

Outils de description et d'analyse prosodiques du français parlé

Jean-Philippe Goldman
Anne Catherine Simon

Vendredi 23 mai 2008
Séminaire ProGram



UNIVERSITÉ
DE GENÈVE

UCL
Université
catholique
de Louvain



Ce qu'on attend de nous...

- Caractériser prosodiquement (ou identifier) des zones fonctionnelles, dans des unités de discours
- Principalement
 - le focus
 - la zone post-focale

Questions que nous nous posons...

- segmentation préalable en unités syntaxiques ou informationnelles?
- hypothèses sur l'interface prosodie – syntaxe – discours ?
- définition de "focus", "post-focus" ?

=> outils automatiques d'(aide à l')annotation prosodique

Outils d'aide à l'annotation prosodique (1/2)

outils qui **remplacent** à 100% le traitement manuel

→ tout automatique

outils qui **assistent**, guident le traitement

→ assisté par ordinateur

(Goldman, Lacheret, Simon; PFC 2006)

Outils d'aide à l'annotation prosodique (2/2)

- Outillage
 - gain de temps et fiabilité, reproductibilité
 - ergonomie et mutualisation des données et outils
- Approche inductive, *essentiellement* acoustique
- Cadre théorique
 - modélisation validée par le manuel (auditif)
 - aller-retour avec les modèles sous-jacents
 - modularité des outils (en fonction des cadres théoriques)

Outils (quasi-)automatiques

1. **EasyAlign (Aligner)**
segmentation phonétique, syllabique, lexicale
2. **ProsoGram (Mertens 2004)**
stylisation de l'intonation, visualisation durée et intensité
3. **ProsoProm**
détection de syllabes proéminentes
(voir PFC06, Interspeech07, CMLF2008)
4. **ProsoReport (Phonostylographe)**
description phonostylistique globale
étude « radio vs. lecture » (IDP - sept07, SP - mai08)
5. **ProsoReportDynamique**
description de l'évolution des paramètres prosodiques
application aux corpus ProGram

1. EasyAlign

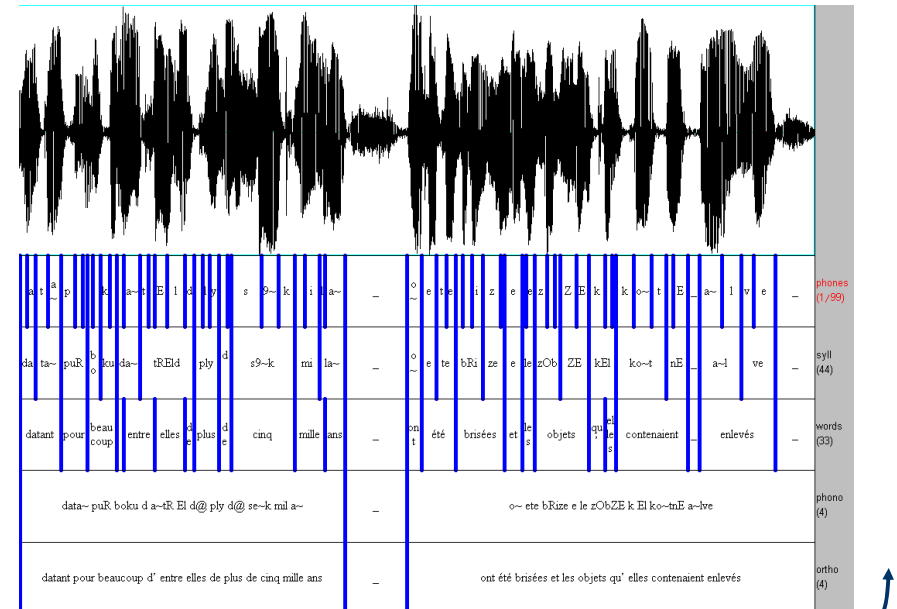
Outil automatique de segmentation phonétique, syllabique et lexicale, sous **Praat**, à partir d'une transcription orthographique (ou phonétique)



Série de scripts organisés dans un sous-menu «EasyAlign»

1. Segmentation en « pseudo-phrases »
Texte + Son → tire *ortho*
2. Phonétisation des « pseudo-phrases »
tire *ortho* → tire *phono*
3. Segmentation en phonèmes (alignement)
tire *phono* → tire *phones* + *syll* + *words*

DEMO



tires : ortho, phono, words, syll, phones

Utilisation, détails internes

1. Segmentation en pseudo-phrases
 - Pose des frontières entre chaque pseudo-phrase
 - Formatage par l'expert du texte en une pseudo-phrase par ligne => durée de chaque pseudo-phrase estimée par sa longueur orthographique et la durée des pauses
 - Vérification manuelle des frontières de phrases
2. Phonétisation
 - Transcrit phonétiquement la tire orthographique (dictionnaire phonétique et règles)
 - Vérification de la transcription en phonèmes
3. Segmentation (i.e. alignement) en phonèmes
 - modèles acoustiques et système HTK (Hidden Markov Model ToolKit) en alignement forcé

Performances et développements

- **Evaluation** : différence moyenne des positions de frontières segmentales (5mn)
 - 2 aligneurs humains: 23 ms
 - EasyAlign vs. humain: 28 ms
- **Efficacité**
 - de 2x à 5x le temps réel (vs. 80x)
- **Développement en cours**
 - Autres langues prises en compte (nl,hb,en,pt-br)
 - Automatisation de la détection des variantes phonologiques (élision, liaison)
 - Ajout d'informations morphosyntaxiques

2. ProsoGram (Mertens 2004)

fonction

segmentation et stylisation de la courbe de F0
représentation de la durée et de l'intensité

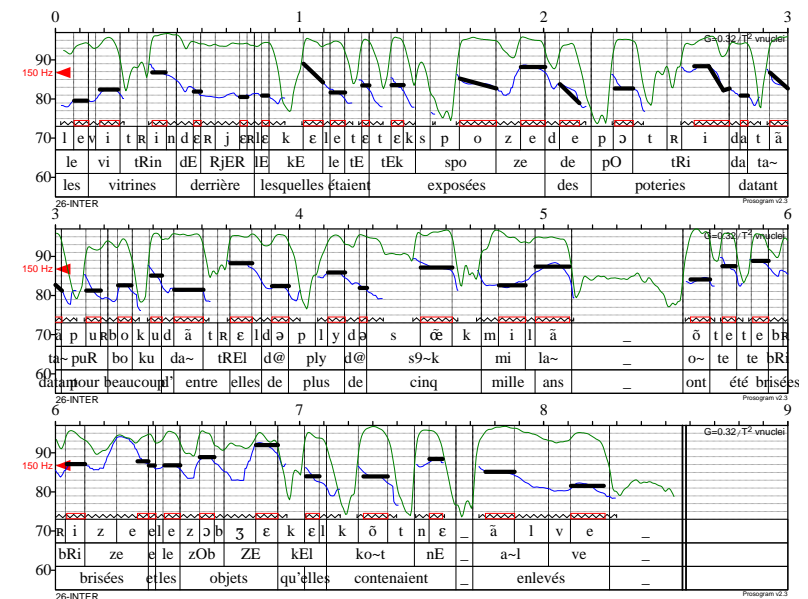
objectifs

aider à la transcription et à l'analyse prosodiques

⇒ à la base des outils développés et présentés ici

Mertens, Piet (2004). Le prosogramme : une transcription semi-automatique de la prosodie, *Cahiers de l'Institut de Linguistique de Louvain* 30/1-3, 7-25.

<http://bach.arts.kuleuven.be/pmertens/prosoqram/>



3. ProsoProm

fonction

détection de syllabes proéminentes dans un enregistrement segmenté en syllabes

objectifs

décrire un enregistrement,
aider l'annotateur,
confirmer/infirmier un modèle,
aider pour la segmentation en macro-unités intonatives

2. ProsoProm

Quels paramètres prosodiques et quels seuils ?

Cible mélodique après stylisation (Prosogram)	+2
Mouvement mélodique	+3.5
Durée syllabique	x2
Pause subséquente	300 ms

Proéminent, mais dans quel contexte ?

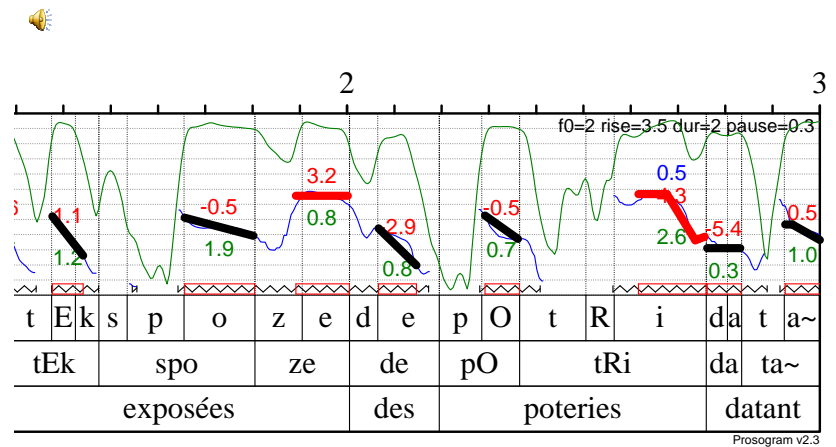
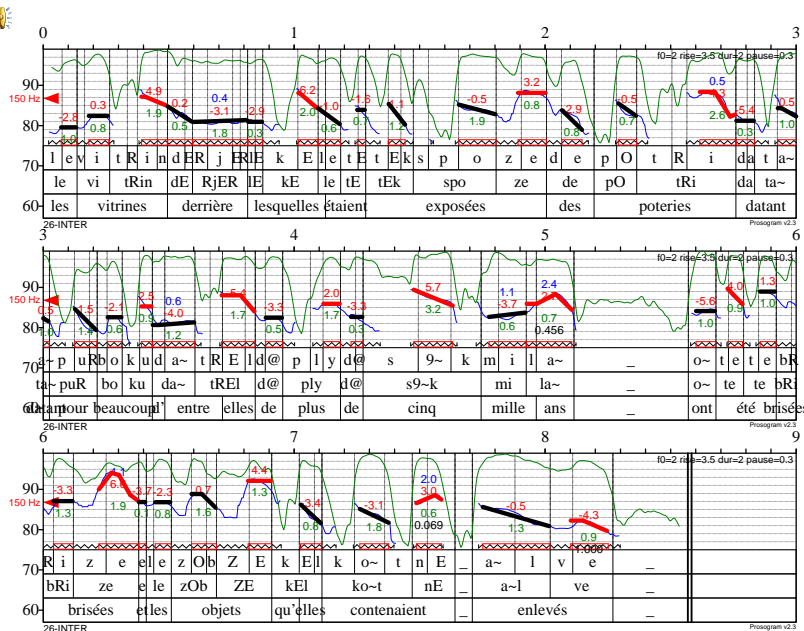
empan de contraste / relativisation des paramètres
indépendance du locuteur; du style; de la langue?

Quelle stratégie de décision ?

explicite : de 1 à n critères suffisants (i.e dépassant un seuil)
statistique : analyse discriminante

Quelle évaluation ?

comparaison syllabe-à-syllabe avec un corpus annoté



4. ProsoReport (Phonostylographe)

- outil de description phonostylistique globale ("portrait")
- **objectifs**
objectiver les propriétés prosodiques
rendre comparable la description d'extraits différents
- "photographie" globale **sans l'évolution ni la variation** des paramètres dans le temps ("détemporalisation" de la parole)
- **détails internes**
basé sur des outils associés
différents degrés de complexité dans la description

ProsoReport : outils associés (1/3)

EasyAlign => alignement syllabique
=> mesures de durée

syllabes	○ nombre de syllabes	
durée	○ durée de parole	s. (avec pauses)
	○ durée d'articulation	s. (sans pauses)
	○ taux d'articulation	% (articul./parole)
	○ débit de parole	syll./s. (pour parole)
	○ débit d'articulation	syll./s. (pour articul.)
	○ durée moyenne syllabes + déviation std	s.
	○ durée moyenne nuclei ⁴ syl. + dév. std	s.
	○ proportion des nuclei pdt l'articulation	%

ProsoReport : outils associés (2/3)

Prosogram => segmentation en "nuclei" et stylisation
=> éviter erreurs de détection

f0 globale	○ moyenne + déviation standard	ST (relatif à 1 Hz)
	○ étendue (inter quantile range)	ST
f0 tons	○ proportion des tons statiques,	%
	○ prop. des tons dynamiques montants,	%
	○ prop. des tons dynamiques descendants	%
	○ prop. des tons dynamiques complexes	%
f0 mvt	○ mouvement intra-nuclei des tons	
	○ dynamiques (excluant les statiques)	ST/syll. et ST/sec.
	○ mouvement absolu intra-nuclei (pour tous les tons)	ST/syll. et ST/sec.
	○ mouvement absolu inter-nuclei (i.e mvt moyen entre les tons)	ST/syll. et ST/sec.
	○ agitation mélodique	ST/syll. et ST/sec.
intensité	○ intensité moyenne dans les nuclei	dB
	○ intensité moyenne hors des nuclei	dB
	○ différence d'intensité entre nuclei et non-nuclei	dB

ProsoReport : outils associés (3/3)





Prosoprom => détection des syllabes proéminentes
=> "densité" des proéminences

Annotation LEX => mots pleins et clitiqes (...)
=> proéminences finales et proéminences initiales

Étude: le style "radio" vs. "lu"

matériau

- 3 chroniques radiophoniques de 2mn (2 femmes , 1 homme; France Inter, France Info)
- relues de manière neutre par une locutrice unique

	L-j	L-a	L-g	L-total	R-total	FR-j	FI-a	FI-g
Nombre de syllabes	772	644	525	1941	1901	755	638	508
Durée de parole	161,295	138,925	122,19	422,41	384,40	149,675	123,21	111,51
								

Hypothèses

par rapport à la "lecture neutre", les chroniques radiophoniques se caractérisent par (Goldman & Auchlin 2006)

- H1 plus grande agitation mélodique
- H2 surarticulation
- H3 quantité plus élevée d'accents initiaux

H1 f0 moyenne – étendue du registre

Tableau 4 : Étendue du registre tonal : étendue totale et étendue calculée en exceptant les 5% de syllabes les plus hautes et les 5% les plus basses (en demi-tons relatifs à 1 Hz)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Étendue f0 max-min	15,8	10,9	23	16,57	18,87	23	15,4	18,2
95%-5%	8,6	8,4	8,7	8,57	12,90	13,8	12,2	12,7

H2 taux d'articulation

Tableau 2 : Durées de parole, durées d'articulation (en sec.) et taux d'articulation (en %)

	L-j	L-a	L-g	Lmoy	R-moy	FR-j	FI-a	FI-g
Durée de parole	161,2	138,9	122,2	140,8	128,1	149,6	123,2	111,5
Durée d'art.	127,8	110,9	99,9	113	112,4	132,8	110,6	93,7
Taux d'art.	79,2	79,8	81,8	80,2	87,5	88,8	89,8	84

taux d'occupation du canal de parole (ratio articulation / silences)

H3 proportion de syllabes proéminentes en position finale vs non finale

Tableau 7 : Syllabes proéminentes : pourcentage de syllabes proéminentes par rapport aux syllabes non proéminentes ; pourcentage de syllabes proéminentes en position d'accent initial (syllabe initiale de mot lexical polysyllabique) ; pourcentage de syllabes proéminentes en position d'accent final (dernière syllabe de mot lexical)

	L-j	L-a	L-g	L-moy	R-moy	FR-j	FI-a	FI-g
Prom.	29,5	33,2	34,5	32,40	37,03	35,6	35,9	39,6
	(228/772)	(214/644)	(181/525)			(269/755)	(229/638)	(201/508)
Prom/i	19,7	18,1	26,3	21,37	31,43	30,1	30,3	33,9
Prom/f	58,6	64,5	58,8	60,63	59,6	59,3	59,4	60,1

=> davantage de proéminences en position initiale de mots lexicaux en parole radio

ProsoReport

Inconvénient

- "aplatit" en une dimension la variation ou l'évolution temporelle des phénomènes étudiés

Avantages

- comparabilité de corpus (tableaux)
- faire émerger des différences non prévues (bottom up)
- choix d'utiliser (ou non) l'information linguistique (LEX)

ProsoReportDynamic

Objectifs

- calculer les modifications prosodiques globales telles qu'elles évoluent dans le cours de la parole
- afficher ces variations

Paramètres analysés

- débit d'articulation
- valeur moyenne de fréquence fondamentale (f0 mean)
- étendue de la fréquence fondamentale (Δf_0)
- taux des syllabes proéminentes / non proéminentes

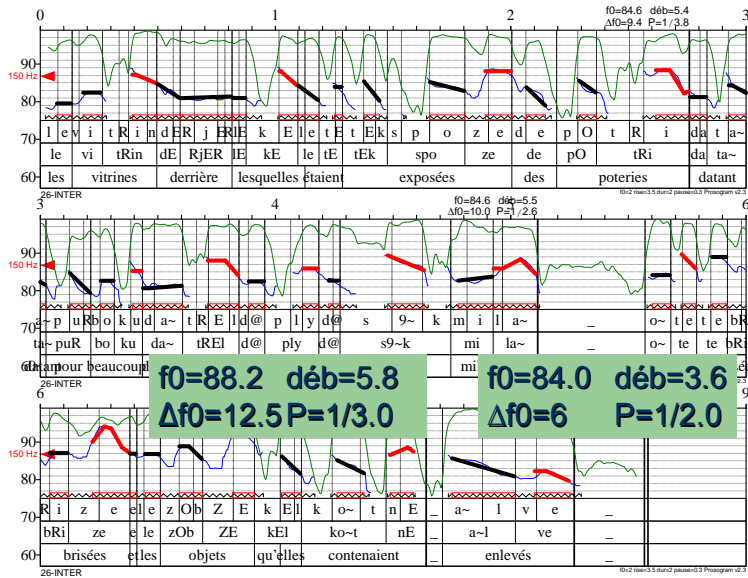
ProsoReportDynamic

Comment mesurer l'évolution des paramètres?

- segmenter l'enregistrement en "unités" (constituants syntaxiques, unités de discours...)
- calculer les paramètres prosodiques par unité
- comparer l'évolution d'une unité à la suivante

Résultats attendus – observations sur

- accélération ou ralentissement du débit
- compression de l'étendue de f0
- sur- ou sous-"accentuation"
- etc.

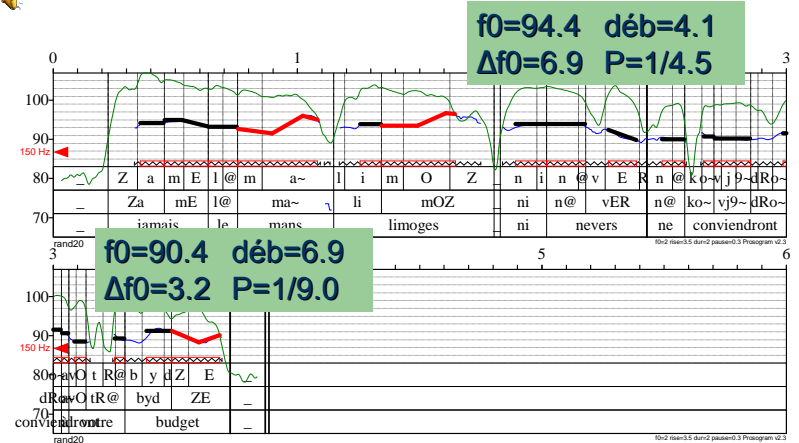


Un exemple analysé

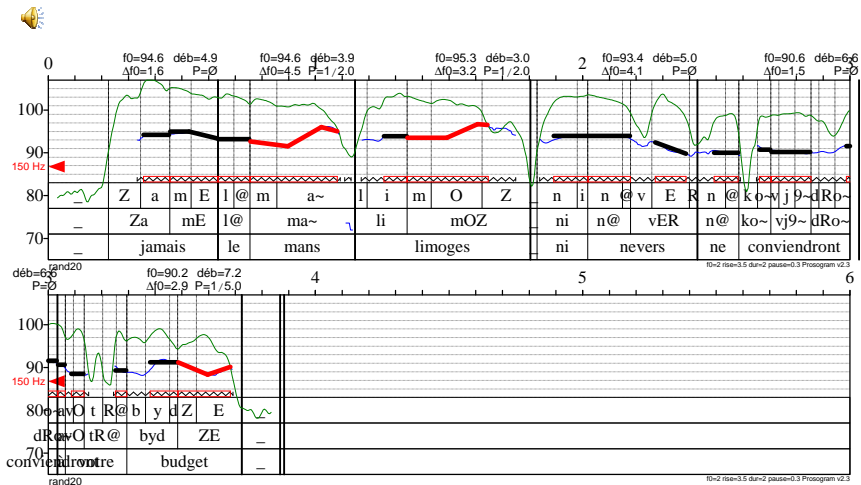
- Delta Lloyd (pub) – maximisation des contrastes

Quelques exemples du corpus Pro-Gram

1. exemple du corpus "coordination – registre"
2. exemple du corpus "ellipse – registre"
3. exemple du corpus "interrogatives (Ester)"



segmentation en 2 unités
compression de l'étendue
accélération du débit

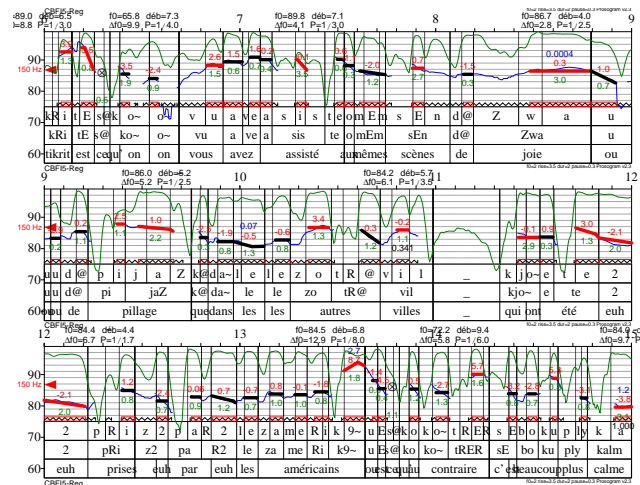


segmentation en constituants syntaxiques

mêmes mouvements globaux du débit, de l'étendue f0

validité? des mesures de f0 moyenne et de Prom

exemple CBF15_Reg

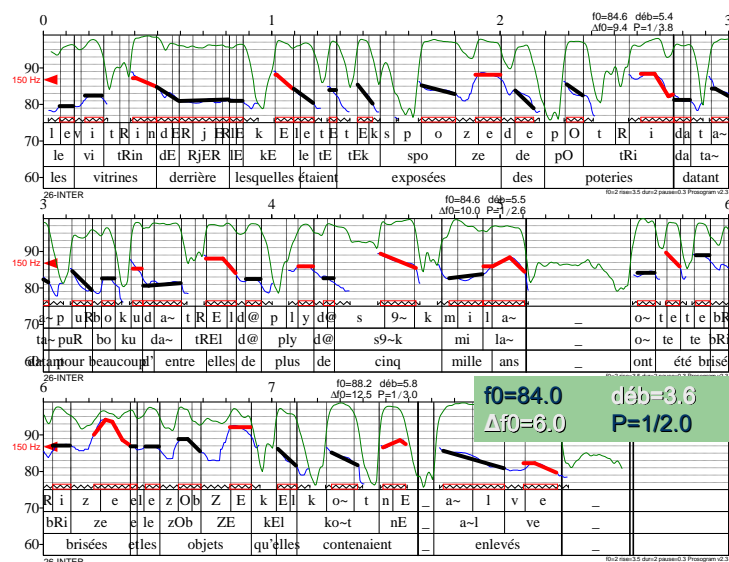


exemple CBF15_Reg

1. vous avez assisté
2. aux mêmes scènes de joie
3. ou de pillage
4. que dans les autres villes
5. qui ont été euh prises
6. par euh les Américains
7. ou est-ce qu'au contraire
8. c'est beaucoup plus calme

f0=89.8	déb=7.1
Δf0=4.1	P=1/3.0
f0=86.7	déb=4.0
Δf0=2.7	P=1/2.5
f0=86.0	déb=5.2
Δf0=5.2	P=1/2.5
f0=84.2	déb=5.7
Δf0=6.1	P=1/3.5
f0=84.4	déb=4.4
Δf0=6.7	P=1/1.7
f0=84.5	déb=6.8
Δf0=12.9	P=1/8.0
f0=72.2	déb=9.4
Δf0=5.8	P=1/6.0
f0=84.0	déb=5.4
Δf0=9.7	P=1/2.5

26-INTER (proéminences)



Conclusions

- outils (semi-)automatiques pour l'analyse prosodique
- choix de l'unité de calcul pour *ProsoReportDynamic*, fonction des hypothèses
- choix des paramètres prosodique à calculer
- ! difficulté de traiter par cet outil des enregistrements trop courts (phrases lues < 3-4 secondes)

merci de votre attention