···"

Ask not what your country can do for ···

Any-to-English speech translation

"El rápido zorro marrón salta sobre ···"

The quick brown fox jumps over ···

Non-English transcription

"언덕 위에 올라 내려다보면 너무나 넓고 넓은 ···"

언덕 위에 올라 내려다보면 너무나 넓고 넓은 ···

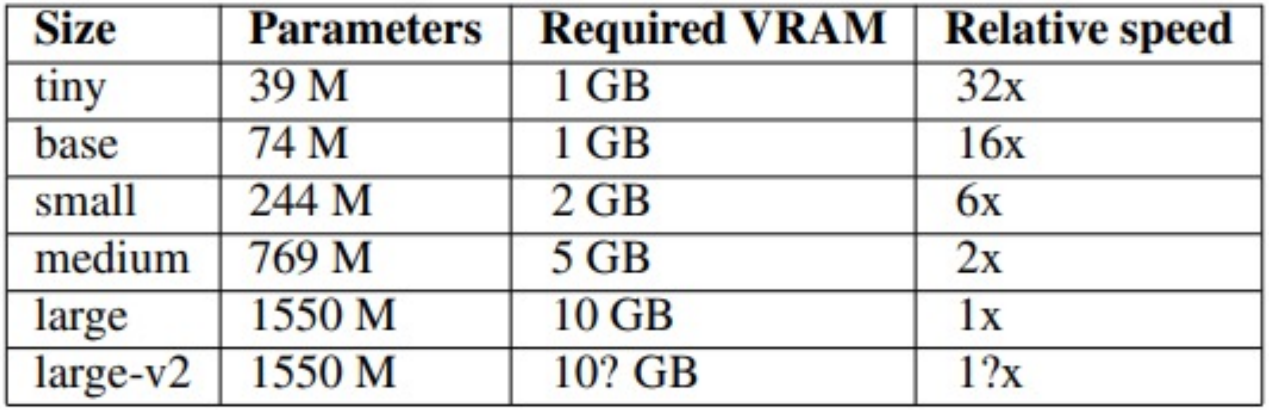
No speech

(background music playing)

∅

Sequence-to-sequence learning
EN
TRANS-CRIBE
0.0
The
quick
brown
...
next-token prediction
MLP
cross attention
self attention
Transformer Encoder Blocks
cross attention
Transformer Decoder Blocks
Sinusoidal Positional Encoding
2 × Conv1D + GELU
Learned Positional Encoding
SOT
EN
TRANS-CRIBE
0.0
The
quick
...
Log-Mel Spectrogram
Tokens in Multitask Training Format

Main parameters of the Whisper models (Radford 2022 + Whisper github)



Size	Parameters	Required VRAM	Relative speed
tiny	39 M	1 GB	32x
base	74 M	1 GB	16x
small	244 M	2 GB	6x
medium	769 M	5 GB	2x
large	1550 M	10 GB	1x
large-v2	1550 M	10? GB	1?x

Table 1: Whisper models tested for this experiment

https://huggingface.co/models?search=openai/whisper

The large-v3 model is trained on 1 million hours of weakly labeled audio and 4 million hours of pseudo-labeled audio collected using large-v2.
https://github.com/openai/whisper/discussions/1762

Initial intuition working on translation and transcription : Interlanguage Retranscription & Named Entity Recognition (NER) Issues

Chomsky

- expected model /ˈtʃɒmski/ French realisation [ʃɔmski] for <Chomsky>
- Different interpretations of different models:
- *Je me ski* (tiny)
- *J'aime ce qui* (base)
- *James Key* (medium)
- *Jomski* (large)
- *Jamsky* (small/large-v2)

Ballier, N. Namdarzadeh, B. Zimina, M. and Yunès, J.-B. (2023) Translating Dislocations or Parentheticals : Investigating the Role of Prosodic Boundaries for Spoken Language Translation of French into English, Proceedings of Machine Translation Summit XIX Vol. 2: Users Track, 119-132. https://files.sciconf.cn/upload/file/20230827/20230827195133_32318.pdf

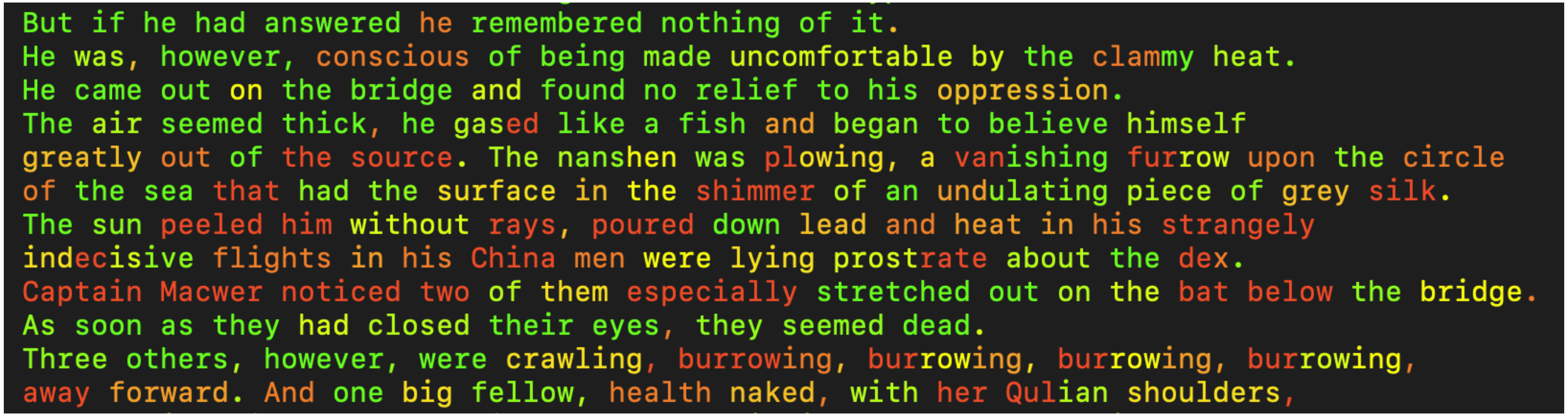
Plausible uses of Whisper for segmental analysis

- Language detection feature for A1 identification (work in progress for A2)
- average subtoken probability score for level/CEFR correlates
- Levenshtein distance as robust measure / correlate to levels
- Tiny/ tiny.en more sensitive to learner variation

- To be more systematically tested for spontaneous speech : Delta between tiny.en (sensitivity to distorsion) and medium for « reference » transcription

-> Papers on global scoring

Analysing confidence scores with C++ implementation of Whisper (Gerganov 2022)

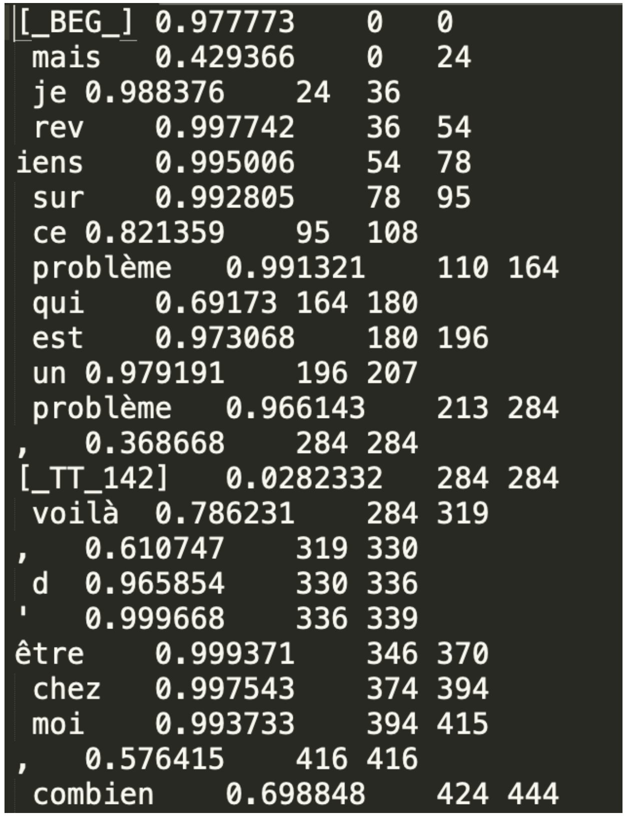


```
But if he had answered he remembered nothing of it.
He was, however, conscious of being made uncomfortable by the clammy heat.
He came out on the bridge and found no relief to his oppression.
The air seemed thick, he gased like a fish and began to believe himself
greatly out of the source. The nanshen was plowing, a vanishing furrow upon the circle
of the sea that had the surface in the shimmer of an undulating piece of grey silk.
The sun peeled him without rays, poured down lead and heat in his strangely
indecisive flights in his China men were lying prostrate about the dex.
Captain Macwer noticed two of them especially stretched out on the bat below the bridge.
As soon as they had closed their eyes, they seemed dead.
Three others, however, were crawling, burrowing, burrowing, burrowing, burrowing,
away forward. And one big fellow, health naked, with her Qulian shoulders,
```

Herculean

https://github.com/ggerganov/whisper.cpp https://github.com/jbyunes/whisper.cpp

Probing Whisper scores with C++ implementation



```
[_BEG_] 0.977773    0   0
 mais   0.429366    0   24
 je 0.988376    24  36
 rev    0.997742    36  54
iens    0.995006    54  78
 sur    0.992805    78  95
 ce 0.821359    95  108
 problème   0.991321    110 164
 qui    0.69173 164 180
 est    0.973068    180 196
 un 0.979191    196 207
 problème   0.966143    213 284
,   0.368668    284 284
[_TT_142]   0.0282332   284 284
 voilà  0.786231    284 319
,   0.610747    319 330
 d  0.965854    330 336
'   0.999668    336 339
être    0.999371    346 370
 chez   0.997543    374 394
 moi    0.993733    394 415
,   0.576415    416 416
 combien    0.698848    424 444
```

https://github.com/jbyunes/whisper.cpp

Reverse engineering : the meaning of Whisper’s special subtokens

- • 50,255 linguistic subtokens, corresponding to English words or fragments for French or graphemes for languages like Persian;
- • special tokens, some of them corresponding to boundaries of the Transformer: the end of text and end of sentence subtokens 50257 *[_EOT_]* and 50258 *[_SOT_]*;
- • 100 extra-tokens labelled *[_ex- tra_token_50259]* to *[_extra_token_50359]*;
- • 7 special tokens are also acknowledged in the literature such as 50360 *[_SOLM_]*, 50361 *[_PREV_]*, 50362 *[_NOSP_]*, 50363 *[_NOT_]* and 50364 *[_BEG_]*. *[_BEG_]* cor- responds to the beginning of the 30 second window when the sound file is processed by Whisper;
- • 1,500 out-of-vocabulary OOV subtokens from *[_TT_1]* to *[_TT_1500]*. they correspond to temporal subtokens

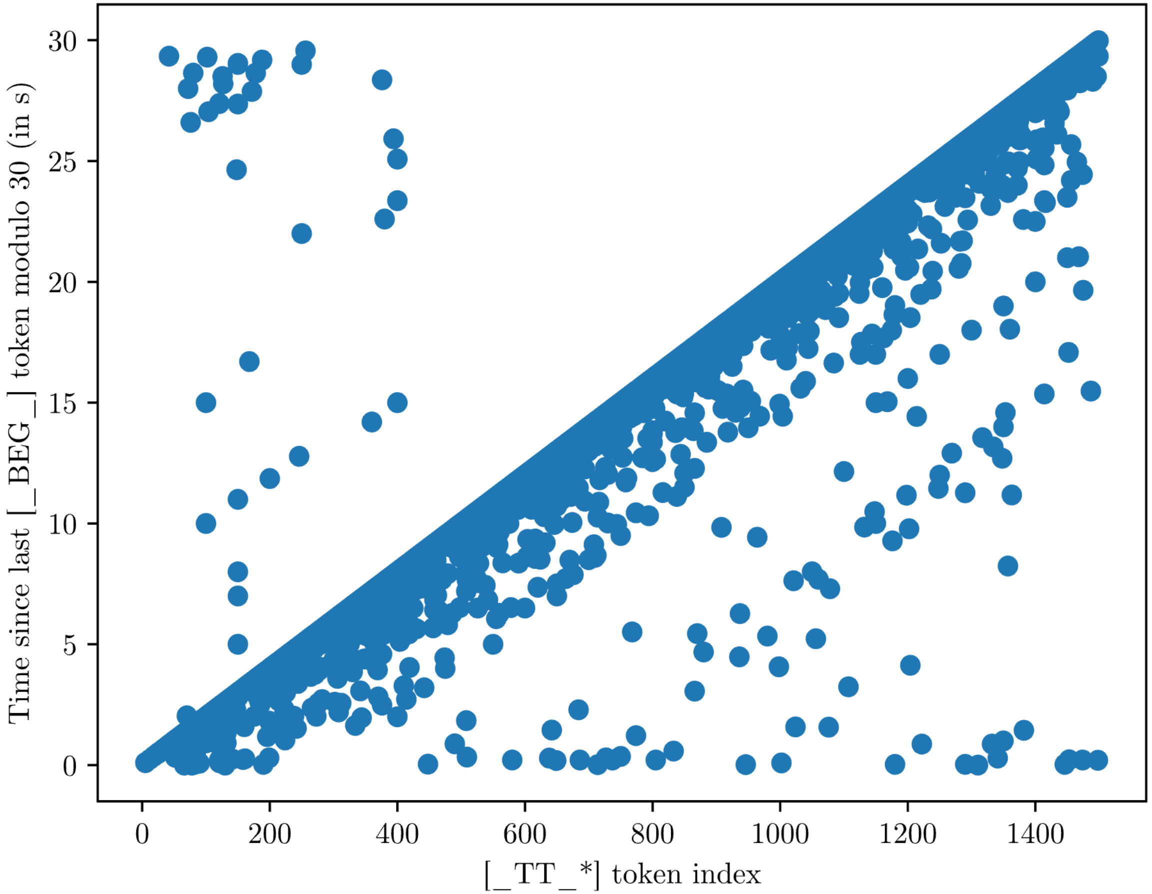


Figure 4: Token indices vs modulated time

MAIN RESULTS (3 papers in one slide)

corpora

- ANGLISH levels predicted
- ISLE levels predicted

tasks

- Language identification task (probability of identification English / L1)
- Mean scores for transcription task

- Metrics : Levensthein distance to expected transcription

Scoring the ANGLISH corpus

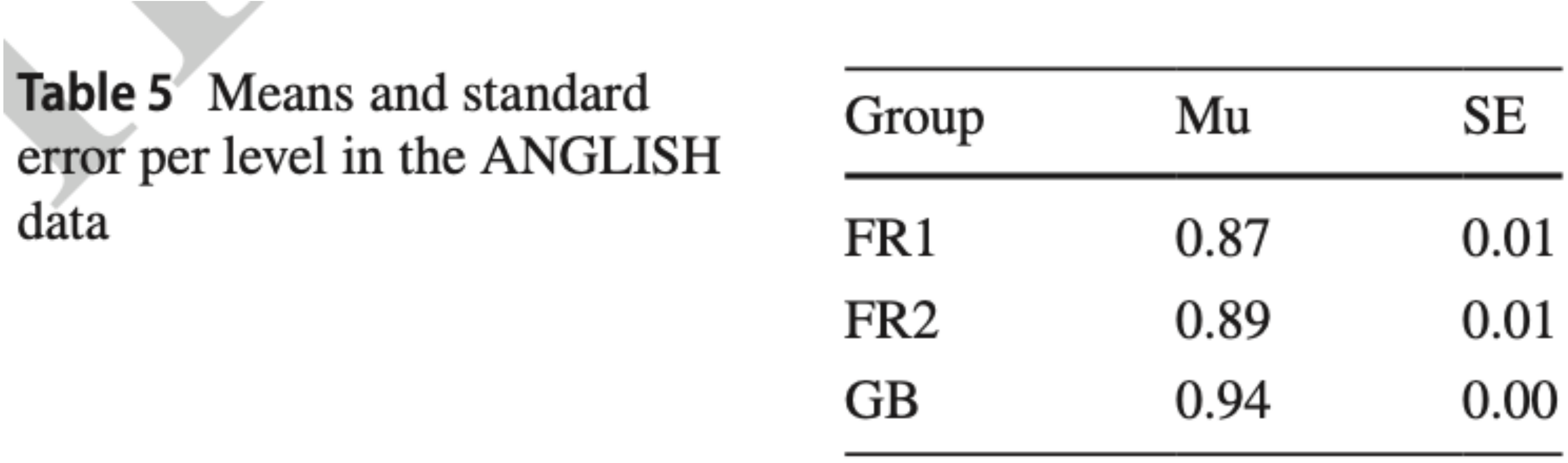


Table 5 Means and standard error per level in the ANGLISH data

Group	Mu	SE
FR1	0.87	0.01
FR2	0.89	0.01
GB	0.94	0.00

Scoring the ANGLISH corpus

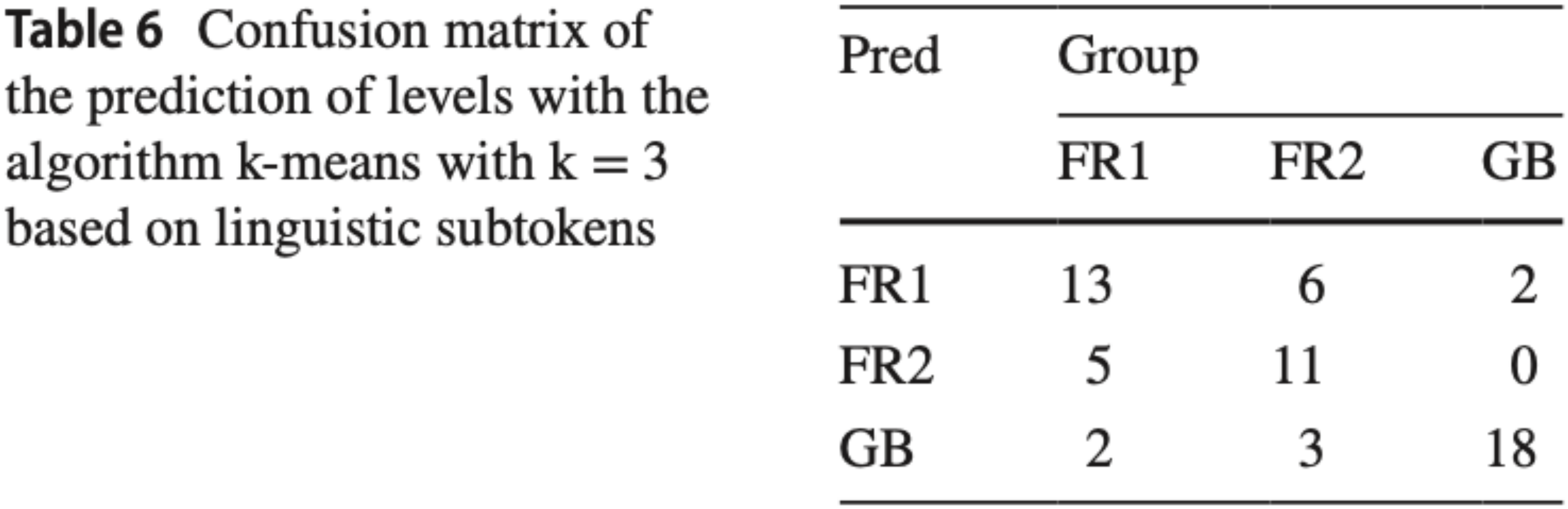


Table 6 Confusion matrix of the prediction of levels with the algorithm k-means with k = 3 based on linguistic subtokens

Pred	Group		
	FR1	FR2	GB
FR1	13	6	2
FR2	5	11	0
GB	2	3	18

« affordance » : ability to capture (mis)realisations locally at the subtoken level

1.Je me ski:	ʒ(ə)m(ə)ski
2.J'aime ce qui:	ʒɛms(ə)ki
3.James Key:]	ʒɛmski
[ʃɔmski] 4.Jomski:	ʒɔmski
5.Jamsky:	ʒamski

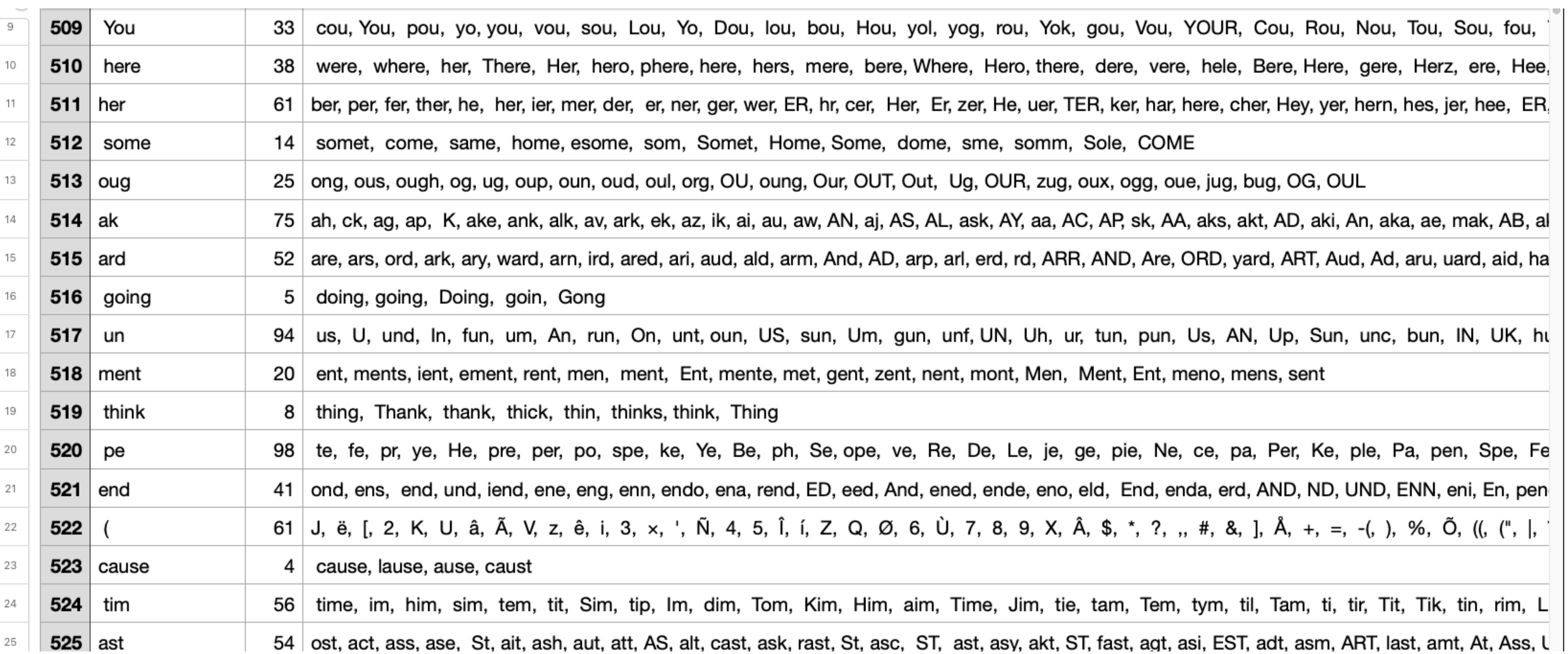
Holistic probability scores vs. Detailed scores for subtokens

-Different phonetic-subtoken mappings for different models

Graphematic affordance: what's in the graphemic representation ('holes' in the Whisper dictionary / net)?

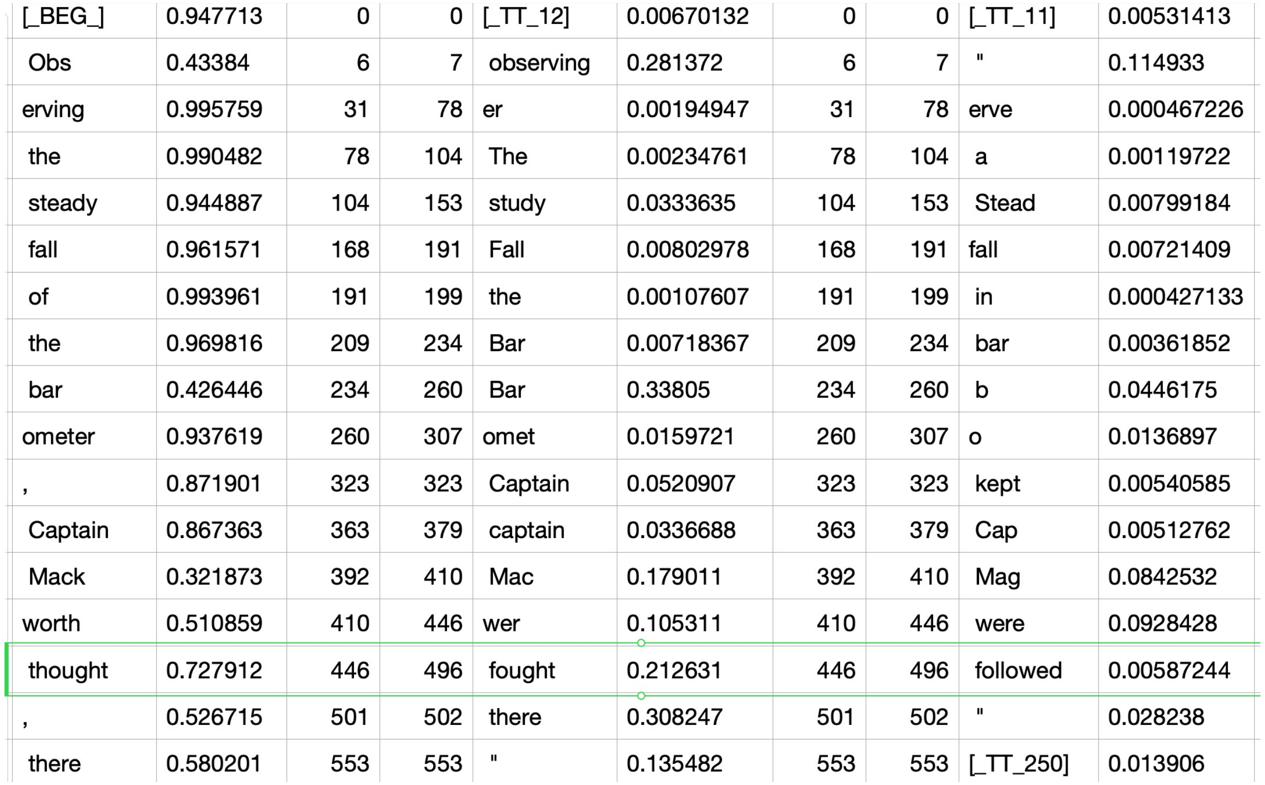
- JANUS WP 2.1
- (pilot) phonological neighbour density

Phonological neighbourhood density (WiP)



509	You	33	cou, You, pou, yo, you, vou, sou, Lou, Yo, Dou, lou, bou, Hou, yol, yog, rou, Yok, gou, Vou, YOUR, Cou, Rou, Nou, Tou, Sou, fou,
510	here	38	were, where, her, There, Her, hero, phere, here, hers, mere, bere, Where, Hero, there, dere, vere, hele, Bere, Here, gere, Herz, ere, Hee,
511	her	61	ber, per, fer, ther, he, her, ier, mer, der, er, ner, ger, wer, ER, hr, cer, Her, Er, zer, He, uer, TER, ker, har, here, cher, Hey, yer, hern, hes, jer, hee, ER,
512	some	14	somet, come, same, home, esome, som, Somet, Home, Some, dome, sme, somm, Sole, COME
513	oug	25	ong, ous, ough, og, ug, oup, oun, oud, oul, org, OU, oung, Our, OUT, Out, Ug, OUR, zug, oux, ogg, oue, jug, bug, OG, OUL
514	ak	75	ah, ck, ag, ap, K, ake, ank, alk, av, ark, ek, az, ik, ai, au, aw, AN, aj, AS, AL, ask, AY, aa, AC, AP, sk, AA, aks, akt, AD, aki, An, aka, ae, mak, AB, al
515	ard	52	are, ars, ord, ark, ary, ward, arn, ird, ared, ari, aud, ald, arm, And, AD, arp, arl, erd, rd, ARR, AND, Are, ORD, yard, ART, Aud, Ad, aru, uard, aid, ha
516	going	5	doing, going, Doing, goin, Gong
517	un	94	us, U, und, In, fun, um, An, run, On, unt, oun, US, sun, Um, gun, unf, UN, Uh, ur, tun, pun, Us, AN, Up, Sun, unc, bun, IN, UK, hu
518	ment	20	ent, ments, ient, ement, rent, men, ment, Ent, mente, met, gent, zent, nent, mont, Men, Ment, Ent, meno, mens, sent
519	think	8	thing, Thank, thank, thick, thin, thinks, think, Thing
520	pe	98	te, fe, pr, ye, He, pre, per, po, spe, ke, Ye, Be, ph, Se, ope, ve, Re, De, Le, je, ge, pie, Ne, ce, pa, Per, Ke, ple, Pa, pen, Spe, Fe
521	end	41	ond, ens, end, und, iend, ene, eng, enn, endo, ena, rend, ED, eed, And, ened, ende, eno, eld, End, enda, erd, AND, ND, UND, ENN, eni, En, pen
522	(	61	J, ë, [, 2, K, U, â, Ã, V, z, ê, i, 3, ×, ', Ñ, 4, 5, Î, í, Z, Q, Ø, 6, Ù, 7, 8, 9, X, Â, $, *, ?, ,, #, &,], Å, +, =, -(,), %, Õ, ((, (", |,
523	cause	4	cause, lause, ause, caust
524	tim	56	time, im, him, sim, tem, tit, Sim, tip, Im, dim, Tom, Kim, Him, aim, Time, Jim, tie, tam, Tem, tym, til, Tam, ti, tir, Tit, Tik, tin, rim, L
525	ast	54	ost, act, ass, ase, St, ait, ash, aut, att, AS, alt, cast, ask, rast, St, asc, ST, ast, asy, akt, ST, fast, agt, asi, EST, adt, asm, ART, last, amt, At, Ass,

WiP : « phonetic » neighbours in alternative predictions



[_BEG_]	0.947713	0	0	[_TT_12]	0.00670132	0	0	[_TT_11]	0.00531413
Obs	0.43384	6	7	observing	0.281372	6	7	"	0.114933
erving	0.995759	31	78	er	0.00194947	31	78	erve	0.000467226
the	0.990482	78	104	The	0.00234761	78	104	a	0.00119722
steady	0.944887	104	153	study	0.0333635	104	153	Stead	0.00799184
fall	0.961571	168	191	Fall	0.00802978	168	191	fall	0.00721409
of	0.993961	191	199	the	0.00107607	191	199	in	0.000427133
the	0.969816	209	234	Bar	0.00718367	209	234	bar	0.00361852
bar	0.426446	234	260	Bar	0.33805	234	260	b	0.0446175
ometer	0.937619	260	307	omet	0.0159721	260	307	o	0.0136897
,	0.871901	323	323	Captain	0.0520907	323	323	kept	0.00540585
Captain	0.867363	363	379	captain	0.0336688	363	379	Cap	0.00512762
Mack	0.321873	392	410	Mac	0.179011	392	410	Mag	0.0842532
worth	0.510859	410	446	wer	0.105311	410	446	were	0.0928428
thought	0.727912	446	496	fought	0.212631	446	496	followed	0.00587244
,	0.526715	501	502	there	0.308247	501	502	"	0.028238
there	0.580201	553	553	"	0.135482	553	553	[_TT_250]	0.013906

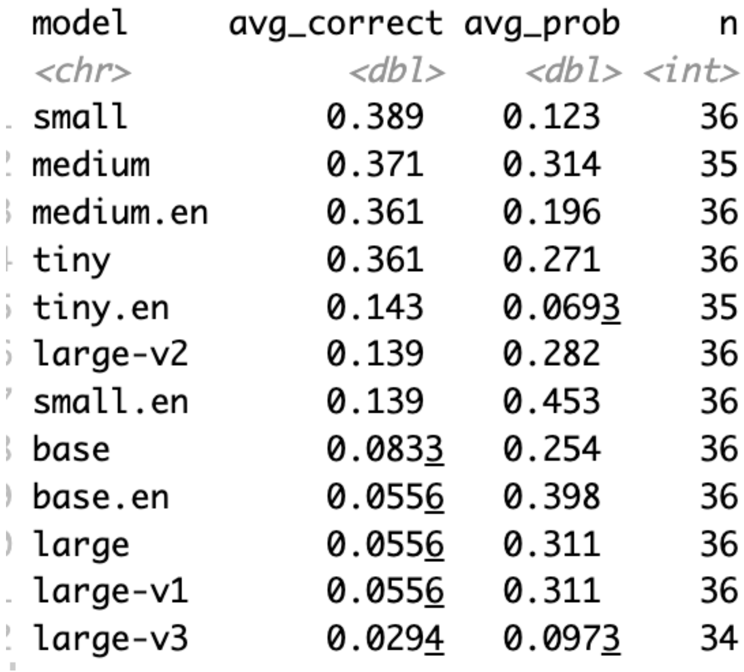
Phonetic sensitivity : subtoken transcription robustness

- Pilot study : VOT (Ballier & Fullerton, 2024, Fullerton & Ballier, to be resubmitted)

- Calibration studies on the signal-to-subtoken mapping : investigating the paradigm: multilingual vs. Native model sensitivity (retranscriptions of the same .wav input)

Using probability as a proxy (work in Progress: Maelle Bourbon & colleagues)

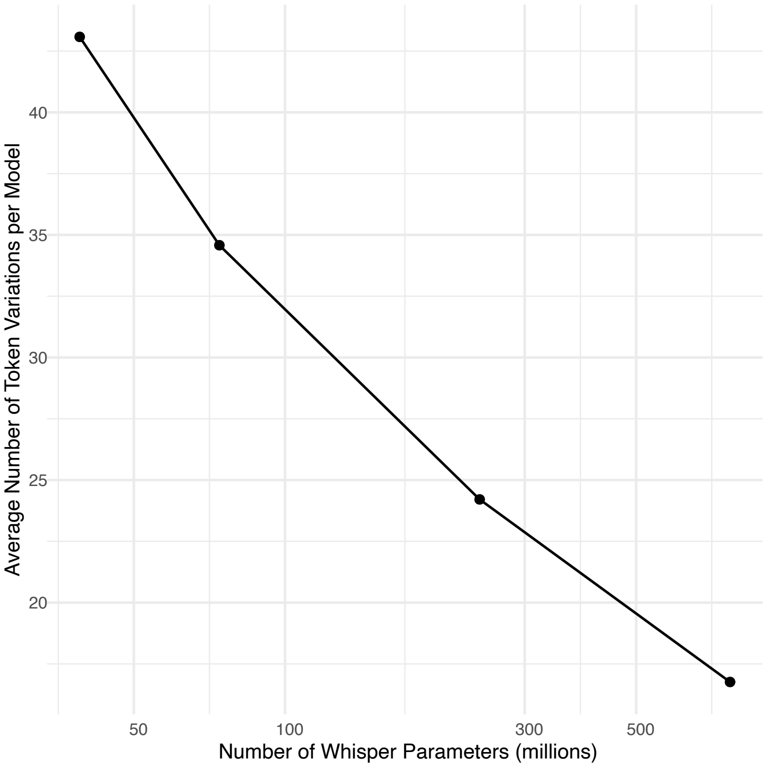
Model sensitivity (Fullerton, in progress)



```
model       avg_correct avg_prob     n
<chr>             <dbl>    <dbl> <int>
small            0.389    0.123     36
medium           0.371    0.314     35
medium.en        0.361    0.196     36
tiny             0.361    0.271     36
tiny.en          0.143    0.0693    35
large-v2         0.139    0.282     36
small.en         0.139    0.453     36
base             0.0833   0.254     36
base.en          0.0556   0.398     36
large            0.0556   0.311     36
large-v1         0.0556   0.311     36
large-v3         0.0294   0.0973    34
```

Role of Size in models for sensitivity?

(character error rate)



INCLSP2023

Model calibration

(Ballier et al. 2024)

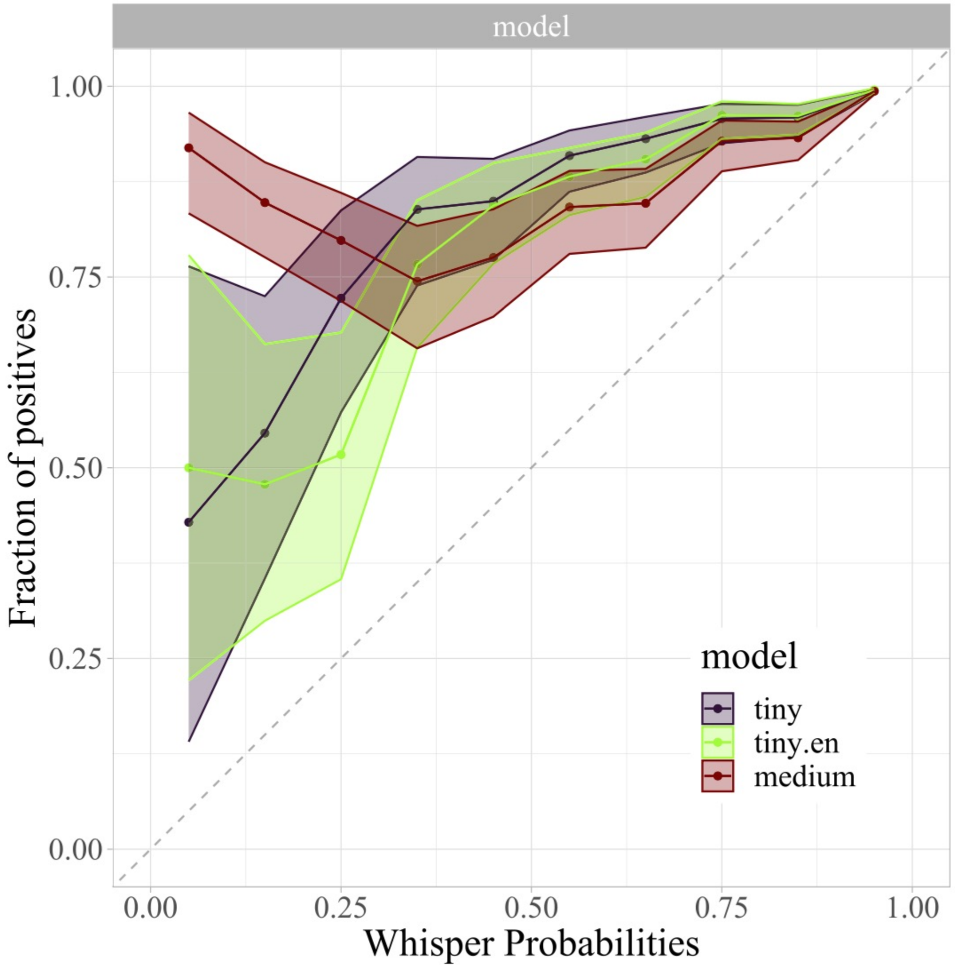
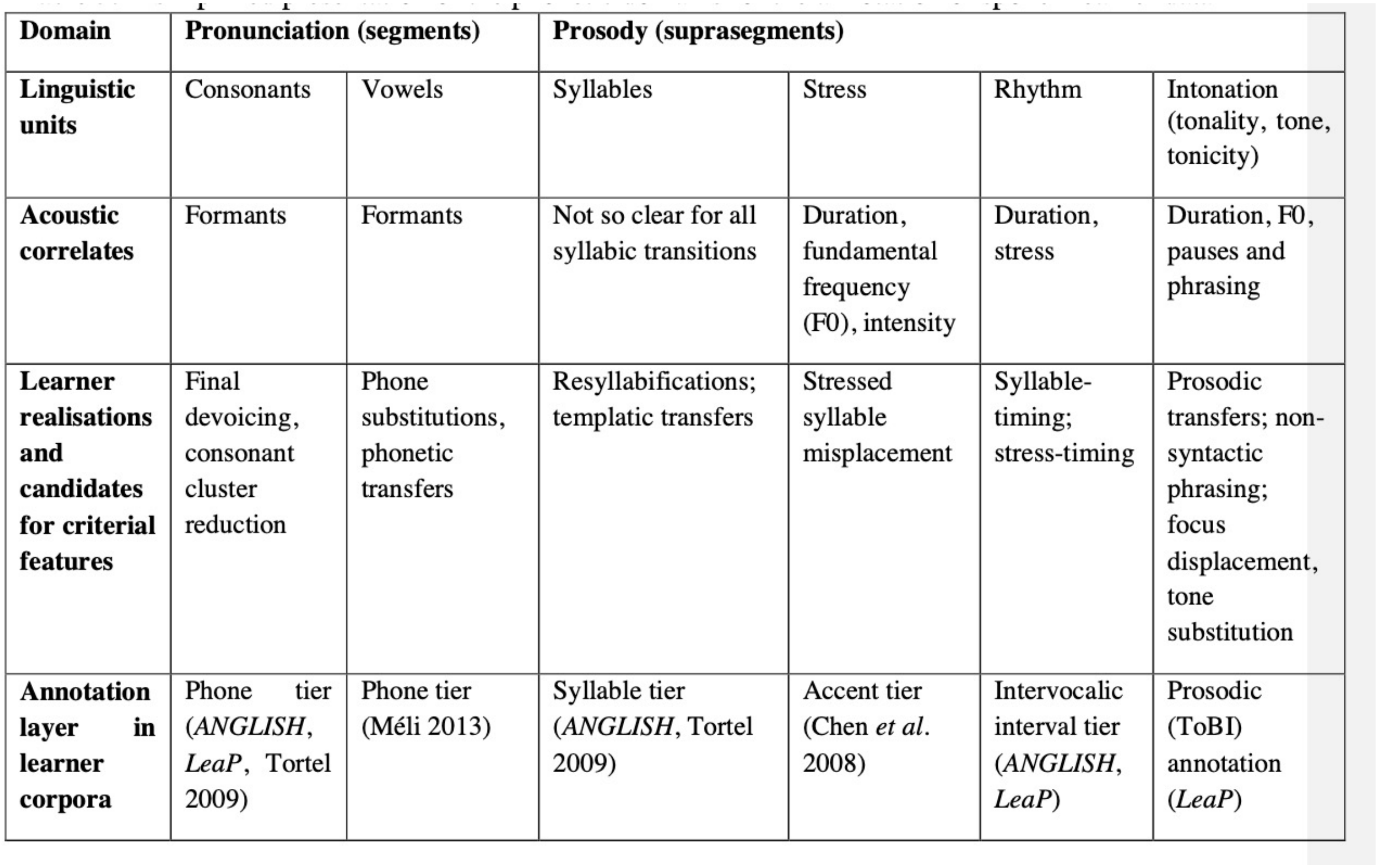


Fig. 4 Calibration curve for three Whisper models for the transcription of the learner #003 from the ISLE corpus

Disc: Speech tokenisers and the issue of discretisation of speech / descriptors

- « criterial feature » (Hawkins & Buttery, 2010) ??
- Discrete phenomenon for CEFR boundaries ?
- Continuous scales for CEFR « descriptors » ??
- RQ1: prosodic domains as criterial features?
- RQ2 Matching speech (sub)tokens with criterial feature?

MDD, LLMs and error typologies (after Ballier& Martin 2015)



Domain	**Pronunciation (segments)**		**Prosody (suprasegments)**			
Linguistic units	Consonants	Vowels	Syllables	Stress	Rhythm	Intonation (tonality, tone, tonicity)
Acoustic correlates	Formants	Formants	Not so clear for all syllabic transitions	Duration, fundamental frequency (F0), intensity	Duration, stress	Duration, F0, pauses and phrasing
Learner realisations and candidates for criterial features	Final devoicing, consonant cluster reduction	Phone substitutions, phonetic transfers	Resyllabifications; templatic transfers	Stressed syllable misplacement	Syllable-timing; stress-timing	Prosodic transfers; non-syntactic phrasing; focus displacement, tone substitution
Annotation layer in learner corpora	Phone tier (*ANGLISH*, *LeaP*, Tortel 2009)	Phone tier (Méli 2013)	Syllable tier (*ANGLISH*, Tortel 2009)	Accent tier (Chen *et al*. 2008)	Intervocalic interval tier (*ANGLISH*, *LeaP*)	Prosodic (ToBI) annotation (*LeaP*)

<-

(RHYTHM) : capturing fluency with Whisper?

- We need filled pause transcriptions for PHON* tasks ()
- Subtokens for *heu / erm / ahem*
- The interlanguage of filled pauses (Chlébowski& Ballier 2022)

Capturing the 3Ts with Whisper?

1. **TONE**
2. **TONICITY**
3. **TONALITY**

TONE: retraining Whisper with Momel INTSINT

Indirect performance punctuation

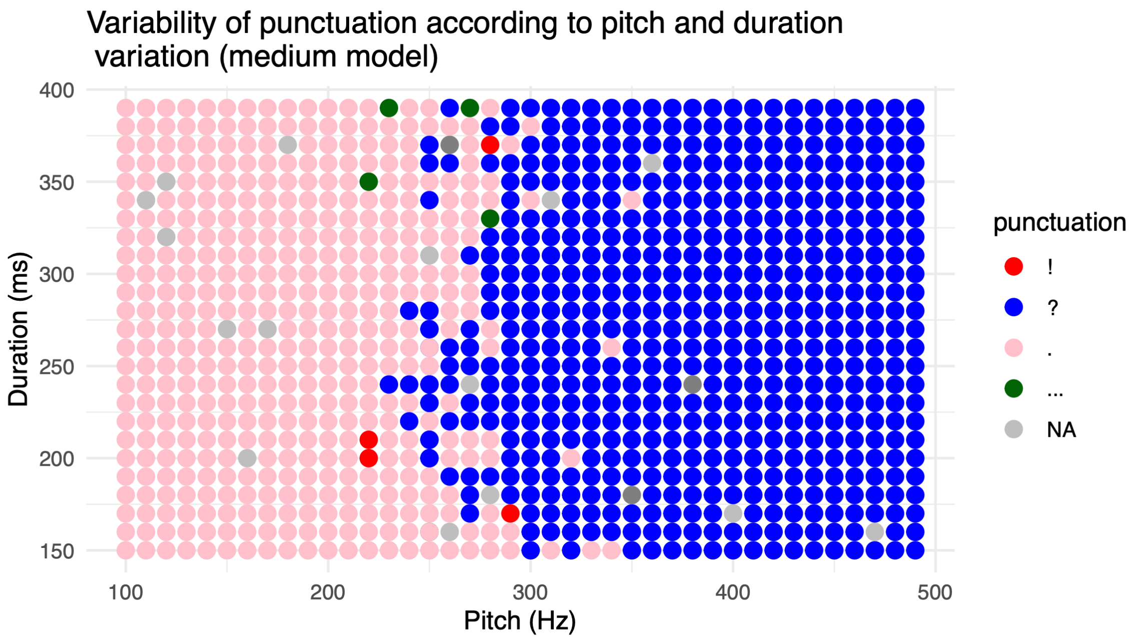
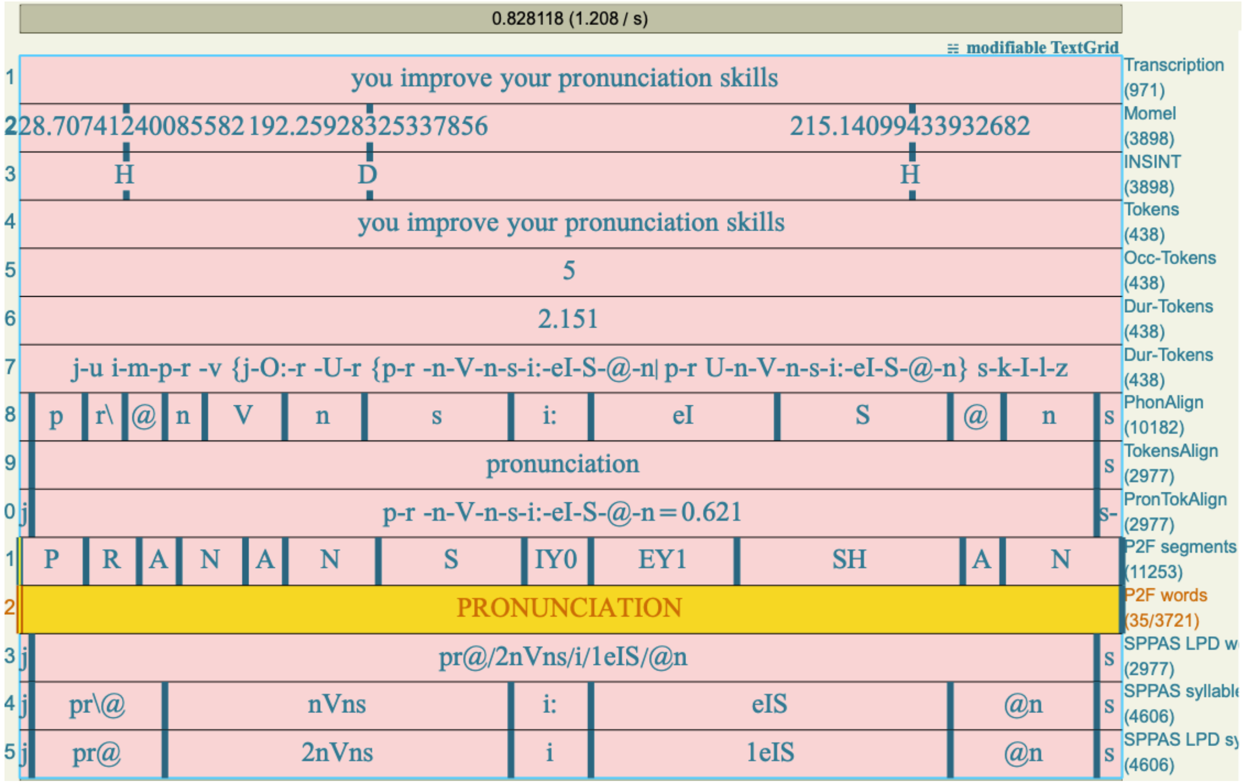


Figure 3: Whisper Response (medium model) for English data as a function of duration (ms) and pitch (Hz)

Janus WP 2.3

https://openreview.net/pdf?id=gCOm8dwzeg

PEASYV pipeline (Ballier & Méli, 2023)



0.828118 (1.208 / s)

1	you improve your pronunciation skills	Transcription (971)
2	228.70741240085582 192.25928325337856 215.14099433932682	Momel (3898)
3	H D H	INSINT (3898)
4	you improve your pronunciation skills	Tokens (438)
5	5	Occ-Tokens (438)
6	2.151	Dur-Tokens (438)
7	j-u i-m-p-r -v {j-O:-r -U-r {p-r -n-V-n-s-i:-eI-S-@-n| p-r U-n-V-n-s-i:-eI-S-@-n} s-k-I-l-z	Dur-Tokens (438)
8	p r\ @ n V n s i: eI S @ n s	PhonAlign (10182)
9	pronunciation s	TokensAlign (2977)
0	j p-r -n-V-n-s-i:-eI-S-@-n = 0.621 s-	PronTokAlign (2977)
1	P R A N A N S IY0 EY1 SH A N	P2F segments (11253)
2	PRONUNCIATION	P2F words (35/3721)
3	j pr@/2nVns/i/1eIS/@n s	SPPAS LPD w (2977)
4	j pr\@ nVns i: eIS @n s	SPPAS syllable (4606)
5	j pr@ 2nVns i 1eIS @n s	SPPAS LPD sy (4606)

PEASYV OUTPUTS

- .PDF diagnostic files
- .CSV Datasets
- INTSINT labels : "[code] the intonation of an utterance by means of an alphabet of 8 discrete symbols constituting a surface phonological representation of the intonation: T (Top), M (mid), B (bottom),H (Higher), L (Lower), S (Same), U (Upstepped), D (Downstepped)" [Hirst 2006].

Fine-tuning for Tones : learning INTSINT

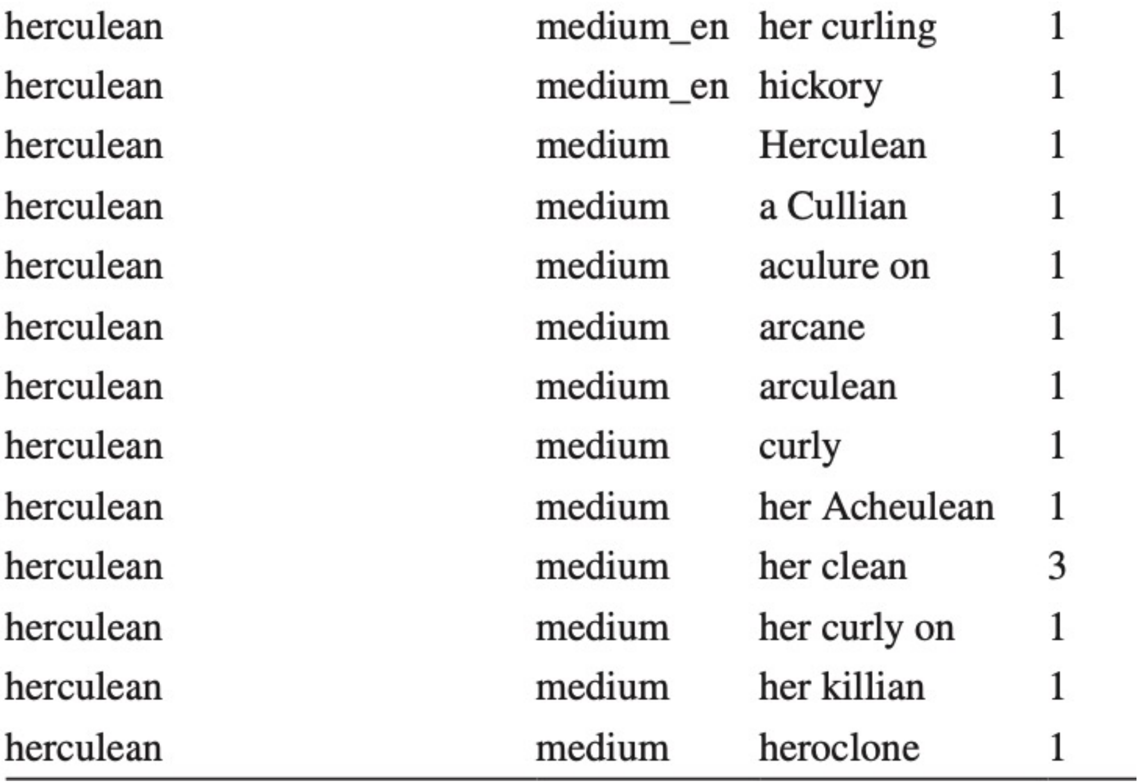
- Learning INSTINT labels with special subtokens
- Pbm : time-stamped (point tier) -> associated to syllables / next subtoken

TONICITY

- testing Whisper sensitivity to shifting tonicity with translations tasks of shifting tonicity : *HE did it. He did it.*
- Partial test with compounds vs. Phrases (*greenhouse vs. Green house*)
- The reanalysis hypothesis (Ballier et al, 2024 *IJST*) : correlation of misplaced stresses and alternative respelling?

- To be discussed / tested : retrain Whisper with capitals for stressed syllables? / AIML tags for prominence???

To be tested : the reanalysis hypothesis



herculean	medium_en	her curling	1
herculean	medium_en	hickory	1
herculean	medium	Herculean	1
herculean	medium	a Cullian	1
herculean	medium	aculure on	1
herculean	medium	arcane	1
herculean	medium	arculean	1
herculean	medium	curly	1
herculean	medium	her Acheulean	1
herculean	medium	her clean	3
herculean	medium	her curly on	1
herculean	medium	her killian	1
herculean	medium	heroclone	1

Frequency constraints :

Is reanalysis more frequent for free (vs. bound) tokens?

- Distribution of polysyllables for the CONRAD dataset (*IJST*)

Free subtokens: Maximum free tokens length: 13 Average free tokens length: 4.35

- 5 characters: *dirty, about*
- 6 characters: *steady, simple, belief, wisdom, county*
- 7 characters: *Captain, thought, weather*

- 8 characters: *knocking, implying, moderate, informed, circular*
- 9 characters: *precisely, authority, questions, conscious, vanishing*
- 10 characters: *experience, moderately, discomfort, atmosphere,*
- 11 characters: *disturbance, information, necessarily*
- 12 characters: *accomplished, catastrophic*
- 13 characters*:* *comprehension, uncomfortable*

Sample bound tokens for each length:

- 1 character: *,, r, ,, ., .*
- 2 characters: *Wh, ir, 's, .", ac*
- 3 characters: *Obs, âĢĶ, lys,*

- 4 characters: *aman, ated, ones, oons*
- 5 characters: *isive, lling, ously, oping*
- 6 characters: *erving, ometer*

- 7 characters: *putable, ulating*

TONALITY (Chunking)

- To be tested : TT special tokens for time stamps in the ANGLISH corpus
- Partial proxy : punctuation signs
- Big issues with Whisper « segments »

Fine-tuning with the Aix-Marsec corpus : minor | vs. Major || boundaries (Arnold & Ballier, 2019 ⟨hal-04012540⟩)

TONALITY

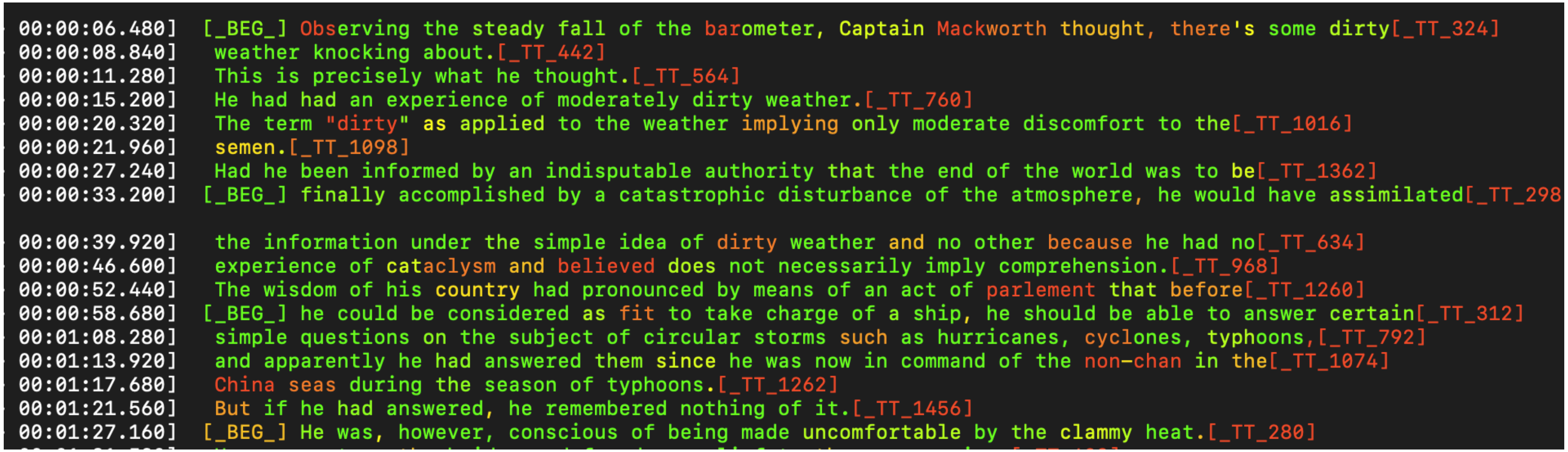
- Can we use the Whisper time stamps as a correlate for tonality?

[TT_] as a phonetic boundary ??

, and . as phonological boundary ??

- Is a weak probability associated to [TT_] the signal that the chunking is a bit unfortunate?

TONALITY (tiny.en model)



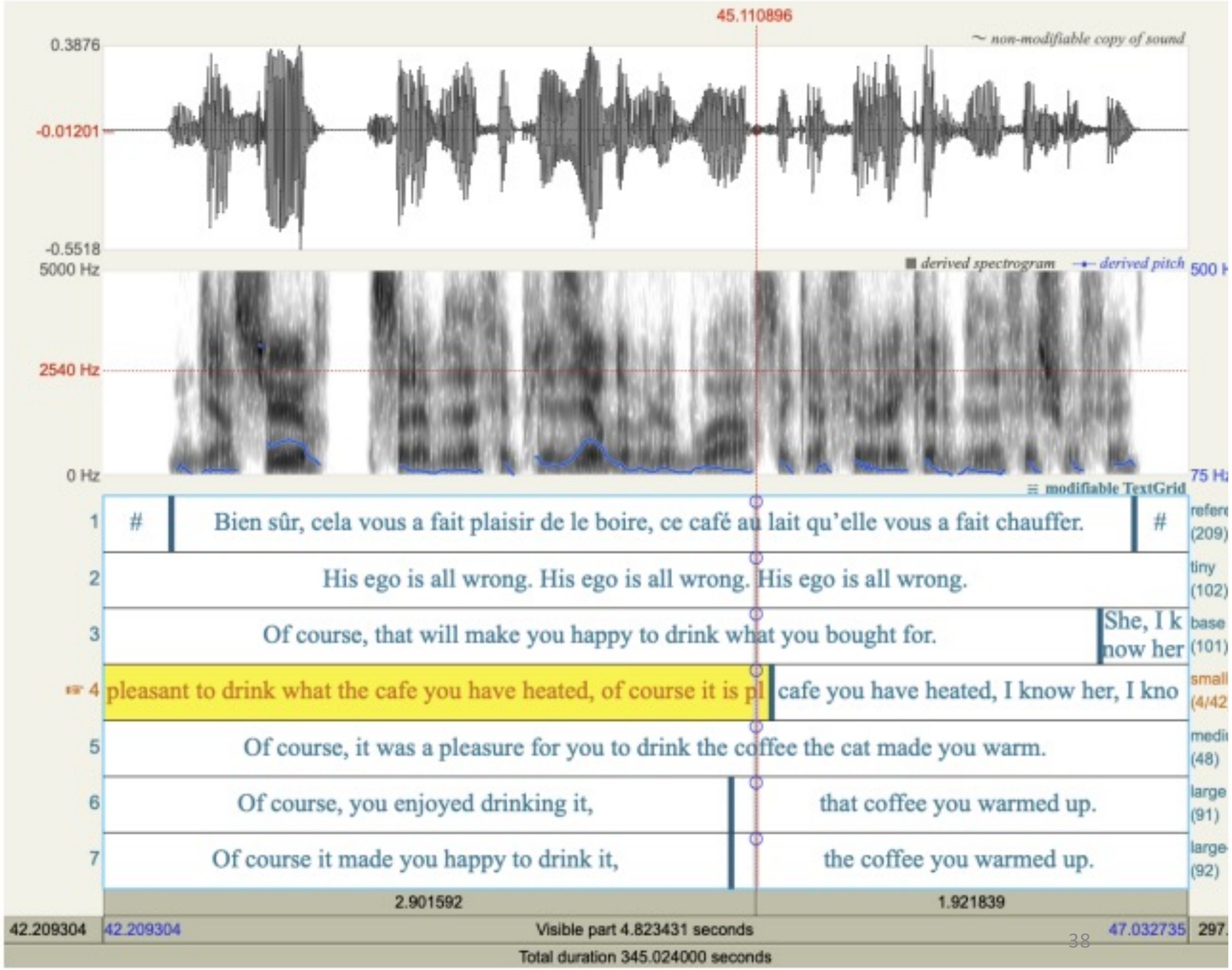
```
00:00:06.480]  [_BEG_] Observing the steady fall of the barometer, Captain Mackworth thought, there's some dirty[_TT_324]
00:00:08.840]   weather knocking about.[_TT_442]
00:00:11.280]   This is precisely what he thought.[_TT_564]
00:00:15.200]   He had had an experience of moderately dirty weather.[_TT_760]
00:00:20.320]   The term "dirty" as applied to the weather implying only moderate discomfort to the[_TT_1016]
00:00:21.960]   semen.[_TT_1098]
00:00:27.240]   Had he been informed by an indisputable authority that the end of the world was to be[_TT_1362]
00:00:33.200]  [_BEG_] finally accomplished by a catastrophic disturbance of the atmosphere, he would have assimilated[_TT_298

00:00:39.920]   the information under the simple idea of dirty weather and no other because he had no[_TT_634]
00:00:46.600]   experience of cataclysm and believed does not necessarily imply comprehension.[_TT_968]
00:00:52.440]   The wisdom of his country had pronounced by means of an act of parlement that before[_TT_1260]
00:00:58.680]  [_BEG_] he could be considered as fit to take charge of a ship, he should be able to answer certain[_TT_312]
00:01:08.280]   simple questions on the subject of circular storms such as hurricanes, cyclones, typhoons,[_TT_792]
00:01:13.920]   and apparently he had answered them since he was now in command of the non-chan in the[_TT_1074]
00:01:17.680]   China seas during the season of typhoons.[_TT_1262]
00:01:21.560]   But if he had answered, he remembered nothing of it.[_TT_1456]
00:01:27.160]  [_BEG_] He was, however, conscious of being made uncomfortable by the clammy heat.[_TT_280]
```

Qualitative Analysis:
Transcription

- Variability of Whisper segmentation of time intervals across models
- Absence of pause (#)
- SRT files: not easy to

use for subtitles

(overlap for base model)



1 # | Bien sûr, cela vous a fait plaisir de le boire, ce café au lait qu'elle vous a fait chauffer. | #

2 His ego is all wrong. His ego is all wrong. His ego is all wrong.

3 Of course, that will make you happy to drink what you bought for. | She, I know her

4 pleasant to drink what the cafe you have heated, of course it is pl | cafe you have heated, I know her, I kno

5 Of course, it was a pleasure for you to drink the coffee the cat made you warm.

6 Of course, you enjoyed drinking it, | that coffee you warmed up.

7 Of course it made you happy to drink it, | the coffee you warmed up.

Candidates for criterial features ?

- stress shift

A COMMON ROADMAP ?

Test suites (Janus WP 2.1)

- Holistic grading vs. Granular evaluation :
- The Speak&Improve and SpeechOcean datasets

Experiments on duration

- Duration ablation (30 seconds of speech) (Myssik 2011)

ANNOTATED datasets :

- The PARAAF corpus / dataset (UPCité) https://emmanuelferragne.com/project/paraaf/

TO BE TESTED…

Start.boldvoice.com

- BOLDVOICE.COM

REFERENCES

- Ballier, N., Arnold, T., Méli, A., Fullerton, T., & Yunès, J. B. (2024). Whisper for L2 speech scoring. *International Journal of Speech Technology*, *27*, 923-934. https://taylorarnold.org/pdf/2024-whisperl2.pdf <IJST2024>
- Ballier, N., Méli, A., Amand, M., & Yunès, J. B. (2023). Using Whisper LLM for Automatic Phonetic Diagnosis of L2 Speech, a Case Study with French Learners of English. In *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)* (pp. 282-292). https://aclanthology.org/2023.icnlsp-1.30.pdf
- Ballier, N., & Méli, A. (2024, October). Investigating Acoustic Correlates of Whisper Scoring for L2 Speech Using Forced Alignment with the Italian Component of the ISLE corpus. In *13th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2024)* (Vol. 13). https://aclanthology. org/volumes/2024. nlp4call-1/

- Ballier, N., Burin, L., Namdarzadeh, B., Ng, S., Wright, R., & Yunès, J. B. (2024). Probing whisper predictions for French, English and Persian transcriptions. In *7the International Conference on Natural Language and Speech Processing* (Vol. 7, pp. 129-138). Association for Computational Linguistics. https://hal.science/hal-04912112/document
- Namdarzadeh, B., & Ballier, N. (2024, draft). Audio LLM subtokens as encapsulated" knowledge": the case of Persian subtoken graphemic representations in Whisper. In *Grapholinguistics in the 21st century—From graphemes to knowledge*. https://hal.science/hal-04927138/document

Papers in the making

- VOT and Whisper model sensitivity

Probing the relevance threshold of Whisper predictions for the transcription task of Persian,French and English

- Phonology of semantic speech tokens

CONCLUSION : Q&A

Demo?

🚀 Let's discuss!

Open access models, local analysis possible

Thanks for the invitation

Special thanks to Sylvain for organising this!!