

English and French VOT of Initial Stops in Bilingual and L2 Speakers:

Comparing Automatic and Manual Measurements



Marguerite CAMERON

Martine Adda-Decker & Barbara Kühnert



My Thesis Project

- Analyzing the pronunciation of French-English speakers in a multi-lingual context
- Using Voice Onset Time (VOT) as a convenient evaluation of speaker performance and way to compare my findings to existing literature
- Thousands of tokens = need for automation

Dr. VOT & The Present Study

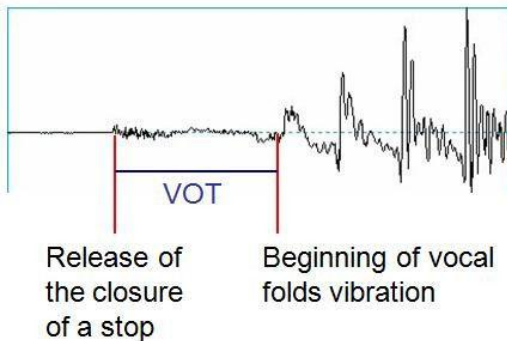
- Shrem, Goldrick, & Keshet (2019) – Deep-learning for measuring VOT intervals of word-initial stop consonants in naturalistic speech
- Tested as being more accurate than the authors' previous model
- Capable of measuring both positive and negative VOT

Goals :

- Confirm the accuracy of Dr.VOT on French and bilingual data
- Ultimately apply Dr.VOT to corpora of spontaneous speech in FR/EN

Voice Onset Time (VOT)

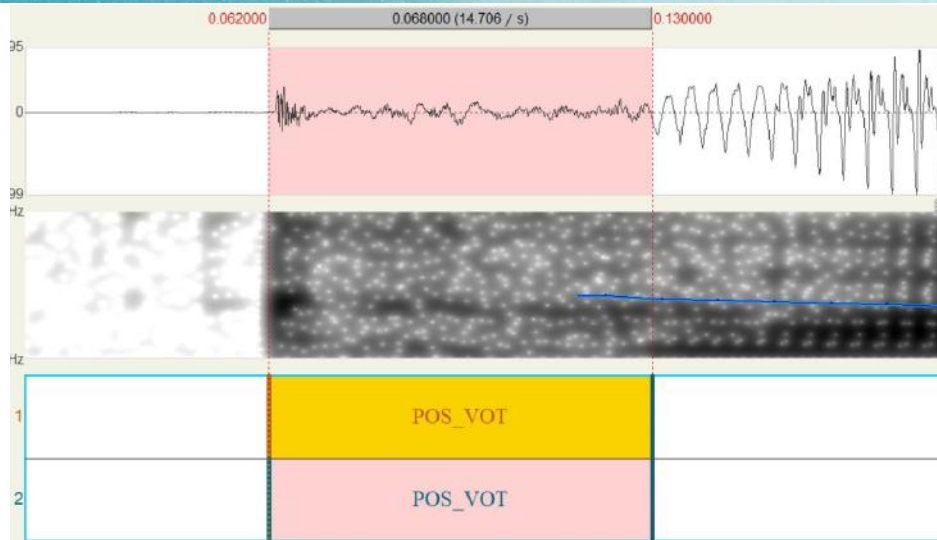
- The period between the release of a stop consonant and the onset of vocal fold vibration (voicing)
- Positive VOT : voicing starts after the stop release
- Negative VOT : voicing starts before the stop release



<https://splab.net/apd/k400/>

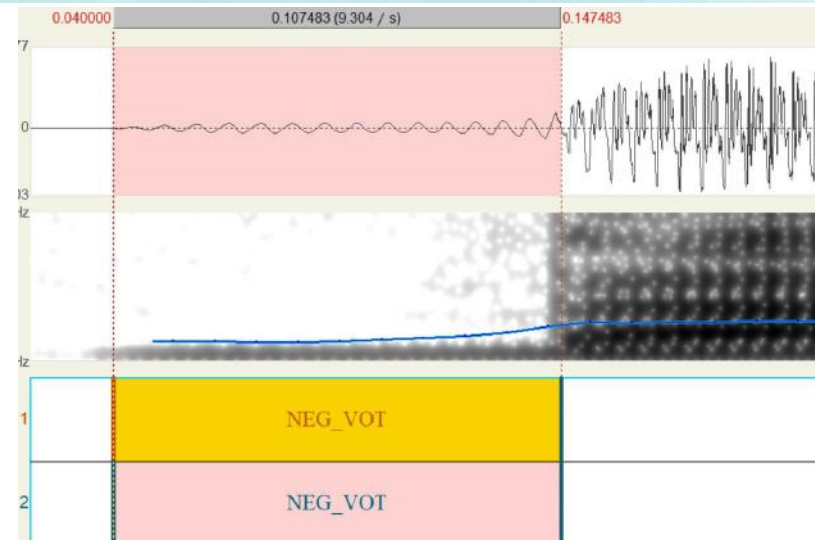
Voice Onset Time (VOT)

Positive VOT



"calf" in English

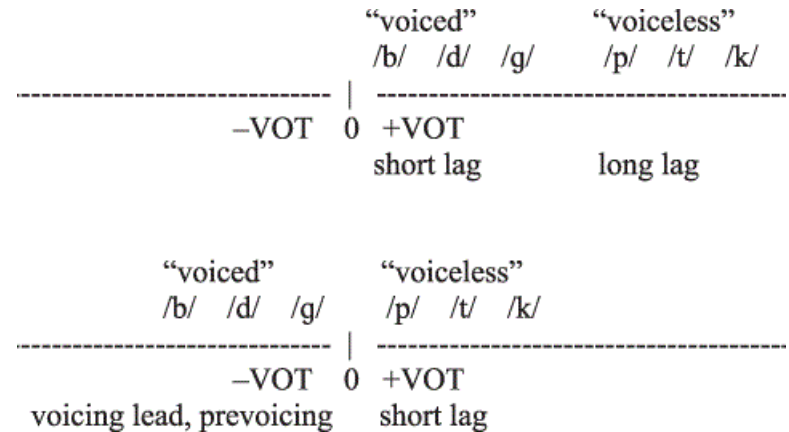
Negative VOT



"bague" in French

Word-Initial VOT as a Comparison Tool

- English and French both use the phonemes /p t k b d g/
- « voiced » vs « voiceless » thresholds for word-initial stop consonants are language-dependent
- English : short lag for voiced stops and long lag for voiceless stops (often aspirated)
- French : prevoicing for voiced stops and short lag for voiceless stops (generally unaspirated)

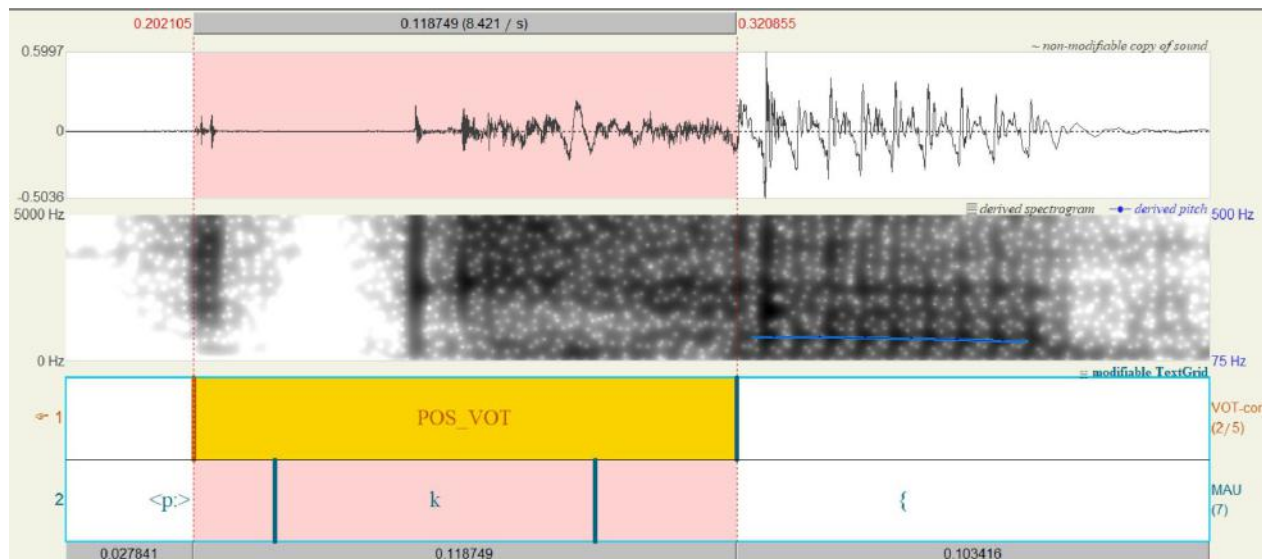


Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). EFFECTS OF AGE AND SPEAKING RATE ON VOICE ONSET TIME: The Production of Voiceless Stops by Near-Native L2 Speakers. *Studies in Second Language Acquisition*, 37(1), 71-100. doi:10.1017/S0272263114000151

VOT of Voiceless Consonants

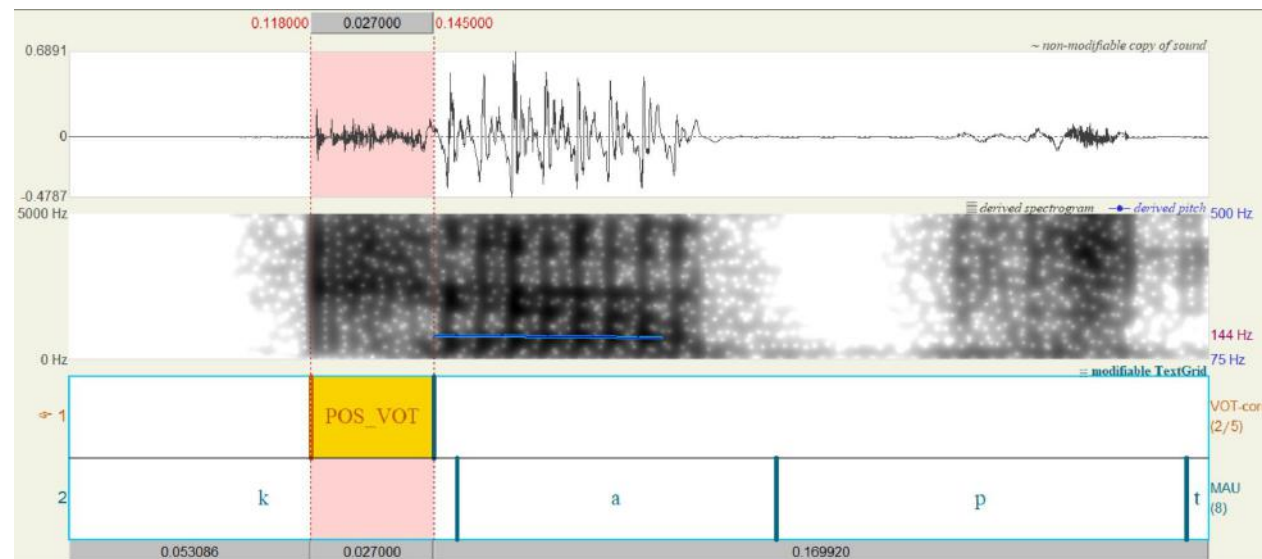
EN vs FR

FEC Corpus



"captor" EN

VOT = 119ms
(long lag)



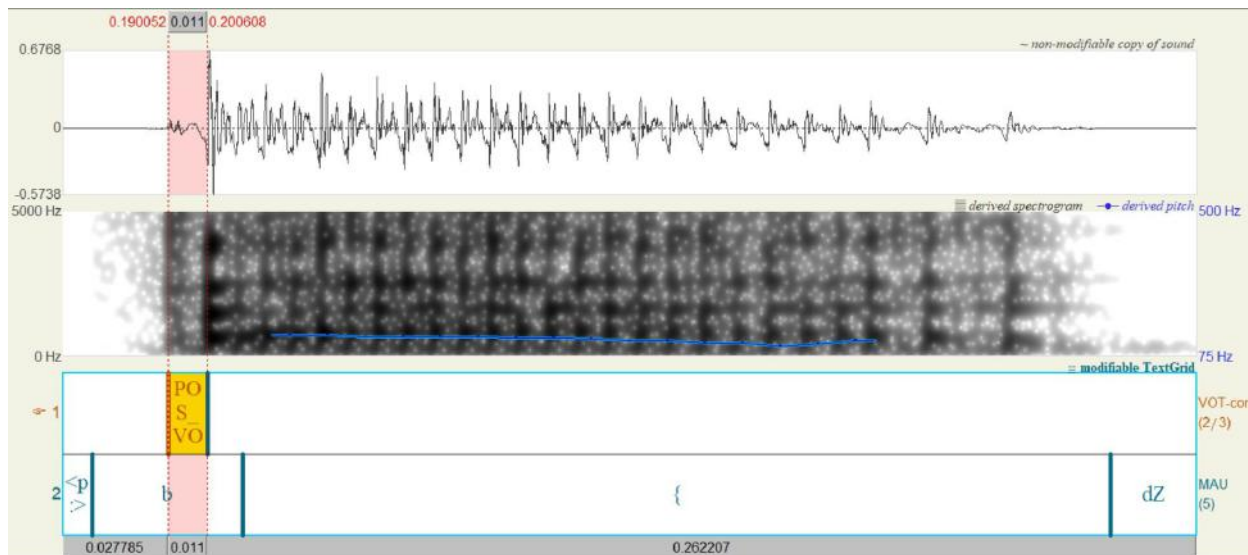
"capteur" FR

VOT = 27ms
(short lag)

VOT of Voiced Consonants

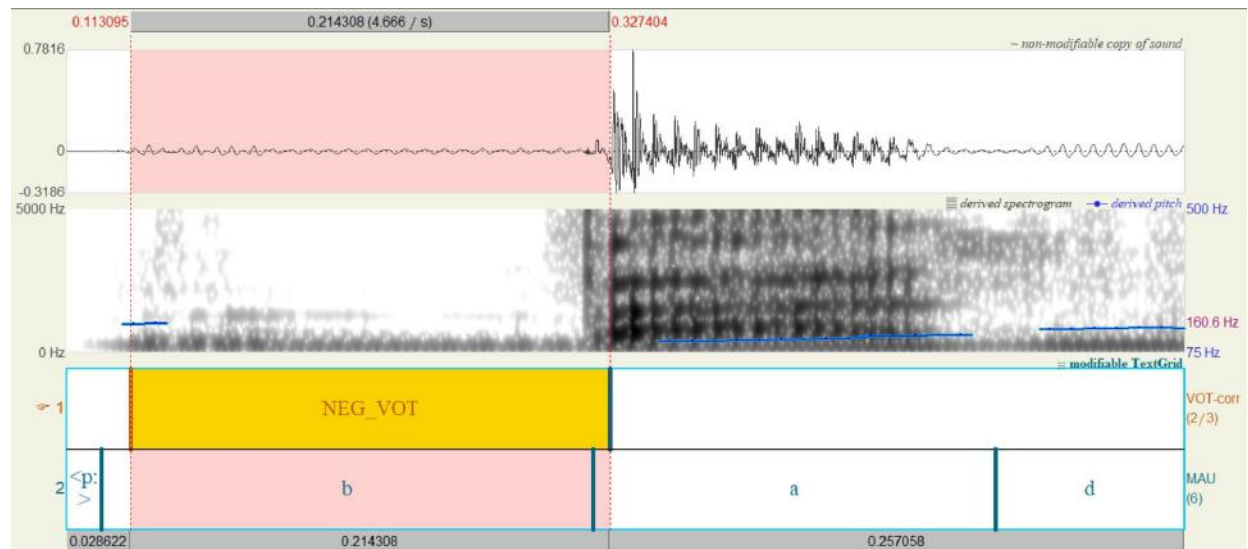
EN vs FR

FEC Corpus



"badge" EN

VOT = 11ms
(short lag)



"badge" FR

VOT = -214ms
(pre-voicing)

French-English Cognates Corpus (FEC Corpus)

- Isolated words, read speech
- 1148 tokens (968 English and 180 French)
- Word-initial stops in a Consonant-Vowel (CV) context
- 6 speakers (3 English L1, 2 French L1, 1 English L2)
- Corpus designed for teaching pronunciation (cleanaccent.com)

- Annotated by Dr.VOT, then verified by hand.
- Automatic measures <5ms different than manual measures were not adjusted

Are Expected Differences Between English and French VOT Observed ?

Subset	P	P	T	T	K	K	B	B	D	D	G	G
# of Tokens	174	16	101	8	199	55	270	55	134	8	24	2
Manual Mean	87.7	20.3	107.08	22.37	94.5	36.6	-43.2	-153.3	-36.6	-112.1	-28.9	-127.5
Auto Mean	85.7	20.3	103.5	22.37	91.9	34.6	-39.5	-96.1	-32.9	-108.8	-24.8	-117.5

EN

FR

Comparison of English vs French manual and automatic VOT measurements on the FEC corpus

Are Mean Automatic VOT Measures Significantly Different from Manual Corrections ? French & English Data

Subset	POS_VOT	NEG_VOT	P	T	K	B	D	G
# of Tokens	821	225	190	109	254	325	142	26
Manual Mean	66.1	-137.8	87.7	100.9	82.0	-61.8	-40.9	-36.5
Auto Mean	61.1	-116.1	85.7	103.6	79.5	-39.5	-37.2	-32.0
P-value	0.02597	7.075e-07	0.5089	0.5562	0.4285	0.00046	0.6968	0.8305
Significant?	Yes	Yes	No	No	No	Yes	No	No

Comparison of manual and automatic VOT measurements on the FEC corpus (English and French data combined), tested for significance by phoneme.

Note: POS_VOT and NEG_VOT are not directly correlated to voiceless and voiced consonants

Are Mean VOT Measures Significantly Different for English Data ?

Subset	P	T	K	B	D	G
# of Tokens	174	101	199	270	134	24
Manual Mean	87.7	107.08	94.5	-43.2	-36.6	-28.9
Auto Mean	85.7	103.5	91.9	-39.5	-32.9	-24.8
P-value	0.5089	0.3968	0.3638	0.5503	0.7035	0.8502
Significant?	no	no	no	no	no	no

Comparison of manual and automatic VOT measurements on the FEC corpus (English data), tested for significance by phoneme.

Are Mean VOT Measures Significantly Different in French Data ?

Subset	P	T	K	B	D	G
# of Tokens	16	8	55	55	8	2
Manual Mean	20.3	22.37	36.6	-153.3	-112.1	-127,5
Auto Mean	20.3	22.37	34.6	-96.1	-108.8	-117,5
P-value	1	1	0.3719	1.669e-06	0.8084	0.6521
Significant?	no	no	no	yes	no	no

Comparison of manual and automatic VOT measurements on the FEC corpus (French data), tested for significance by phoneme.

Errors and Limitations of Dr.VOT

- Important shortcomings :

Vastly incorrect intervals of negative VOT in French data

Regular exclusion of release aspiration for voiced English stops

- Minor shortcomings :

Automatic intervals often a few milliseconds shorter than manual corrections

Rare mislabeling of « positive » or « negative » VOTs

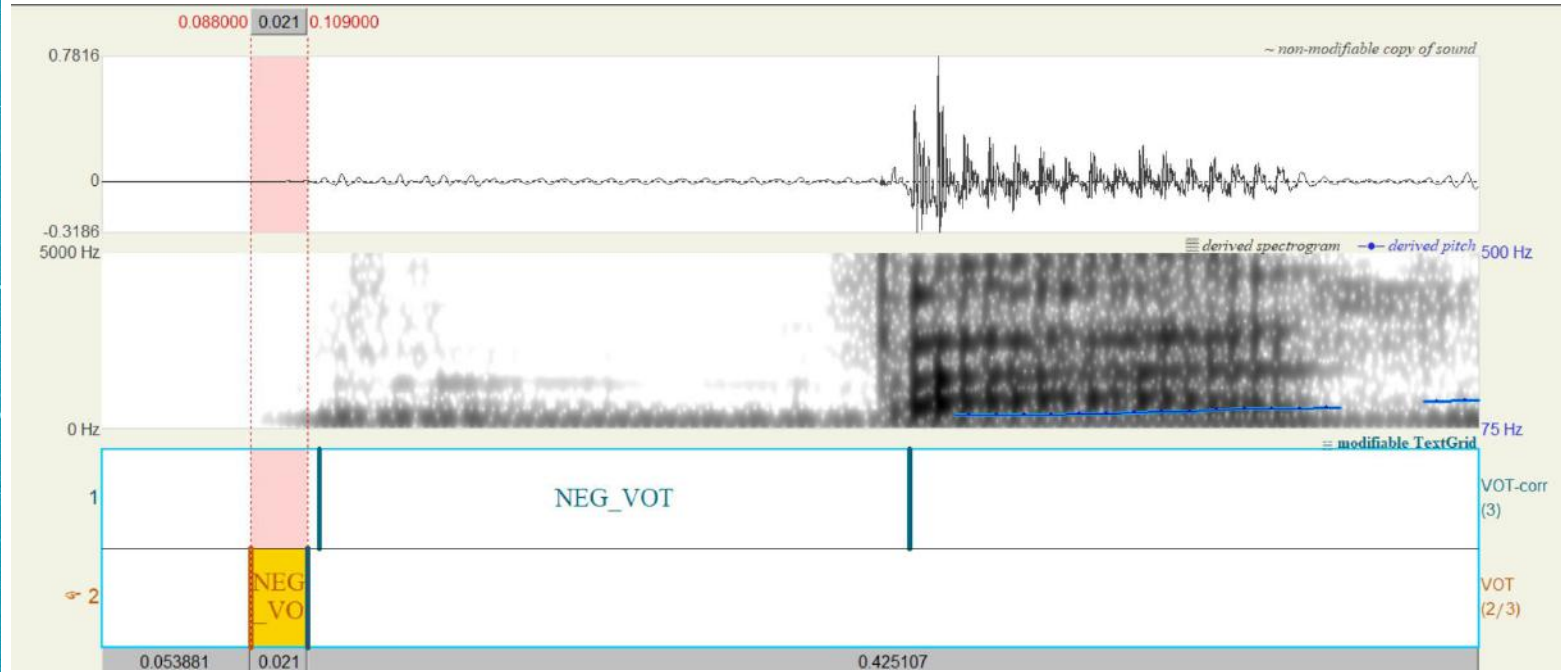
Incorrect Interval for French NEG VOT

« badge »

French, FEC Corpus

Tier 1 : Manual VOT

Tier 2 : Auto VOT



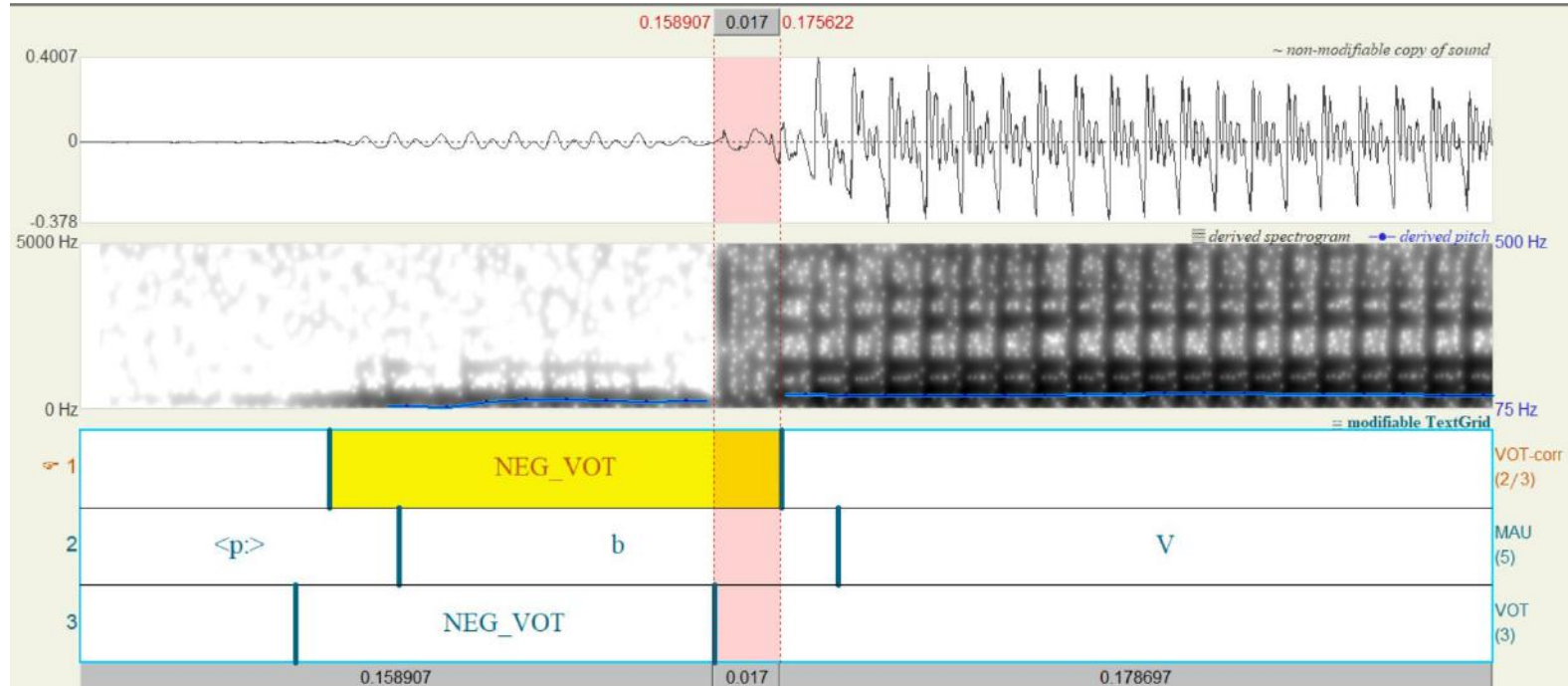
Correction to Include Burst Aspiration

« buzz »

English, FEC Corpus

Tier 1 : Manual VOT

Tier 3 : Auto VOT



The VOT interval measured by DrVOT occasionally omitted the release burst of voiced consonants, typically yielding only a small difference with the manual measure.

Such errors were still systematically corrected, as the release burst is part of the consonants' articulation, and it could be interesting to see to what degree this affected VOT measurements

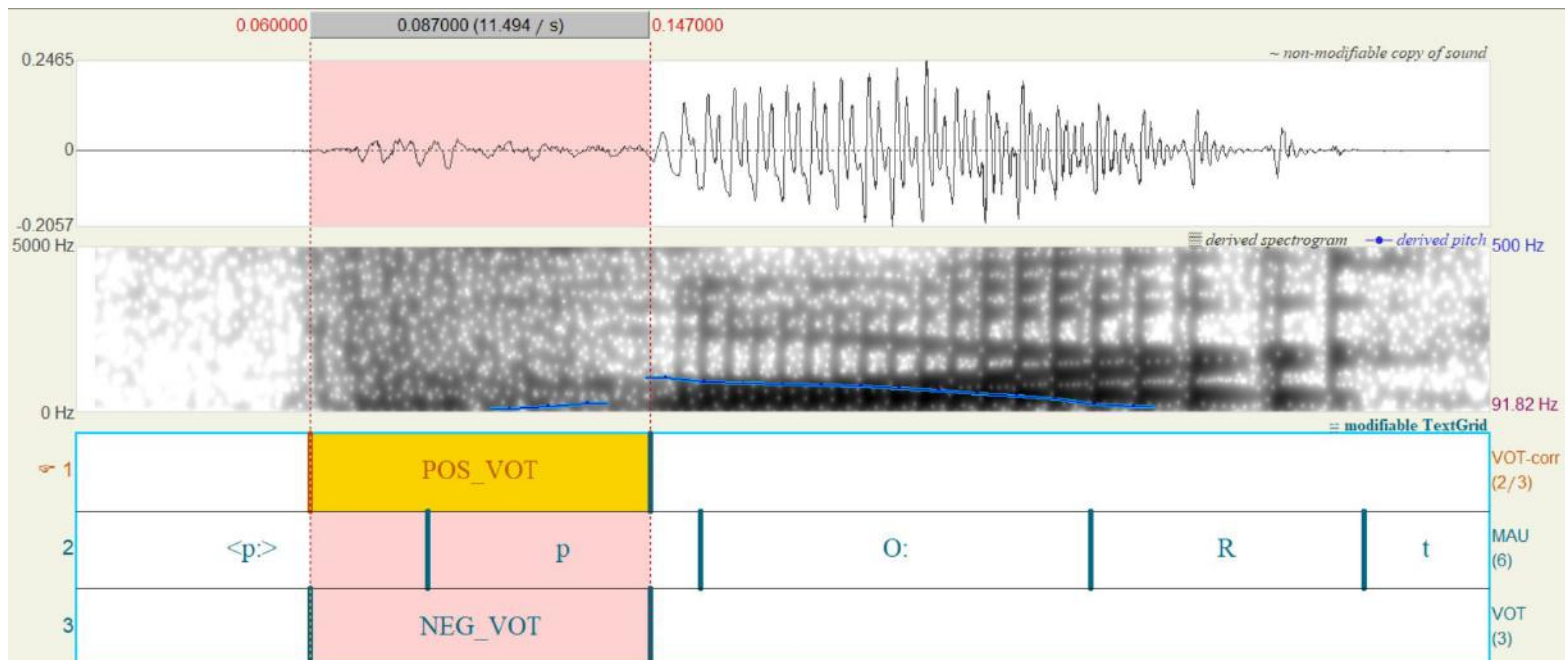
Correcting VOT Type Label

« port »

English, FEC Corpus

Tier 1 : Manual VOT

Tier 3 : Auto VOT



Total POS_VOT that should be labelled as NEG_VOT : 11

Total NEG_VOT that should be labelled as POS_VOT : 7

Dr. VOT Going Forward in My Thesis Project

- Dr. VOT is limited to word-initial consonants, and is therefore not usable for word-median consonants
- Dr. VOT is so far untested on continuous speech
- It will be necessary to verify VOT intervals of French data
- Dr. VOT was trained on English data
- Sheer interval length is a limited perspective of VOT

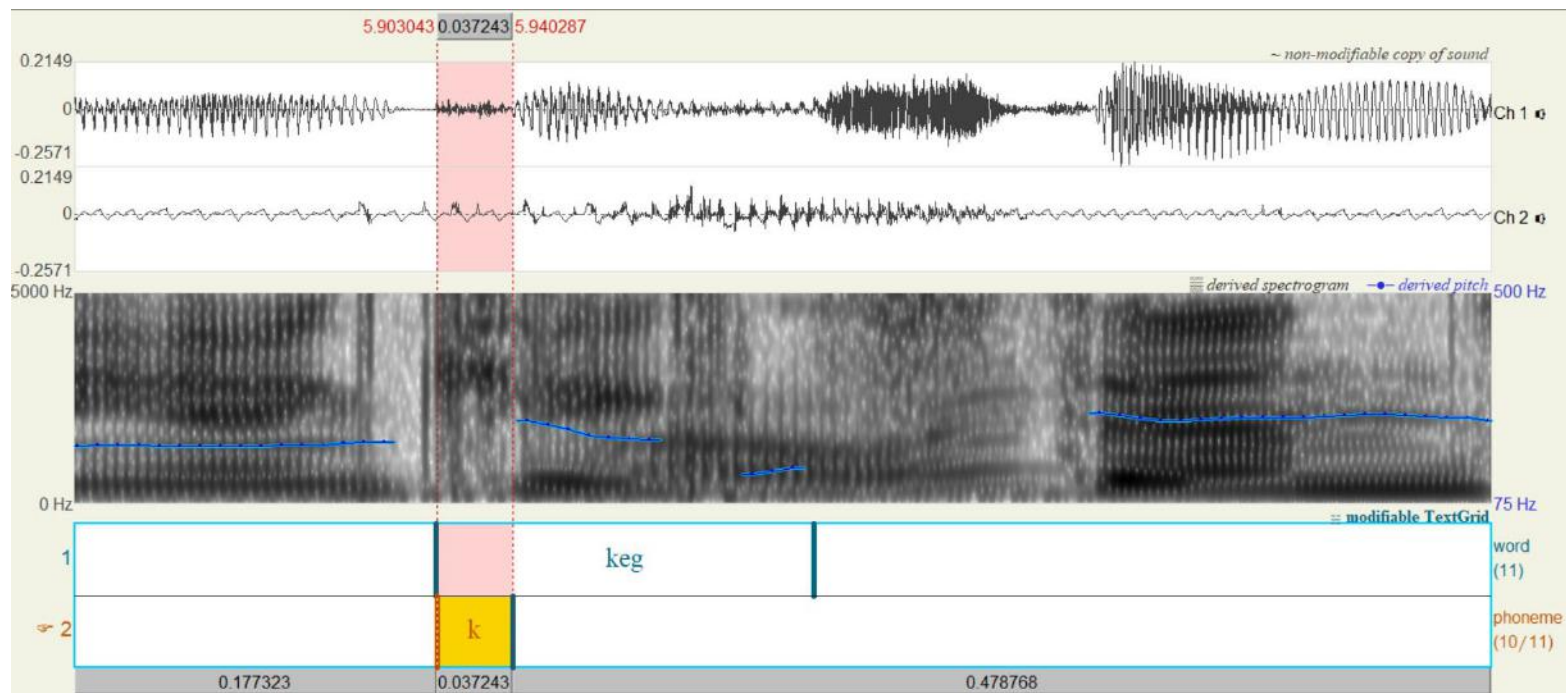
Data for Further Study : Code-Switching

- English-French Primed Code-Switching (EFPCS) Corpus :
dataset of English and French code-switching (CS) tokens from bilingual conversations
- Spécificités des Interactions verbales dans le cadre de Tandems linguistiques Anglais-Français (SITAF) Corpus :
CS tokens from English and French L2 learner conversations

Code-switching

- Concurrent use of multiple languages in a single speech act,
 - Common for multilingual speakers, distinct from loanwords
 - More recently studied from a phonetics / phonology perspective
-
- EFCS Corpus : Represents a communication decision by bilinguals
 - SITAF Corpus : Represents a necessity by L2 learners

EFPCS Corpus



“keg” in English, with Manual VOT on Tier 1
*Part of the utterance : “Un tonneau de bière, tu sais ? Et un **keg stand** c’est quand...”*

SITAF Corpus

File (Original Name)	Time Stamp	Speaker	Transcription / Description
13_mix_A09_F09_jeu_2_a	00:18	F09	"Ca veut dire quoi, ça ?"
	01:19	F09	"I think to pay more, it will res- restreindre re- restrict"
	02:40	F09	"But here, also, we have teachers who have doctorats and aggrégations "
	06:18	F09 & A09	<p>F09 - ...sometimes it's ern an arnaque How would you say? Une arnaque A09 - It's a scam F09 - A scame? A09 - You mean like errh it's It's like [c'est] like they're tr- like it's kind like fraud almost? "Arnaque" [oui] is not like "fraude"? F09 - Yes A09 - Ok F09 - Errj It's like err...err...French we say familiarly err "une boite à fric" je sais pas si vous avez une exp-, [ok] have you got any expression for I mean something you just pay and you have the same err the same quality of teaching that A09 - A scam F09 - Or sometimes [yeah] it's worst A09 - Ok F09 - Yeah. And I, my BTS it was free but i know in other schools it was err it was err expensive err about err two c'est two hundred c'est ça, deux milles ? A09 - Two hundred dollars wait err no two thousand two thousand, sorry dollars. Euh j'ai du mal</p>

Taken from a conversation task, in which participants had to determine to which degree they agreed on a subject.

References

- Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75–86
- Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). EFFECTS OF AGE AND SPEAKING RATE ON VOICE ONSET TIME: The Production of Voiceless Stops by Near-Native L2 Speakers. *Studies in Second Language Acquisition*, 37(1), 71-100. doi:10.1017/S0272263114000151
- Y. Shrem, M. Goldrick, and J. Keshet, “Dr. vot: Measuring posi-tive and negative voice onset time in the wild,” in *Proceedings of Interspeech*, 2019.
- Jacqueline Vaissière, Weilin Shen, Sachie Shioya. © 2015-2019 Cleanaccent (<https://cleanaccent.com/>)
- Horgues, C. & Scheuer, S. (2018). L’exploitation d’un corpus d’interactions en tandem anglais/français pour mieux comprendre les enjeux de la rétroaction corrective entrepairs. *Alterstice*, 8, (1), 63–81. <https://doi.org/10.7202/1052609ar>
- Cameron, M. (2019) *Voice Onset Time of French-English Code-Switching* (Master’s Thesis). Université de Paris 7 Diderot, Paris



Questions & Feedback