

## Problématique

L'alcool a un effet au niveau acoustique de la parole. En intégrant les résultats de l'extraction de données linguistiques selon 13 variables phonétiques, nous tentons de détecter automatiquement, avec un algorithme de classification, si un segment est produit par une personne sobre ou intoxiquée.

### Question de recherche

Quelle segmentation, entre une découpe par phonèmes et par unités prosodiques, permet une meilleure performance en matière de détection automatique de l'intoxication à l'alcool dans la parole?

### Hypothèse

La segmentation par phonèmes permettra une meilleure détection automatique de l'intoxication à l'alcool dans la parole, puisque la plupart des variables phonétiques utilisées sont plus appropriées pour les phonèmes.

## Méthodes

Nous utilisons le *Alcohol Language corpus* (ALC), qui est composé d'une collection d'enregistrements de 162 locuteurs et locutrices de l'allemand à l'état sobre et à l'état alcoolisé (0,05% à 0,25%).

Le ALC est segmenté phonémiquement par le partiteur automatique du Bavarian Archive for Speech Signals (BAS). Nous avons réalisé automatiquement la segmentation par unités prosodiques en utilisant les silences ( $<p>$ ) comme bornes.

### Variable dépendante

Le taux d'alcool dans le sang rendu catégoriel: sobre pour  $x < 0,08\%$  et alcoolisé pour  $x \geq 0,08\%$

Nous avons extrait les données du signal sonore par un script Praat sur les deux segmentations avec les mêmes variables phonétiques. Après avoir produit un modèle linéaire général (GLM), présenté dans *Résultats*, avec l'entièreté des variables indépendantes sur les deux segmentations pour déterminer les variables significatives, nous ressortons un modèle pour chaque segmentation qui a toutes deux les mêmes 10 variables pour respecter l'objectif de comparaison de notre recherche.

### Variables indépendantes

Le F1 pris au troisième quart, le F2 pris aux premier et troisième quart, le F0, la durée, l'intensité, le centre de gravité, l'écart type, l'aplatissement (Kurtosis) et l'asymétrie (Skewness).

La classification est faite par régression logistique binomiale.

### Formule

$$Y = C_1X_1 + C_2X_2 + C_3X_3 + \dots$$

## Résultats: Segmentation par phonèmes

Premier Modèle pour la segmentation par phonèmes après normalisation des variables :

Variable	Estimate	Std.Error	z value	Pr(>  z )
(Intercept)	0.25979	0.01260	20.619	$< 2e - 16$ ***
CF1_c	-0.01290	0.01434	-0.900	0.3680
CF1_1_4	-0.00490	0.01199	-0.409	0.6828
CF1_3_4	-0.02622	0.01217	-2.155	0.0312 *
CF2_c	0.00903	0.01524	0.592	0.5535
CF2_1_4	-0.06712	0.01305	-5.145	2.67e-07 ***
CF2_3_4	-0.06543	0.01251	-5.231	1.68e-07 ***
F0_U	0.75872	0.00950	79.861	$< 2e - 16$ ***
Duree	2.03187	0.19318	10.518	$< 2e - 16$ ***
Intensite	-0.27103	0.01261	-21.497	$< 2e - 16$ ***
Centre_de_gravite	0.16717	0.02289	7.304	2.78e-13 ***
Ecart_type	-0.47763	0.02102	-22.726	$< 2e - 16$ ***
Skewness	-2.63766	0.09766	-27.008	$< 2e - 16$ ***
Kurtosis	4.07230	0.37725	10.795	$< 2e - 16$ ***

AIC: 2 355 726

La représentation est faite en observant les *Odds Ratio* du modèle pour la segmentation phonémique. Le modèle est sélectionné selon les variables significatives qui sont représentées ci-dessus par les astérisques.

Variable	N	Odds ratio	p
CF1_3_4	1707510	0.98 (0.96, 0.99)	0.004
CF2_1_4	1707510	0.93 (0.91, 0.94)	<0.001
CF2_3_4	1707510	0.93 (0.91, 0.95)	<0.001
F0_U	1707510	2.14 (2.10, 2.17)	<0.001
Duree	1707510	1.40 (1.21, 1.63)	<0.001
Intensite	1707510	0.74 (0.73, 0.76)	<0.001
Centre_de_gravite	1707510	1.20 (1.15, 1.25)	<0.001
Ecart_type	1707510	0.60 (0.58, 0.62)	<0.001
Skewness	1707510	0.06 (0.05, 0.07)	<0.001
Kurtosis	1707510	94.39 (47.41, 188.44)	<0.001
(Intercept)		1.34 (1.31, 1.37)	<0.001

Fig. 1: Segmentation phonémique

### Matrice de confusion

État de l'individu	Prédictions	
	Sobre	Intoxiqué
Sobre	410296	373777
Intoxiqué	428776	494661

Le tableau ci-dessus indique les nombres de prédictions du modèle utilisant une segmentation du signal par phonèmes. Lorsque le modèle prédit que la personne est sobre, mais qu'elle est en fait intoxiquée, ou vice versa, le résultat est erroné et affiché en rouge. Cela nous montre que le modèle de régression logistique binomiale prédit que le sujet est sobre sur 49.14% des segments, ayant raison 48.89% du temps. Il prédit que le sujet est intoxiqué sur 50.85% de l'échantillon, ayant raison 56.95% du temps.

Cela résulte en un pourcentage de bonne réponse total de 53.00%.

## Résultats: Segmentation par unités prosodiques (UP)

Premier Modèle pour la segmentation par UP après normalisation des variables :

Variable	Estimate	Std.Error	z value	Pr(>  z )
(Intercept)	0.50036	0.04479	11.172	$< 2e - 16$ ***
CF1_c	-0.06026	0.03058	-1.971	0.048765 *
CF1_1_4	0.08990	0.03007	2.989	0.002798 **
CF1_3_4	0.01315	0.03024	0.435	0.663692
CF2_c	-0.10071	0.03056	-3.295	0.000983 ***
CF2_1_4	-0.11389	0.03017	-3.775	0.000160 ***
CF2_3_4	-0.11305	0.03070	-3.683	0.000231 ***
F0_U	0.47860	0.03356	14.260	$< 2e - 16$ ***
Duree	-0.47688	0.10116	-4.714	2.43e-06 ***
Intensite	-0.31607	0.03680	-8.590	$< 2e - 16$ ***
Centre_de_gravite	1.06826	0.17625	6.061	1.35e-09 ***
Ecart_type	-0.95953	0.09945	-9.648	$< 2e - 16$ ***
Skewness	-3.01602	0.34101	-8.844	$< 2e - 16$ ***
Kurtosis	1.71469	0.84593	2.027	$< 2e - 16$ *

AIC: 152 536

La représentation est faite en observant le *Odds Ratio* du modèle pour la segmentation par UP. Pour respecter une comparaison juste, les mêmes variables phonétiques sont utilisées.

Variable	N	Odds ratio	p
CF1_3_4	110716	1.02 (0.97, 1.09)	0.434
CF2_1_4	110716	0.93 (0.88, 0.97)	0.003
CF2_3_4	110716	0.88 (0.83, 0.93)	<0.001
F0_U	110716	1.60 (1.50, 1.71)	<0.001
Duree	110716	0.61 (0.50, 0.75)	<0.001
Intensite	110716	0.77 (0.72, 0.82)	<0.001
Centre_de_gravite	110716	3.09 (2.20, 4.36)	<0.001
Ecart_type	110716	0.36 (0.30, 0.44)	<0.001
Skewness	110716	0.05 (0.02, 0.09)	<0.001
Kurtosis	110716	6.13 (1.13, 31.27)	0.031
(Intercept)		1.53 (1.42, 1.65)	<0.001

Fig. 2: Segmentation par unités prosodiques

### Matrice de confusion

État de l'individu	Prédictions	
	Sobre	Intoxiqué
Sobre	16905	15555
Intoxiqué	35999	42257

Le tableau ci-dessus indique les nombres de prédictions du modèle utilisant une segmentation du signal par unités prosodiques. Lorsque le modèle prédit que la personne est sobre, mais qu'elle est en fait intoxiquée, ou vice versa, le résultat est erroné et affiché en rouge. Cela nous montre que le modèle de régression logistique binomiale prédit que le sujet est sobre sur 47.78% des segments, ayant raison 31.95% du temps. Il prédit que le sujet est intoxiqué sur 52.22% de l'ensemble, ayant raison 73.09% du temps.

Cela résulte en un pourcentage de bonne réponse total de 53.44%.

## Discussion

En utilisant le *Alcohol Language Corpus*, nous avons présenté deux exécutions d'un modèle de régression logistique binomiale dans le but de comparer deux segmentations linguistiques: une par unités phonémiques et une autre par unités prosodiques. Il ne semble pas y avoir de différence entre le taux de prédictions correctes d'un modèle de régression logistique binomiale selon la segmentation du signal sonore. Il nous est donc impossible de confirmer ni d'infirmer notre hypothèse de départ qui prévoyait de meilleurs résultats sur la segmentation par phonèmes.

### Futures améliorations

- Une future analyse par régression logistique multi-niveaux permettrait de prendre en considération les différences interindividuelles des réactions à l'intoxication, en considérant chaque individu comme facteurs de notre variable dépendante.
- Une comparaison plus systématique avec des segmentations non linguistiquement justifiées, comme un découpage par dix millisecondes, pourrait être accomplie.
- En plus d'ajouter d'autres variables phonétiques plus pertinentes, l'ajout de variable comme le nombre d'erreurs de prononciation, d'hésitations et d'omission aurait pu participer à l'amélioration de l'algorithme de classification.
- Un système d'apprentissage automatique par réseau de neurones pourrait, si ce n'est répondre à la question par celui-ci, permettre de faire une sélection plus performante des variables intégrées dans l'analyse.

## Conclusion

Nous avons tenté d'aborder la problématique de la détection automatique de l'intoxication à l'alcool par la parole en comparant deux segmentations. Nous utilisons le *Alcohol Language Corpus* pour l'analyse. La classification est faite par une régression logistique binomiale pour chaque segmentation. Des matrices de confusions nous indiquent ensuite que les pourcentages de bonnes prédictions de nos modèles, soit 53.00% pour les phonèmes et 53.44% pour les unités prosodiques, sont insuffisants. Notre hypothèse prévoyant qu'une segmentation du signal sonore par phonèmes permettrait une meilleure détection automatique n'a pu ni être confirmée ni infirmée. Une analyse multi-niveaux qui prend en compte les locuteurs, l'utilisation d'un autre système d'apprentissage automatique, une sélection plus attentive des variables et l'ajout d'un autre type de segmentation préalable au choix de la langue observée, comme les syllabes, pourraient représenter des pistes intéressantes d'analyses ultérieures et provoquer de meilleurs résultats.