

Analyse et représentation de signaux acoustiques modulés

Application à la mesure du vibrato musical

Bertrand DAVID

Département du Traitement de Signal et des Images
Télécom Paris (ENST)

La présentation commence par l'écoute de l'ambiance sonore d'une forêt du Costa Rica, où les sons d'oiseaux et l'écoulement d'un ruisseau ressortent plus particulièrement. L'écoute est suivie simultanément et illustrée par une représentation du son dans un plan temps fréquence. Cette représentation permet de situer aussi bien dans le temps et la fréquence les sons entendus, qui sont autant d'événements dans ce plan. Certains sont bien localisés d'un point de vue temporel et d'autres présentent une évolution notable de leur contenu fréquentiel, que notre oreille perçoit.

Au delà du vibrato, cette présentation est l'occasion d'une discussion plus générale sur les signaux modulés, c'est à dire des signaux dont une ou plusieurs caractéristiques évoluent au cours du temps.

L'analyse et la représentation de ce type de signaux seront exposées de façon assez simple, afin de mettre l'accent en particulier sur l'utilisation de cet outil de représentation du son dans le plan temps/fréquence. Le second but est de décrire le vibrato musical, et les grandeurs qui y sont associées.

Le plan de la présentation consiste en l'écoute de quelques exemples de signaux modulés dont des vibratos musicaux, puis en l'exposé du principe et du fonctionnement des différents outils de représentation temps-fréquence comme le spectre et le spectrogramme, enfin la présentation des principaux modèles de modulation et les grandeurs descriptives du vibrato. Pour finir, un panorama des outils de représentation sera exposé, aboutissant à une discussion sur les précautions à prendre avec ces outils.

Introduction, exemples

On peut distinguer plusieurs types de sons dans lesquels nous percevons des modulations :

- Les sons modulés « écologiques » (animaux ou objets).

On entend avant tout des variations en fréquence (chants d'oiseaux, bruit d'une porte qui grince), mais aussi des variation d'intensité comme c'est le cas pour le bruit d'une grenouille.

- Les effets synthétiques, produits par modification de sons enregistrés ou par synthèse *ex nihilo*. Une première catégorie concerne les sons sans enveloppe ou avec enveloppe, entre lesquels on perçoit la modification d'intensité. Un autre type de modulation, non développée mais néanmoins intéressante, est la modulation en anneau. Elle consiste à faire passer l'enveloppe en dessous de la position 0, de façon à avoir des valeurs positives et négatives, ce qui occasionne des « changements de signe » du son, autrement dit des changements de phase.

Un autre exemple de synthèse couramment utilisée (cartes son du commerce) est celle obtenue par modulation de fréquence .

- Les sons musicaux vus en détail au cours de cette conférence (voix chantée instruments à vent, instruments à cordes) où le vibrato est produit par des instrumentistes qui agissent physiquement sur leur instrument.

Différentes phases d'un son musical

Cela nous amène à essayer de distinguer différentes phases d'un son musical. Dans le premier extrait sonore de l'introduction représenté dans le plan temps fréquence, certains événements sont nettement localisés, mais on peut également percevoir un fond sonore, qui évolue de façon assez continue (le bruit de la rivière). Deux types de sons sont donc à distinguer: ceux qui varient rapidement et ceux qui restent relativement constants. Pour un signal, on parlera plutôt de stationnarité ou d'instationnarité. Pour revenir à un modèle très simple, il est possible de distinguer différentes phases d'un son : l'attaque et la décroissance qui la suit, à variations rapides, puis une partie plus stable, et enfin une partie d'extinction, également rapide. Cette partie stable ne peut être qualifiée de complètement « stationnaire » car c'est dans cette partie que se développe souvent le vibrato. Or celui-ci dépend du temps, il représente une évolution des caractéristiques du signal, donc un écart par rapport à la stationnarité.

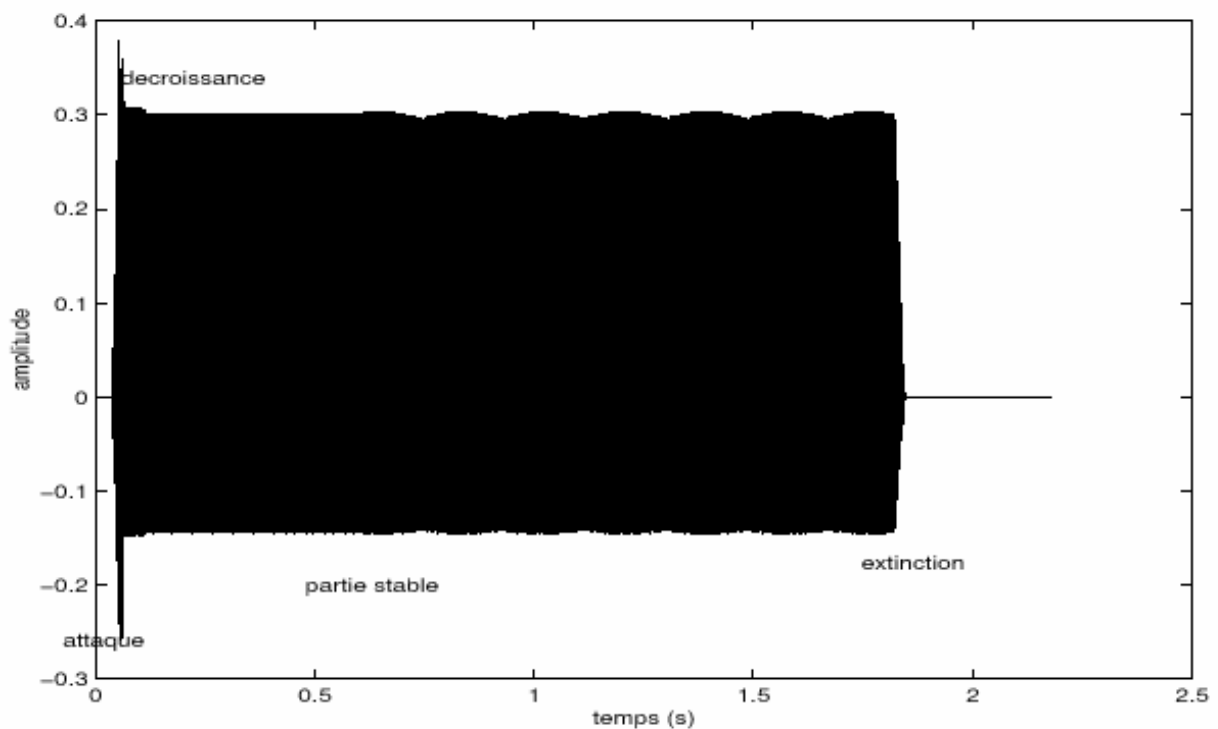


FIG 1 : Représentations des différentes parties temporelles d'un son

Il existe des outils d'analyse et de représentation du signal : le spectre et le spectrogramme.

Spectre

Le spectre est un outil bien connu et très répandu. Pour une sinusoïde infinie, toute l'énergie du spectre est concentrée à une fréquence donnée, c'est à dire la fréquence de la sinusoïde.

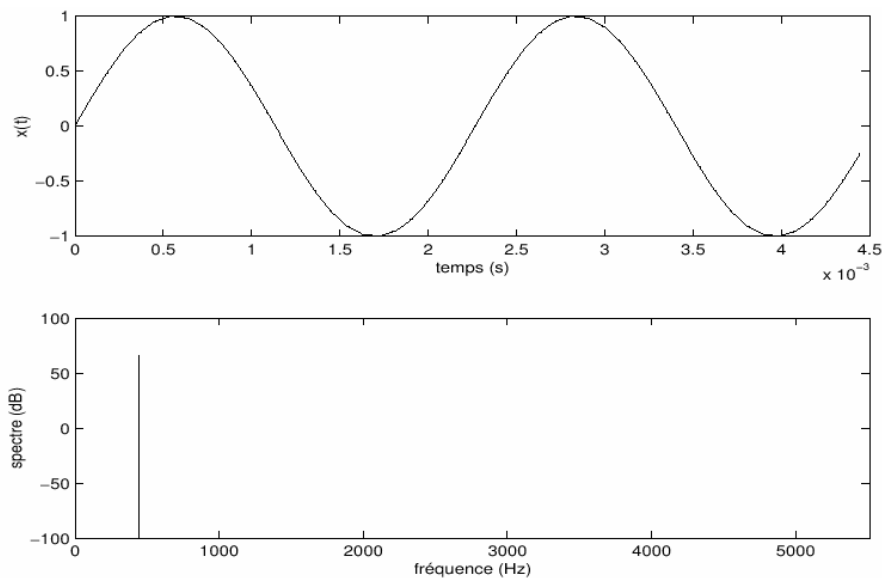


FIG 2 : Représentation temporelle et spectrale d'une sinusoïde pure

Si on tronque ce sinus, l'énergie va d'une part se répartir autour de la fréquence de la sinusoïde (c'est ce qu'on appelle l'étalement spectral), d'autre part on observe la présence d'énergie dans toutes les fréquences (Ce phénomène est appelé fuite spectrale). Etalement et fuite sont liés à la troncature.

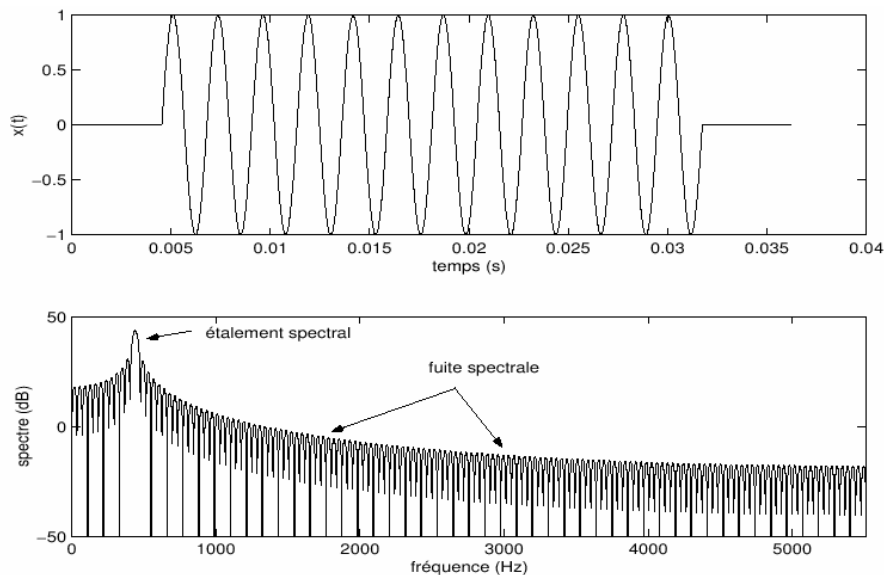


FIG 3 : Représentation temporelle et spectrale d'une sinusoïde par une fenêtre rectangulaire

Si on réalise la troncature de façon non rectangulaire, mais en « fenêtrant » le signal, les transitions dans le signal sont alors plus douces. Cela limite la fuite spectrale mais en

revanche augmente l'étalement en fréquence. Ce point est important pour comprendre le rôle des fenêtres d'analyse. Si la fenêtre a des discontinuités fortes, les fuites spectrales vont être importantes, mais l'étalement moindre. Si on prend une fenêtre de discontinuité plus douce, on va au contraire obtenir un étalement plus grand, mais moins de fuites spectrales.

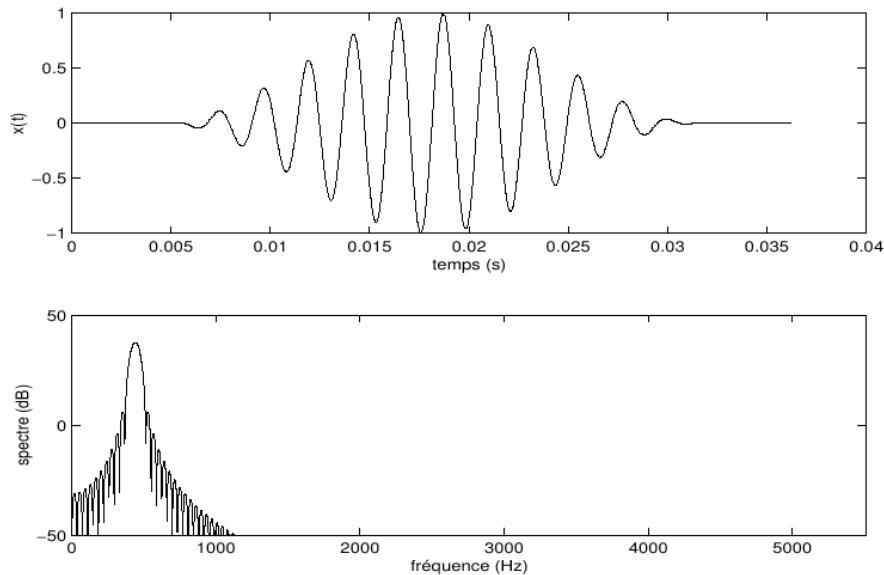


FIG 4 : Représentation temporelle et spectrale d'une sinusoïde fenêtrée par une fenêtre non rectangulaire

Spectrogramme

L'intérêt du spectrogramme est de pouvoir représenter le spectre en évoluant dans le temps. Le nom scientifique de la fonction mathématique associée à cet outil, plus communément appelé « spectrogramme », est la Transformée de Fourier à Court Terme (TFCT). Ce nom provient de l'analyse effectuée sur des fenêtres de support temporel fini. Une autre dénomination de cette représentation est « sonagramme ». Il s'agit d'une marque déposée Kay Electronics.

Le principe du spectrogramme est de « découper » le son en tranches ou trames qui se recouvrent. Pour chacune de ces tranches on calcule une transformée de Fourier. Ce spectre est alors représenté à un temps correspondant à celui du centre de la fenêtre, sous forme d'un code de couleur. Sur l'exemple suivant, le jaune correspond aux amplitudes les plus fortes, le bleu/violet aux amplitudes les plus faibles. On a ainsi une idée de l'aspect du spectre au temps t . A chaque calcul du spectre, le signal est fenêtré de façon à pouvoir régler à la fois la fuite et l'étalement spectral.

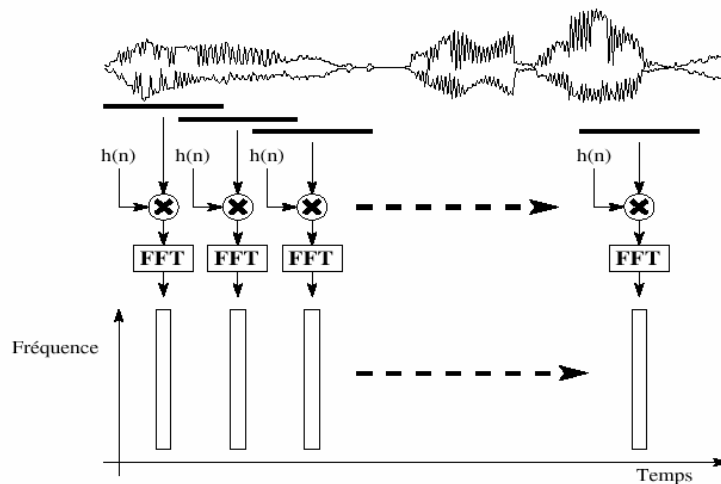


FIG 5 : Principe de calcul et de représentation d'un spectrogramme

Voici un exemple de voix chantée. Il s'agit d'un glissando C5-E5 réalisé par une soprano. L'analyse a été effectuée avec une fenêtre de Hanning de longueur 23 ms. Sur le spectrogramme, on observe bien le glissando et le vibrato de la chanteuse.

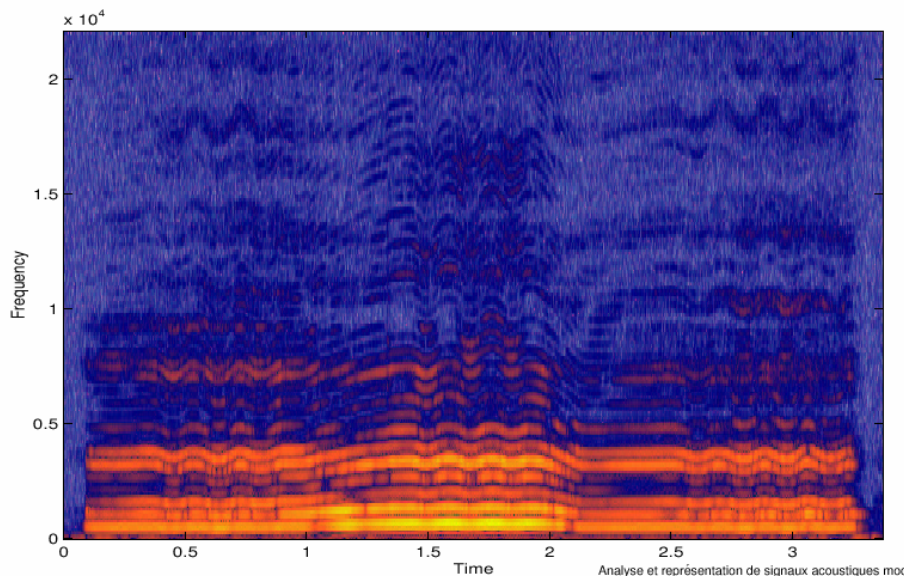


FIG 6 : Spectrogramme d'un extrait vocal (glissando C5-E5, soprano) avec une fenêtre rectangulaire de 23ms

Pour mesurer le vibrato, on serait tenté de réduire la longueur de la fenêtre dans le temps pour gagner en précision et suivre au mieux les variations du spectre. En réalité, si on réduit la longueur des fenêtres, l'étalement spectral augmente, par conséquent la largeur des raies sur le spectrogramme aussi, ce qui perturbe finalement la mesure, car on ne distingue plus distinctement les différentes trajectoires dans le spectrogramme.

Si on revient à la même longueur de fenêtre que dans le premier exemple, tout en utilisant une fenêtre rectangulaire au lieu de la fenêtre « douce » de Hanning, l'étalement spectral est plus faible et les lignes sur le spectrogramme plus fines, mais on a par contre des fuites spectrales beaucoup plus importantes, caractérisées par un manque de contraste dans la représentation du spectrogramme.

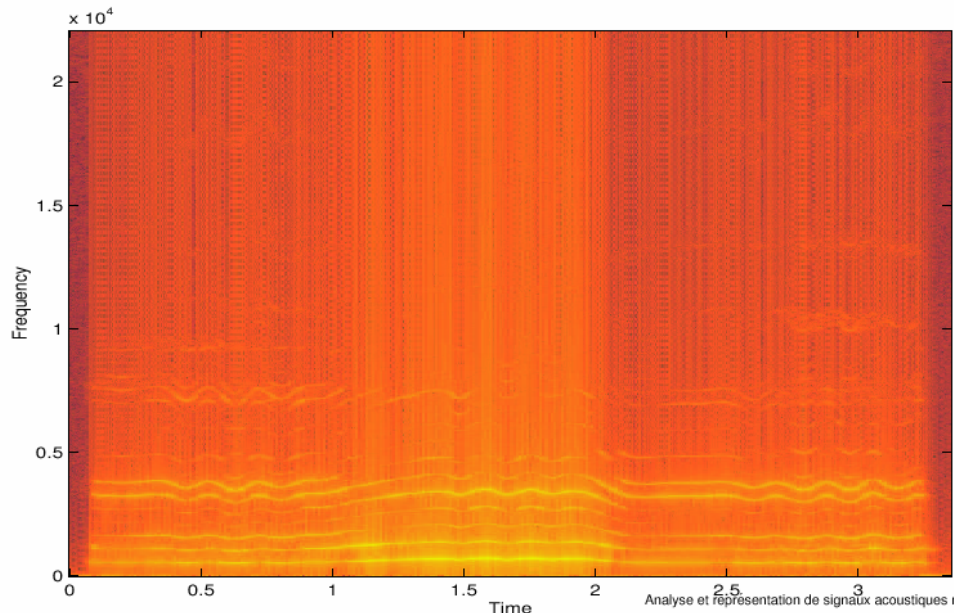


FIG 7 : Spectrogramme d'un extrait vocal (glissando C5-E5, soprano) avec une fenêtre de Hanning de 23ms

Les points importants quant à l'utilisation du spectrogramme sont donc:

- la longueur de fenêtre pour ajuster la précision temporelle, au prix d'un étalement spectral qui peut devenir rédhibitoire
- le choix de la fenêtre qui va conditionner le contraste du spectrogramme, pour une longueur de fenêtre donnée.

différents modèles de modulation

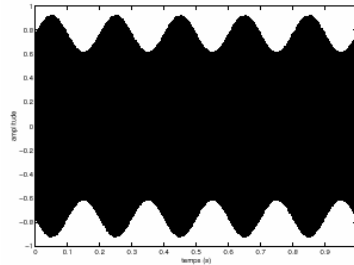
Pour faire simple, moduler un signal consiste à faire varier dans le temps une ou plusieurs des caractéristiques d'un signal stationnaire, sachant que, dans de nombreux signaux physiques, ces paramètres sont liés. Leur modulation n'est donc pas toujours exclusive.

Il est possible d'avoir une variation de l'enveloppe du son, ou une variation de sa périodicité. Les deux paradigmes qui y sont associés sont la modulation d'amplitude et la modulation de fréquence, plus généralement nommée modulation de phase.

Modulation d'amplitude

D'un point de vue mathématique, moduler un sinus en amplitude avec une « profondeur » d'environ 20% revient simplement à écrire l'amplitude du signal comme

une entité dépendant du temps, et variant autour d'une valeur moyenne avec une certaine profondeur et une certaine fréquence.



$$x(t) = A(t) \sin(2\pi f_0 t)$$

$$A(t) = A_0[1 + 0.2 \cos(2\pi f_m t)]$$

FIG 8: Principe de la modulation d'amplitude.

Vici la représentation de l'analyse avec une fenêtre de 20 ms d'une sinusoïde modulée en amplitude.

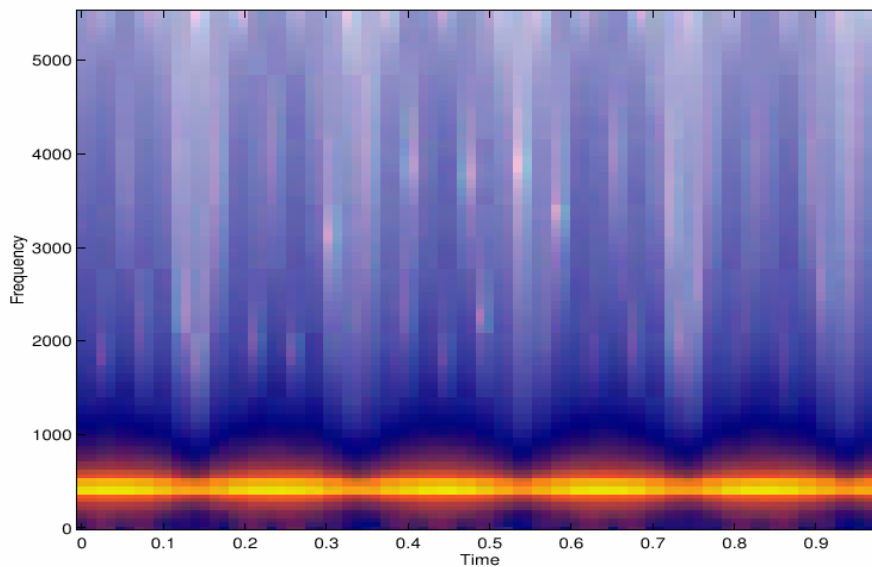
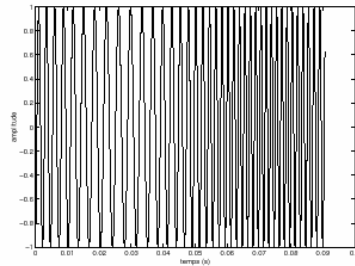


FIG 9 : Spectrogramme de la modulation d'amplitude d'une sinusoïde, avec une fenêtre de 20 ms

La modulation d'amplitude se repère par les variations de couleur sur le spectrogramme et par les éventuelles variations de fuites spectrales associées.

Modulation de fréquence ou de phase

Si on module la fréquence d'une sinusoïde avec la même profondeur que la modulation d'amplitude précédente, le signal résultant est alors le sinus d'une phase qui dépend du temps. La phase est l'intégrale de la fréquence qui, à l'inverse, est la dérivée de la phase à un facteur près. Si on souhaite reconstruire la phase à partir de la modulation de fréquence, il faut alors calculer l'aire située sous la fonction fréquence.



$$x(t) = \sin(\phi(t))$$

$$f(t) = f_0[1 + 0.2 \cos(2\pi f_m t)]$$

$$\phi(t) = 2\pi \int_0^t f(u) du$$

FIG 10 : Principe de la modulation de fréquence ou de phase

Voilà la représentation d'une sinusoïde modulée en fréquence, analysée avec une longueur de fenêtre de 0,1 s.

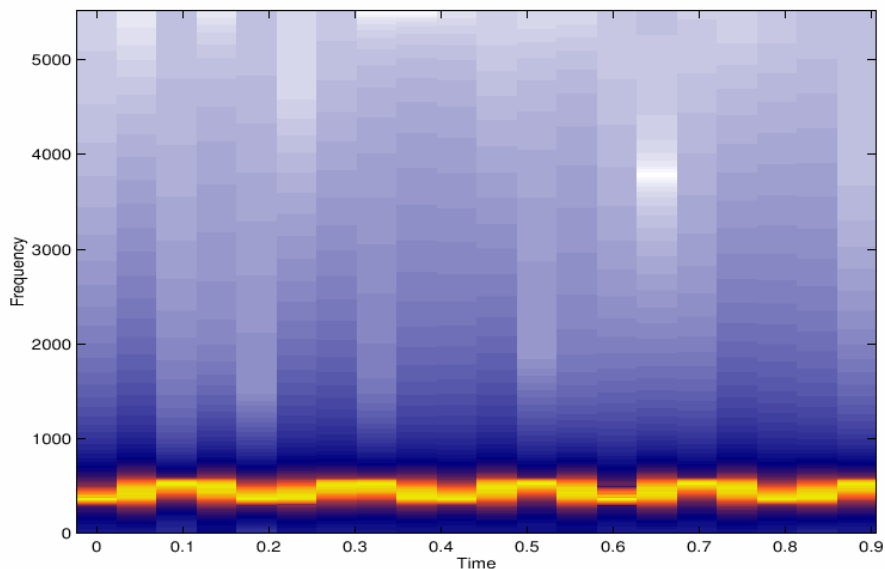
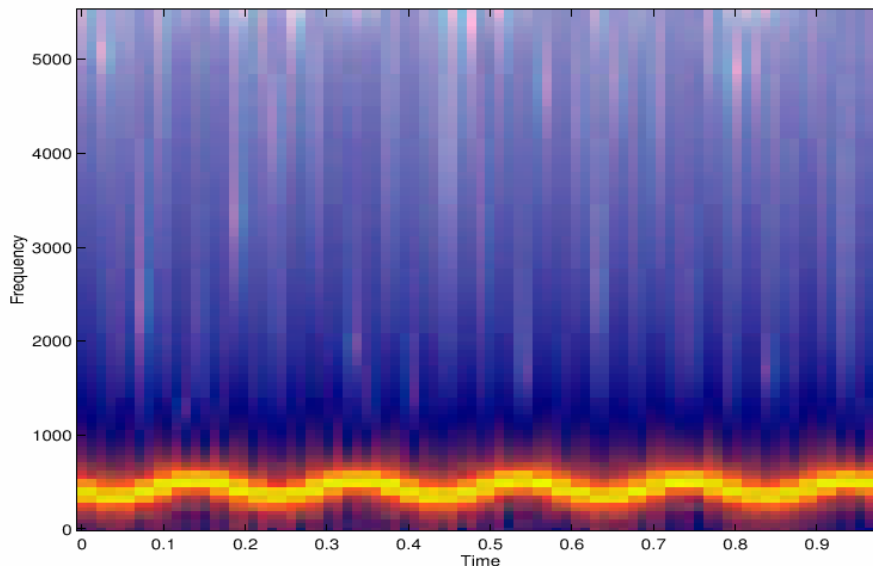


FIG 11 : Spectrogramme de la modulation en fréquence d'une sinusoïde, avec une fenêtre de 100 ms

Voilà le résultat avec une fenêtre de 20 ms.



**FIG 12 : Spectrogramme de la modulation en fréquence d'une sinusoïde, avec une
fenêtre de 20 ms**

Le deuxième spectrogramme permet de mieux d'observer la modulation de fréquence. Plus précis temporellement, il nous permet de suivre le vibrato, malgré l'étalement spectral. Il est donc nécessaire de prendre garde au choix de la longueur de fenêtre : la précision sera affectée si on diminue trop la largeur de la fenêtre d'analyse.

Il est très important de retenir qu'un compromis est indispensable à trouver entre la précision en temps et en fréquence. Si on augmente la précision en temps, la longueur de fenêtre diminue mais la précision en fréquence est amoindrie, car le spectre est étalé.

Le cas de la modulation de fréquence des signaux périodiques est légèrement différent du modèle précédent simplement sinusoïdal. On n'a alors plus seulement un sinus dans le son mais une série de partiels multiples du fondamental. Si on module le fondamental, sa période devient variable avec le temps donc la fréquence fondamentale f_1 est une modulation autour d'une fréquence f_0 moyenne, avec une certaine profondeur. Les harmoniques étant des multiples du fondamental, les fréquences des partiels de rang n sont situés autour de $n*f_0$, avec une profondeur multiple de n .

$$f_1(t) = f_0 [1 + 0,2 \cos(2\pi f_m t)]$$

$$f_n = n*f_1 = n*f_0 + 0,2 n*f_0 \cos(2 \pi f_m t)$$

Autrement dit, l'excursion va être d'autant plus grande qu'on observe les partiels de rang haut. Pour évaluer la modulation, il vaut donc mieux l'observer sur les partiels de rang élevé que sur les partiels de rang bas.

Voici un exemple synthétisé à partir de l'extrait de voix chantée précédent, analysé avec une fenêtre de Hanning de 23 ms.

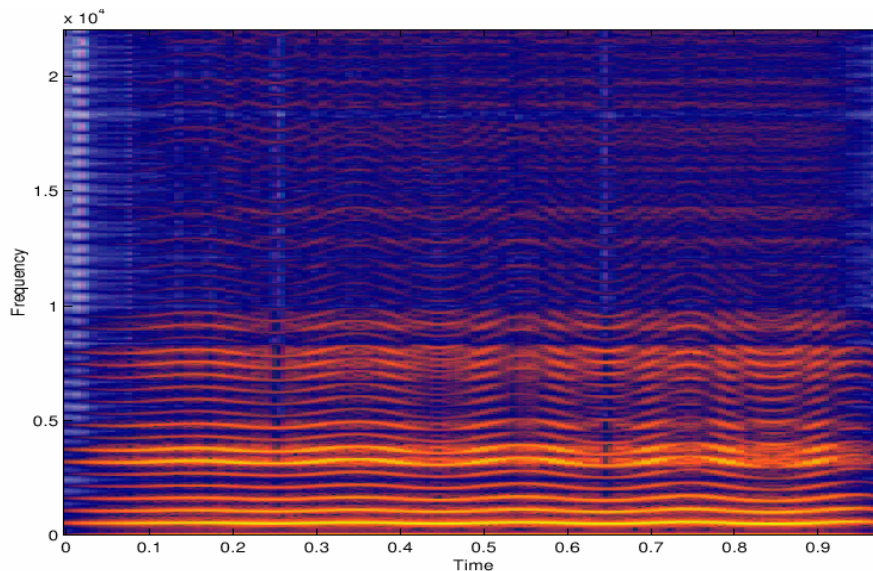


FIG 13 : Spectrogramme de la modulation en fréquence d'un son périodique synthétisé à partir d'un extrait vocal, avec une fenêtre de Hanning de 23 ms.

Grandeurs descriptives du vibrato

Pour un partiel de rang particulier, la mesure du vibrato revient à mesurer la fréquence moyenne de ce partiel et l'excursion Δf , c'est-à-dire l'écart en fréquence. La fréquence maximum sera alors $f_{\text{moy}} + \Delta f$.

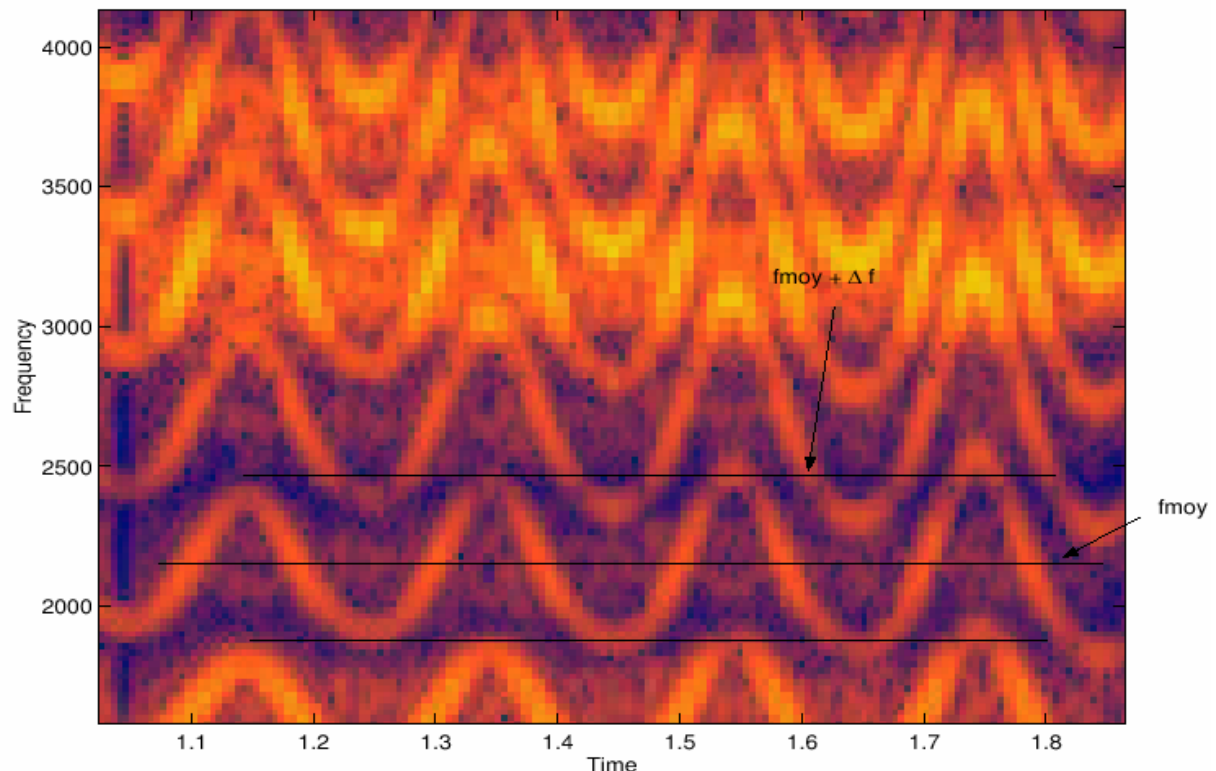


FIG 13 : Zoom sur un partiel de la figure 12

Pour mesurer un vibrato, il faut accéder à :

- sa fréquence ou rapidité, f_m exprimée en Hz. La nommer rapidité évite de confondre fréquence du signal, fréquence du vibrato et excursion. Musicalement, cette grandeur est typiquement de 5 à 7 Hz.
- la fréquence moyenne, c'est-à-dire la position du partiel étudié, de l'ordre de $n \times \text{fondamental}$ ($f_n = n f_1$), où n est le rang de l'harmonique.
- l'étendue, la profondeur ou largeur du vibrato : Cela correspond à l'excursion Δf ramenée en proportion à la fréquence de l'harmonique. Ce rapport est donc constant quel que soit le rang de l'harmonique. $e = \Delta f / f_n$. Ce rapport peut être exprimé en pourcentage, mais cela a plus de sens de l'exprimer en cents pour des sons musicaux. Dans cette unité, la profondeur s'exprime alors par la formule :

$$\text{profondeur} = 1200 \log_2 (1 + e) \text{ avec } f_{\max} = f_n (1 + e)$$

Mesure des grandeurs. Généralités

Ces outils fonctionnent bien dans pour des extraits de voix chantée, moins pour les sons où l'excursion est plus faible comme le vibrato d'instruments à cordes, à cause des problèmes liés à la fuite spectrale et à l'étalement spectral.

Il existe alors d'autres types de méthodes pour mesurer le vibrato :

- le vocodeur de phase, qui utilise la Transformée de Fourier à Court Terme pour effectuer des mesures de fréquence instantanée. Cette méthode consiste à prendre chaque canaux et à « dérouler » la phase, de façon à obtenir la fréquence dans chaque canal par dérivation. On peut ainsi suivre les fréquences instantanées
- le filtrage de Hilbert : La transformée de Hilbert consiste à effectuer un filtrage sélectif à bande latérale unique autour d'un partiel. Il faut donc pouvoir relativement bien l'isoler des autres partiels. Pour cela, on effectue un filtrage qui permet d'extraire l'information de ce partiel, et notamment sa fréquence instantanée. Ce qui nous intéresse est la phase dont on peut obtenir la fréquence instantanée par dérivation.
- la transformée de Wigner-Ville, déjà utilisée pour réaliser des mesures sur des sons de violon. Cela correspond à la transformée de Fourier de la fonction d'autocorrélation locale.
- les méthodes haute-résolution, utiles pour faire des mesures expérimentales sur les vibratos de chanteurs. On part de modèles sinusoïdaux ou quasi sinusoïdaux de signaux avec une exponentielle amortie puis on les décompose en un espace signal et un espace bruit. On peut de cette manière projeter sur l'espace signal et récupérer une information exempte de bruit et ainsi une excellente estimation de la fréquence dans chaque fenêtre. L'avantage de ces méthodes hautes résolution est qu'elles ne sont pas soumises au même compromis temps-fréquence que les méthodes vues précédemment, on peut donc réduire autant que possible la longueur de la fenêtre d'analyse.

conclusion

Cette présentation avait pour but de clarifier le compromis nécessaire entre le temps et la fréquence lors de l'utilisation d'outils d'analyse et de représentation comme le spectre et le spectrogramme.

D'autres outils de mesure des grandeurs descriptives du vibrato sont cités. Ils permettent de traiter des cas où le spectrogramme devient insuffisant.