

# Modélisation Phonétique

Nicolas Gutehrlé

2021

## Table des matières

<b>1</b>	<b>Théorie source-filtre</b>	<b>3</b>
<b>2</b>	<b>Formants</b>	<b>3</b>
<b>3</b>	<b>Spectre et spectrogramme</b>	<b>5</b>
3.1	Voyelles . . . . .	5
3.2	Consonnes . . . . .	11
3.2.1	Occlusives . . . . .	11
3.2.2	Constrictives . . . . .	12
3.2.3	Sonantes, nasales et semi-consonnes . . . . .	13
<b>4</b>	<b>Exercices</b>	<b>16</b>
4.1	Exercice 1 . . . . .	16
4.2	Exercice 2 . . . . .	16

# 1 Théorie source-filtre

Les sons produits lors de la parole sont des sons complexes : les voyelles et les consonnes sonores sont composés de sons périodiques, tandis que les consonnes sourdes sont des sons aperiodiques.

La périodicité des sons voisés provient de la vibration des cordes vocales, situées dans le larynx. Cette vibration est appelée **fréquence fondamentale, F0**. La fréquence fondamentale F0 moyenne d'un homme est de 120 Hz (120 battements par seconde), tandis qu'il est d'environ 240Hz pour une femme. En vibrant, les cordes vocales produisent des harmoniques, qui sont des multiples de la fréquence fondamentale (Pour F0 égal à 120Hz, les premières harmoniques sont 240Hz, 360Hz, 480Hz, 600Hz, ...). Dans la production de la parole, on appelle **bourdonnement fondamental** l'ensemble constitué par la fréquence fondamentale et ses harmoniques.

Les ondes sont produites dans le conduit vocal, qui va du larynx jusqu'à la cavité buccale. Le conduit vocal est clos d'un côté (par les cordes vocales), et ouvert de l'autre (la bouche), comme une bouteille ou un verre. Il possède donc des résonances naturelles.

Plus un conduit clos est petit, et plus ses résonances naturelles seront élevées, donc aiguës (le remplissage d'une bouteille d'eau est un bon exemple). Une voix de femme ou d'enfant sera plus aiguë que celle d'un homme, en raison du larynx légèrement plus haut. Lors de la parole, la taille du conduit vocal change, notamment avec la forme des lèvres, la position de la langue et l'aperture.

Lors de la phonation, ces résonances naturelles vont filtrer le bourdonnement fondamental et renforcer ou diminuer certaines harmoniques du point de vue de l'amplitude. C'est ce qu'on appelle la **théorie source-filtre**, la source étant la vibration des cordes vocales, et les filtres étant les résonances naturelles du conduit vocal.

# 2 Formants

Les harmoniques renforcées par les résonances du conduit vocal sont appelées **formants**. Ces formants, nommés F1, F2, F3, F4, ... sont classés par ordre croissant de fréquences. Ainsi, la fréquence de F1 est inférieure à F2, qui est inférieure à F3, etc. Les formants apparaissent avec les voyelles ainsi qu'avec les consonnes vocaliques (nasales et sonantes).

La fréquence des formants dépend de la longueur de la cavité. Plus une cavité est longue et plus les fréquences seront basses (et inversement). Ainsi, tous les formants sont très sensibles aux modifications de l'appareil phonatoire. Cependant, certains formants sont plus sensibles à la modification d'un élément de la cavité phonatoire que d'autres. Indirectement, chaque formant nous renseigne sur la cavité qui les a renforcée. Ainsi, F1 est associé à l'ouverture de la bouche (trait d'aperture), F2 à la position de la langue (antérieure, postérieure) et F3 à la forme des lèvres (arrondies, étirées).

Le troisième formant correspond à la forme des lèvres (arrondies ou étirées). Il vient abaisser la valeur de F2. Enfin, le quatrième formant correspond aux voyelles nasales. Il est de 600Hz pour toutes les voyelles.

Les formants sont interdépendants. On ne peut donc pas changer la fréquence d'un formant sans changer celle des autres. Ainsi, pour diminuer F2 ou F3, F1 doit d'abord diminuer. De même, l'intensité des formants dépendra de l'intensité de F1.

Les voyelles nasales génèrent ce que l'on appelle des **antiformants**, qui vont réduire l'intensité et la fréquence des autres formants.

Pour visualiser l'impact des formants sur la production de la parole, on peut utiliser un synthétiseur de formants de Klatt, comme ici : <https://www.source-code.biz/klattSyn/>

Tous les paramètres de ce synthétiseur permettent de modifier un aspect du canal phonatoire (source, fréquence fondamentale, gain, ...). Les plus importants pour nous ici sont :

		F1	F2	F3
voy. fermées	i	308	2064	2976
	y	300	1750	2120
	u	315	764	2027
voy. mi- fermées	e	365	1961	2644
	ø	381	1417	2235
	o	383	793	2283
voy. mi- ouvertes	ɛ	530	1718	2558
	œ	517	1391	2379
	ɔ	531	998	2399
voy.ouv.	a	684	1256	2503

FIGURE 1 – Tableau des trois premiers formants des voyelles françaises

**F0** : indique la fréquence fondamentale des cordes vocales (en général, 120Hz pour un homme, 240Hz pour une femme)

**Duration** : durée de son qui sera généré

**Formant freq** : chaque case (jusqu'à F6 et Nasal) permettent d'indiquer la fréquence d'un formant. Les 3 premiers formants sont suffisants en général pour distinguer les voyelles et consonnes

**Synthesize, Play** : Ces boutons permettent respectivement de générer et de jouer le son produit par le synthétiseur avec les paramètres donnés



FIGURE 2 – Visuel d'un synthétiseur de Klatt

### 3 Spectre et spectrogramme

Le spectrogramme est une représentation visuelle à trois dimension du son. L'axe horizontal représente le temps (en ms) tandis que l'axe vertical représente les fréquences (généralement entre 0 et 8000 Hz). Enfin, l'intensité du son est représenté par le degré de couleur ou de noirceur du trait.

Toutes les harmoniques d'un son apparaissent sur le spectrogramme. Les harmoniques renforcées, donc les formants, apparaissent cependant bien plus nettement.

Les sons voisés se caractérisent par une barre de voisement. Celle-ci est visible par une barre légèrement grisée dans les basses fréquences.

Cette barre de voisement n'est pas distinguable avec les voyelles, celles-ci étant forcément sonores. Elle est cependant particulièrement visible avec les consonnes sonores telles que /b m d/. Sur le spectrogramme, cette barre de voisement se repère dans les fréquences les plus basses. Elle précède une zone très chargée en fréquences, qui correspond au relâchement de l'occlusion.

#### 3.1 Voyelles

Les voyelles, en raison de la vibration des cordes vocales, sont des sons périodiques. Pour le formant F1, on remarque plus les voyelles sont fermées (/i e y o/) et plus le formant est bas. A l'inverse, plus les voyelles sont ouvertes (/a ε oe ɔ/) et plus F1 est grand. De même pour F2, plus la langue est avancée et plus F2 est haut

Le triangle ou trapèze vocalique de l'alphabet IPA est obtenu en croisant les valeurs de F1 et F2 de chaque voyelle sur un graphique. Cependant, ce triangle ne prend pas en compte la forme des lèvres (F3) : celui-ci est donc ajouté entre parenthèses à côté de la voyelle.

L'arrondissement des lèvres vient abaisser la fréquence de F2 et F3 (comparer /i/ et /y/)

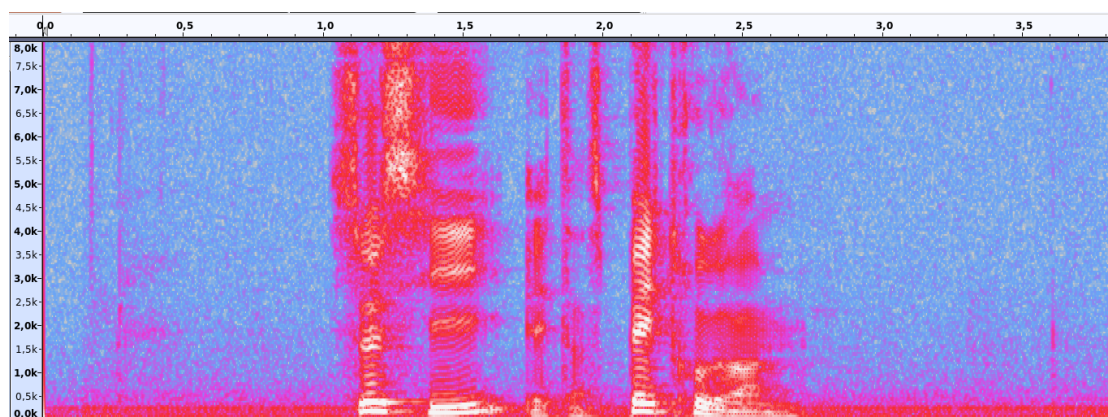


FIGURE 3 – Spectrogramme de la phrase "Ceci est un spectrogramme"

Le formant nasale, situé entre 500HZ et 700Hz en moyenne, vient abaisser les valeurs de tous les autres formants grâce à ce qu'on appelle des **antiformants**.

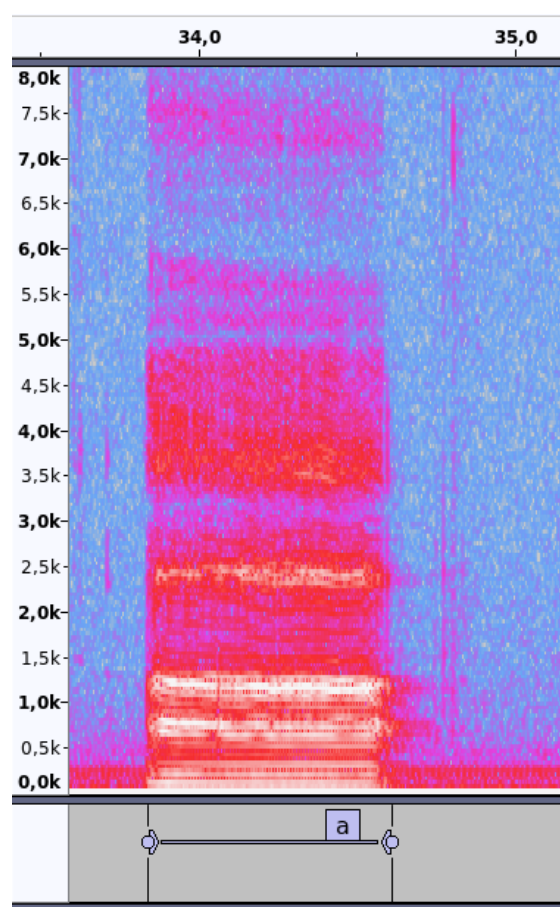


FIGURE 4 – Spectrogramme de voyelle ouverte

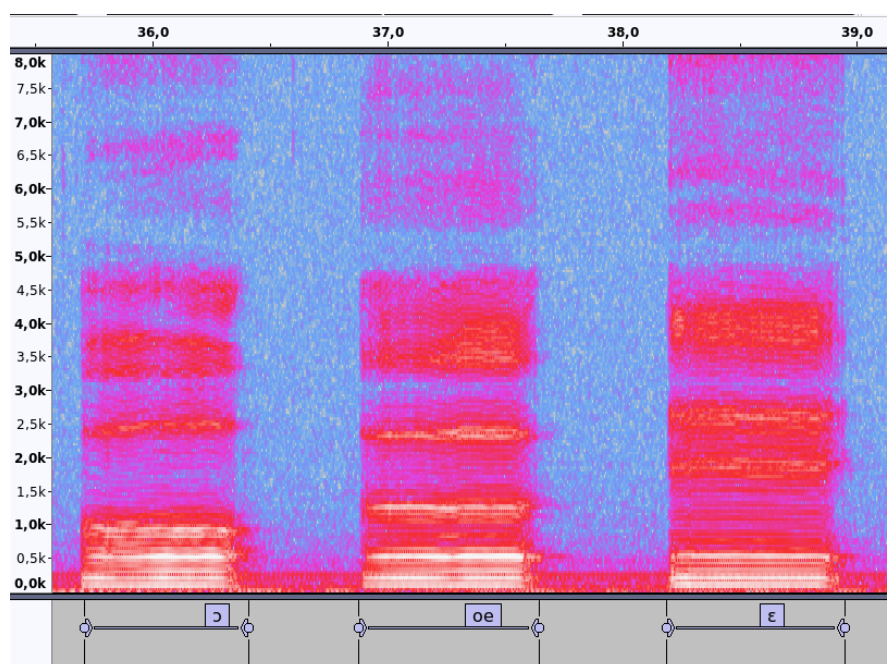


FIGURE 5 – Spectrogramme de voyelles semi-ouvertes



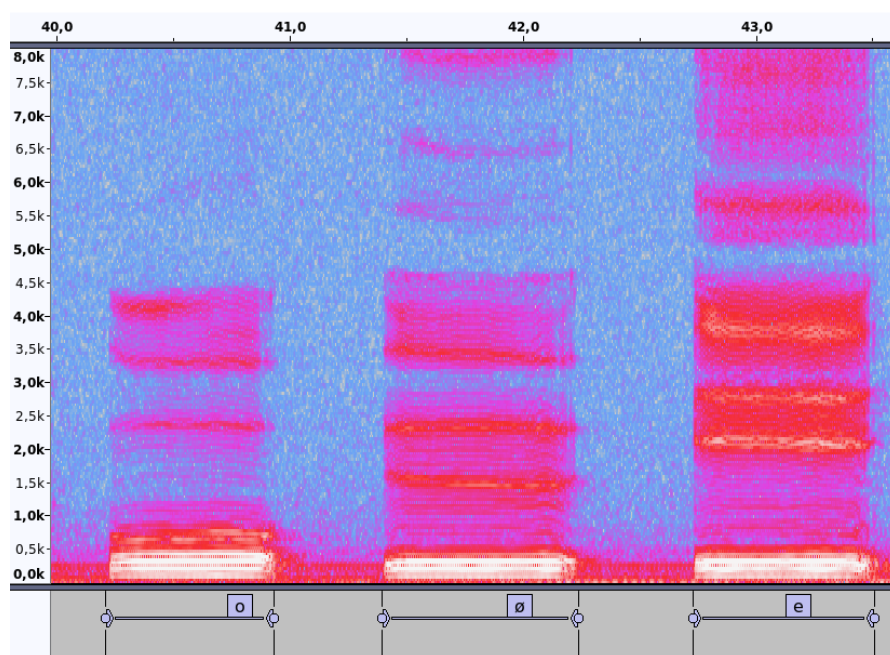


FIGURE 6 – Spectrogramme de voyelles semi-fermées

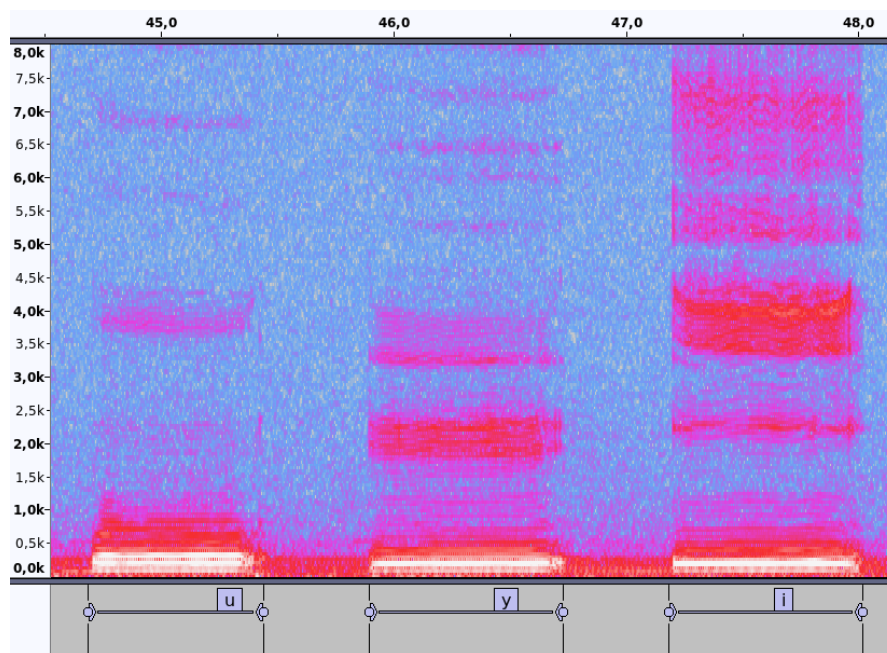


FIGURE 7 – Spectrogramme de voyelles fermées

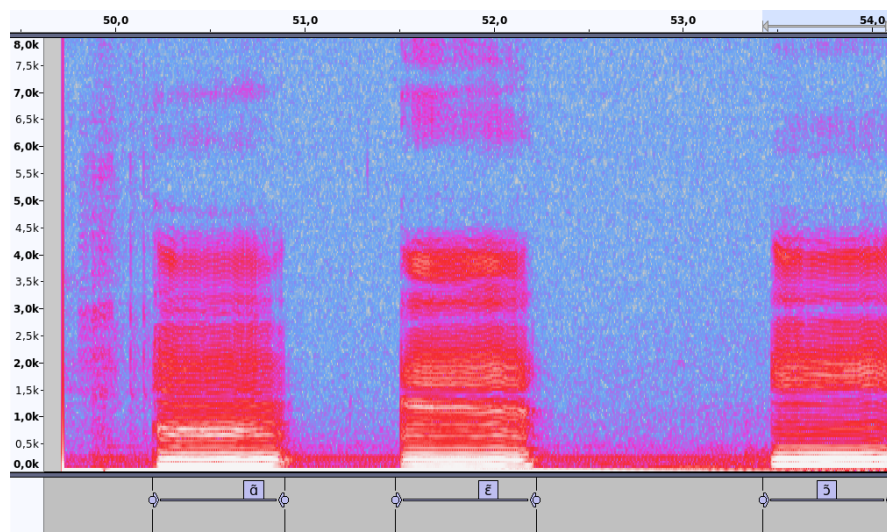


FIGURE 8 – Spectrogramme de voyelles nasales

## 3.2 Consonnes

Les consonnes sourdes sont apériodiques tandis que les consonnes sonores sont périodiques.

Les consonnes sonantes (/l/ et /R/) ainsi que les nasales, en raison de leur caractère vocalique, possèdent des formants, bien que moins intenses.

### 3.2.1 Occlusives

Du point de vue acoustique, les occlusives se constituent de deux étapes : une tenue suivie d'un relâchement. Pour les sourdes, cette tenue correspond à un silence, suivi par un bruit d'explosion lors du relâchement. Ce dernier est apériodique.

Les occlusives sonores sont quant à elles périodiques, en raison de la vibration des cordes vocales. Contrairement aux sourdes, les cordes vocales vibrent lors de la tenue. Cette vibration, qui correspond à la fréquence fondamentale, se voit sur un spectrogramme au travers d'une **barre de voisement**, qui précède l'explosion.

On appelle **Voice Onset Time (V.O.T.)** le temps de déclenchement du voisement par rapport au relâchement. Si les cordes vocales vibrent avant le bruit d'explosion, comme pour les sonores, on dit que le VOT est négatif. À l'inverse, on dit qu'il est positif si les cordes vocales vibrent après l'explosion, comme pour les sourdes.

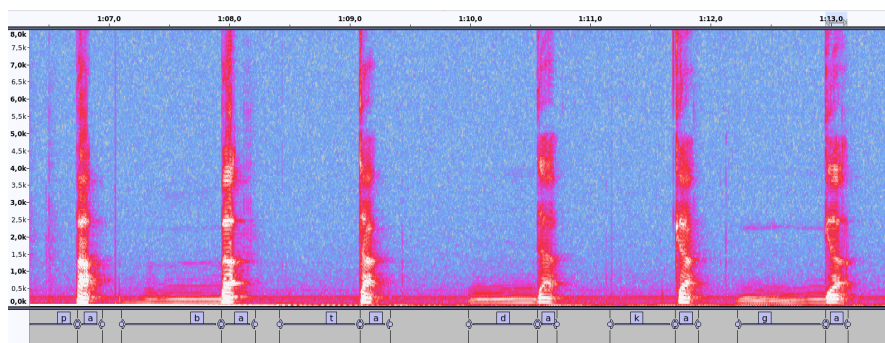


FIGURE 9 – Spectrogramme des occlusives dans le contexte de /a/

### 3.2.2 Constrictives

Contrairement aux occlusives qui se caractérisent par un silence (hors les cordes vocales), les constrictives se caractérisent par un bruit intense lors du blocage de l'air. Ce bruit sera plus ou moins intense suivant le type de constrictive. Ainsi en comparaison, les fricatives /f v/ sont moins intenses (3kHz à 8kHz) que les sifflantes /s z/ (4kHz et 8kHz), qui sont plus intenses que les chuintantes /ʃ ʒ/ (2kHz à 7kHz).

Comme pour les occlusives, les constrictives peuvent être apériodiques si elles sont sourdes et périodiques si elles sont sonores. De plus, l'intensité du bruit dépend également du voisement de la consonne : une consonne sourde sera plus bruyante qu'une consonne sonore (la vibration des cordes vocales venant réduire la pression supra-glottique).

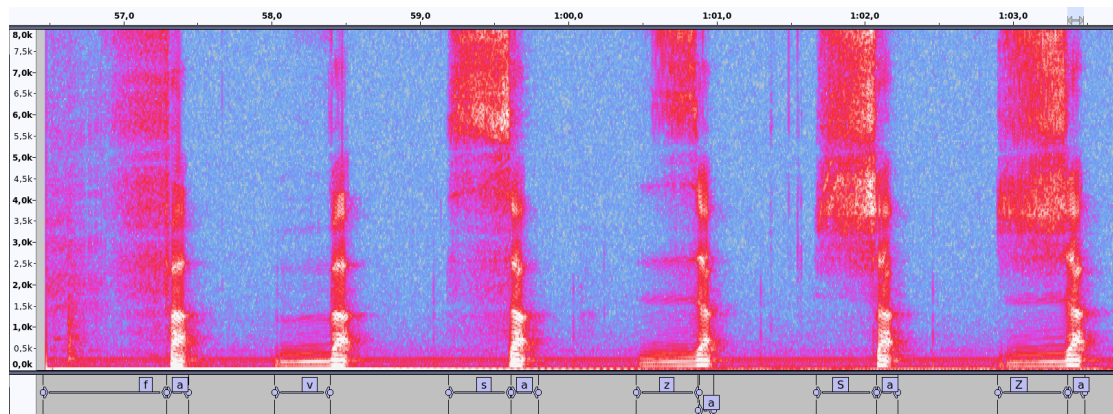


FIGURE 10 – Spectrogramme des constrictives dans le contexte de /a/

### 3.2.3 Sonantes, nasales et semi-consonnes

Les sonantes (latérales, vibrantes), les nasales et les semi-consonnes se différencient des autres consonnes du fait qu'elles soient très vocaliques. Celles-ci sont toutes sonores, et leur constriction est moindre que celle des autres consonnes. À part les semi-consonnes, elles peuvent même jouer un rôle vocalique au sein d'une syllabe.

La latérale /l/ possède des formants. Son format F1 se situe aux alentours des 300Hz, tandis que ces autres formants correspondent à la voyelle qui le suit.

Les nasales /m n/, bien que des occlusives, permettent le passage de l'air par la fosse nasale. Ils provoquent donc des anti-formants, qui vont réduire la fréquence des autres formants.



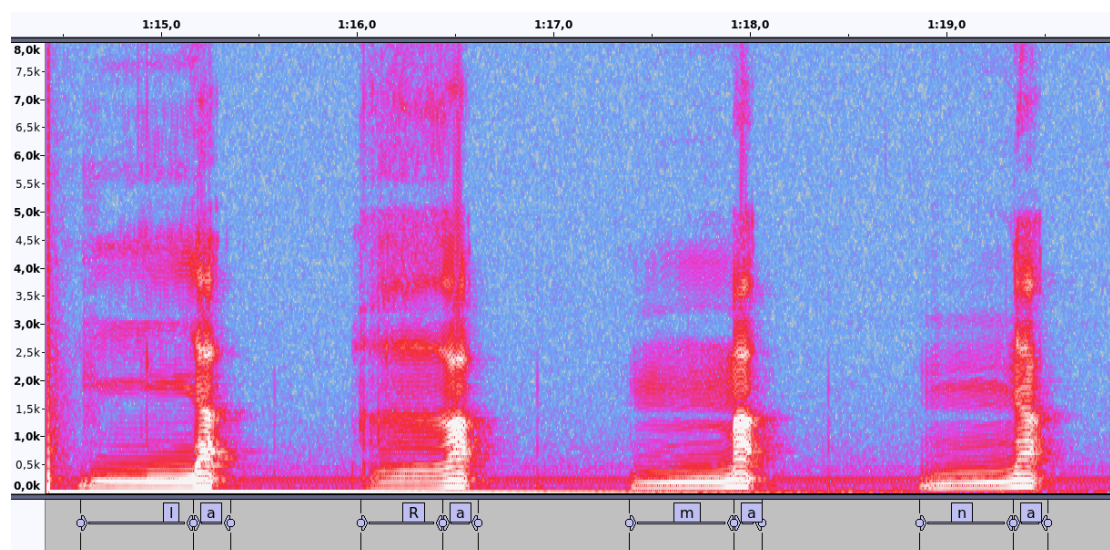


FIGURE 11 – Spectrogramme des sonantes dans le contexte de /a/

Enfin, les semi-consonnes possèdent également des formants, qui sont cependant instables, en raison de la nature même glissante de ces sons.

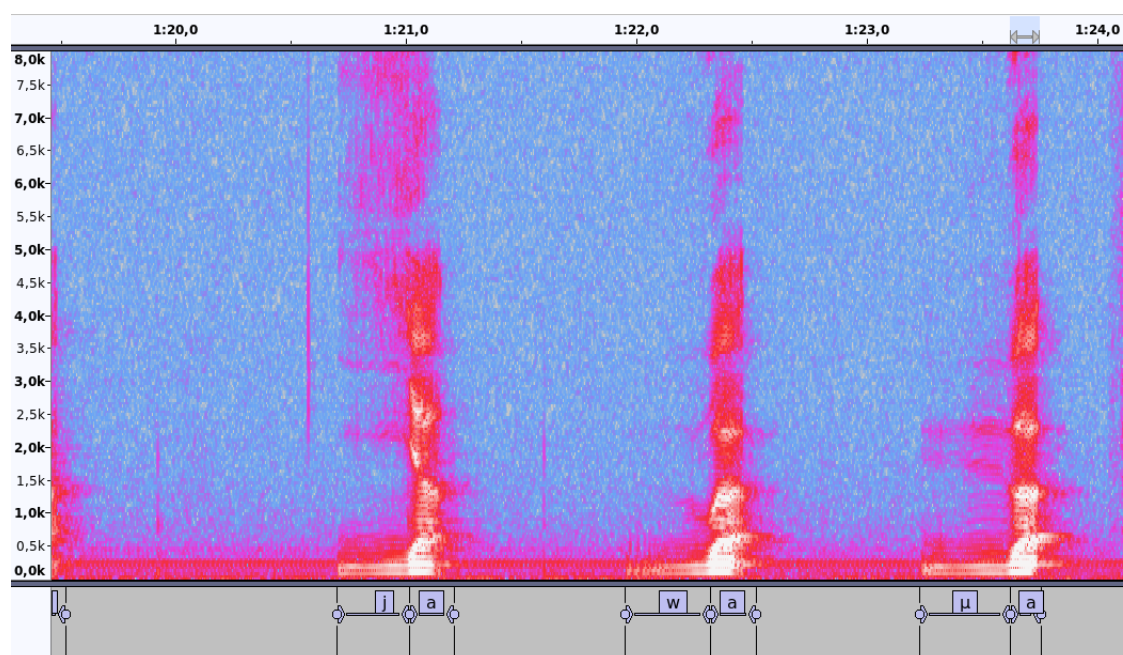


FIGURE 12 – Spectrogramme des semi-consonnes dans le contexte de /a/

## 4 Exercices

### 4.1 Exercice 1

Avec le tableau des formants dans le cours ci-dessus, modifiez les valeurs F0, F1, F2 et F3 dans le synthétiseur de Klatt pour entendre les différences entre les voyelles. N'hésitez pas à seulement changer la valeur de F1 par exemple pour entre la différence par rapport au trait d'aperture. N'hésitez pas non plus à regarder l'oscillogramme et le spectre affichés en dessous.

Le synthétiseur est disponible ici : <https://www.source-code.biz/klattSyn/>

### 4.2 Exercice 2

Dans Audacity, créez un nouveau projet dans lequel vous vous enregistrez en prononçant les mots suivants :

- papa
- maman
- roi
- reine

Une fois ces mots enregistrés, vous annoterez le spectrogramme, en indiquant à quel phonème chaque portion appartient (comme dans les images de ce cours).

- Pour créer un nouveau projet dans Audacity : Fichier Nouveau
- Pour vous enregistrer : cliquez sur le bouton rouge. Assurez vous que votre micro est bien détecté
- Pour passer en vue spectrogramme : à gauche de la piste, cliquez sur la flèche de "Piste audio", puis sur "Spectrogramme"
- Pour annoter : sélectionner la portion du spectrogramme qui vous intéresse, puis faites CTRL-B (Windows et Linux) ou CMD-B (Mac). Entrez ensuite le texte dans l'encart qui s'est ouvert
- Pour écrire en phonétique sur votre ordinateur : Utilisez le site TypeIt (qui propose plusieurs types de claviers, dont l'IPA) : <https://ipa.typeit.org/full/>