

Rapport

Damien Tardieu

14 décembre 2005

Table des matières

1	Introduction	1
2	Addition de signaux anéchoïques	2
2.1	Energie d'une somme de signaux	2
2.2	Cas d'une somme de sinusoides	2
2.3	Conséquence sur le timbre de la somme de deux sons complexes	3
3	L'effet de la salle	3
4	Etude de sons	5
4.1	Fréquence fondamentale	5
4.1.1	Méthode	5
4.1.2	Résultats	5
4.1.3	Conclusion	5
5	Perception de sons simultanés	7
5.1	Sons purs	7
5.2	Sons complexes	7
6	Le modèle de sommation	8

1 Introduction

Lorsque deux ou plusieurs sons sont émis simultanément, ils interagissent entre eux. Cette interaction dépend de plusieurs phénomènes liés aux propriétés des émetteurs (les instruments), du transmetteur (la salle) et du récepteur (l'auditeur). Dans le contexte de l'aide à l'orchestration, nous devons modéliser cette interaction pour obtenir des opérateurs simples nous permettant de combiner des sons, ce qui signifie d'être capable, à partir de la connaissance des propriétés des sons émis, de déterminer les propriétés de la combinaison de ces sons. Nous commençons par étudier l'influence des propriétés de l'émetteur sur l'interaction, nous verrons que deux cas sont à distinguer donnant lieu soit à une sommation des signaux en énergie (ou en puissance) soit à une sommation en amplitude. Ensuite, nous étudierons le

transmetteur, la salle, nous verrons que les caractéristiques aléatoire des salle de concerts modifie l'interaction en effectuant une sorte de moyenne. récepteur, nous verrons que certaine propriété de l'oreille peuvent nous amener aussi à simplifier notre modèle. Enfin, nous décrirons le modèle de sommation de sons choisi.

2 Addition de signaux anéchoïques

2.1 Energie d'une somme de signaux

Dans le cas idéal de signaux émis sans aucun effet de salle regardons comment se comporte l'énergie de la combinaison des signaux.
Soit s_1 et s_2 deux signaux quelconques.

$$\begin{aligned} \text{Energie}(s_1 + s_2) &= E\left(\int_{T_1}^{T_2} (s_1 + s_2)^2 dt\right) \\ &= E\left(\int_{T_1}^{T_2} s_1^2 dt\right) + E\left(\int_{T_1}^{T_2} s_2^2 dt\right) + 2E\left(\int_{T_1}^{T_2} (s_1 s_2)^2 dt\right) \end{aligned}$$

Si s_1 et s_2 sont décorrélés sur $[T_1, T_2]$ alors $E(\int_{T_1}^{T_2} (s_1 s_2)^2 dt) = 0$ donc $\text{Energie}(s_1 + s_2) = \text{Energie}(s_1) + \text{Energie}(s_2)$.

Typiquement nous voyons que pour du bruit, l'addition des signaux se fait en énergie. Dans le cas de signaux corrélés nous voyons donc que la somme dépend de la corrélation. Nous allons maintenant étudier le cas qui nous interesse le plus, c'est à dire les signaux périodiques en regardant le comportement d'une somme de sinusoides.

2.2 Cas d'une somme de sinusoides

Posons $s_k(t) = A_k(t)\cos(\phi_k(t))$ ($k \in 1, \dots, N$).
et $\hat{s}(t) = \sum_{k=1}^N s_k(t) = \hat{A}(t)\cos(\hat{\phi}(t))$, en effet, nous pouvons considérer une somme de partiels comme un seul modulé en amplitude et en phase. En terme perceptif, ceci n'intervient que lorsque les différents partiels ont des fréquences proches (voir paragraphe ...).

Nous cherchons ici $\hat{A}(t)$ et $\hat{\phi}(t)$.

Ainsi,

$$\begin{aligned} \hat{s}(t) &= \sum_{k=1}^N s_k(t) \\ &= \text{Re}\left\{\sum_{k=1}^N A_k(t)e^{j\phi_k(t)}\right\} \\ &= \text{Re}\{R(t) + jI(t)\} \\ &= \hat{A}(t)\cos(\hat{\phi}(t)) \end{aligned}$$

Donc,

$$\hat{A}(t) = \sqrt{R(t)^2 + I(t)^2} \quad (1)$$

$$\hat{\phi}(t) = \text{sign}(I(t)) \arccos(R(t)/\hat{A}(t)) \quad (2)$$

Ainsi pour l'amplitude nous obtenons,

$$\hat{A}(t) = \left(\sum_k A_k(t) \cos(\phi_k(t)) \right)^2 + \left(\sum_k A_k(t) \sin(\phi_k(t)) \right)^2 \quad (3)$$

$$= 2 \sum_k \sum_{n>k} A_k(t) A_n(t) \cos(\phi_k(t) - \phi_n(t)) + \sum_k A_k^2 \quad (4)$$

ce qui nous donne pour deux partiels ($N = 2$) :

$$\hat{A}(t)^2 = 2A_1(t)A_2(t)\cos(\phi_1(t) - \phi_2(t)) + A_1(t)^2 + A_2(t)^2 \quad (5)$$

ou avec $\phi_i(t) = 2\pi f_i(t) + \Phi_i$

$$\hat{A}(t)^2 = 2A_1(t)A_2(t)\cos(2\pi(f_2(t) - f_1(t)) + \Phi_2 - \Phi_1) + A_1(t)^2 + A_2(t)^2 \quad (6)$$

Ainsi, si $f_1 = f_2$, l'amplitude est constante et égale a une valeur comprise entre $A_1(t) + A_2(t)$ et $A_1(t) - A_2(t)$ selon la différence de phase entre les signaux et en moyenne égale a $\sqrt{A_1(t)^2 + A_2(t)^2}$. Si $f_1 \neq f_2$ alors l'amplitude oscille entre $A_1(t) + A_2(t)$ et $A_1(t) - A_2(t)$ autour de $\sqrt{A_1(t)^2 + A_2(t)^2}$ à une fréquence $f_{mod} = f_2 - f_1$. En pratique, f_1 et f_2 ne sont jamais égale, le critère est donc de savoir si la période de modulation est très supérieure à la durée du son, si c'est le cas l'amplitude pourra être considérée comme constante.

2.3 Conséquence sur le timbre de la somme de deux sons complexes

Le timbre d'un son dépend en partie des relations d'amplitudes entre ses différents partiels. Or les amplitudes des partiels d'une somme de sons peuvent varier significativement, soit en produisant des modulation (ou battements) selon la différence de phase. Sachant que les relations de phase entre deux sons émis par des instruments de musique ne sont pas controlable. Par exemple, un duo jouant à l'unisson devrait produire un sons différent à chaque fois puisque les différence de phase ne sont jamias les même. Ou bien dans le cas d'une différence de fréquence suffisante, il devrait produire des battements très fort. Or, en pratique, il semble d'une part que le timbre du duo reste identique quelques soient les relations de phase et d'autre part, que si des battement peuvent être percu, il ne sont pas aussi fort que ce que laisse prévoir le paragraphe précédent.

Dans le paragraphe suivant nous allons voir comment l'effet de salle agit sur la somme de sons.

3 L'effet de la salle

Rappelons pour commencer quelques éléments d'acoustique des salles. Nous pouvons étudier la propagation du son dans une salle de deux manières différentes, soit du point de vue modale soit du point du vue "géométrique" en utilisant les rayons sonore.

Les mode propres dépendent de la géométrie de la salle, cependant, lorsque la fréquence

croit, la densité fréquentielle des modes propres augmente. Ainsi si un son pur est émis, la réponse est constituée de la contributions d'un grand nombre de mode de phase aléatoire, la répartition de l'énergie dans la salle est alors donnée par une variable aléatoire indépendante des caractéristiques de la salle. La fréquence à partir de laquelle on suppose ce modèle est appelé fréquence de Schroeder : $f_s = \sqrt{T/V}$. Ainsi, il est clair que lorsque la fréquence des signaux est supérieure a la fréquence de Schroeder, le résultat en un point de la salle est en fait la somme d'un grand nombre de composante de phase aléatoire, ainsi, la composante variable en cosinus a tendance à être moyenné et donc a s'annuler.

L'autre point de vue et le point de vue géométrique où l'on regarde le trajet des reflections du son sur les différentes parois. Au bout d'un certain temps appelé temps de mélange, le nombre de réflexion devient trop élevé pour pour pouvoir étudier chaque réflexion séparément, ici encore, l'énergie en un point de la salle est modélisé par une variable aléatoire et nous pouvons aboutir à une conclusion similaire à la précédente, c'est à dire que le son résultant est constitué de la contribution de plusieurs réflexions de phase aléatoire aboutissant pour notre duo à un moyennage.

Considérons l'interaction entre n composantes (modes propres ou réflexions) issues de plusieurs sources de même fréquence. L'amplitude résultante peut être décrite par l'équation 3.

ainsi à un instant t fixé,

$$\hat{A} = \left(\sum_k A_k \cos(\phi_k) \right)^2 + \left(\sum_k A_k \sin(\phi_k) \right)^2 \quad (7)$$

Regardons maintenant comment se comporte \hat{A} à un instant donné selon la position du récepteur (ou de manière équivalent, de la source). la phase des différente composantes désignée par la suite ϕ_k peuvent être modélisées par suite de variables aléatoires indépendantes de distribution uniforme entre 0 et 2Π .

La suite X_k telles que $X_i = \cos(\Phi_i)$ est une suite de variables aléatoires indépendantes à valeur dans $[-1, 1]$ avec $E\{X_i\} = \mu_i = 0$ et $Var\{X_i\} = \sigma_i^2$.

Maintenant posons $C = \sum_k X_k$

D'après la version de Liapounov du théorème central limite, la variable aléatoire C converge vers une loi normale de moyenne $\mu_1 = \sum_i \mu_i = 0$ et de variance $\sigma_c^2 = \sum_i \sigma_i^2$.

En faisant le même raisonnement en remplaçant le cosinus par un sinus, nous obtenons le même résultat pour le deuxième terme de \hat{A} . C'est à dire, que si nous appelons ce terme S , S converge vers une loi normale de même paramètres que C .

Ainsi $\hat{A} = C^2 + S^2$ est donc la somme de deux variables aléatoire gaussienne élevé au carré.

Ainsi nous voyons que, pour notre cas, la salle atténue fortement les modulation ou les rend plus complexe, atténue les effet de phase en faisant tendre le résultat vers une moyenne. De plus la distribution spatiale de l'énergie étant aléatoire, l'information arrivant au deux oreille est différente et chaque petit mouvement de l'instruemntiste va changer la réponse en un point.

4 Etude de sons

4.1 Fréquence fondamentale

Nous étudions ici la précision des hauteurs joués sur une base de donnée de sons. Les sons sont extraits de la base de donnée Studio En Ligne. Nous n'étudions que des instrument entretenu jouant des sons stationnaires. Les instruments, au nombre de 11 sont les suivants : Clarinette en Sib, basson, hautbois, flute, trompette en do, cor en fa, trombone, violon, alto, violoncelle, contrebasse. Chaque instrument joue toutes les notes de sa tessiture dans plusieurs mode de jeu différents à plusieurs dynamiques. Les doigtés quart de ton sont aussi analysés pour les instruments a vents. Les sons ont été enregistrés dans l'espace de projection à l'IRCAM en donnant a chaque fois une référence à l'instrumentiste et éventuellement, pour certain doigté difficile (quart de ton), en cherchant la manière de jouer qui permet la note la plus juste possible. Ainsi nous pouvons considérer que les résultat obtenu dans la suite de ce chapitre donne une bonne idée de la justesse des instrumentiste en situation de concert, situation dans laquelle les instrumentistes ne peuvent pas porter autant d'attention à la justesse mais se corrigent en s'écoutant les uns les autres.

4.1.1 Méthode

La fondamentale de chaque son est analysée par l'algorithme f_0 [DR93]. Pour éviter les erreur d'octave, la recherche est limitée à un petit intervalle autour de la fréquence fondamentale de référence donnée par la note jouée. Pour chaque sons, nous sélectionons la partie stable en retirant les zones contenant des écart à la médiane de plus d'un quart de ton puis en selectionnant 1 seconde au milieu du son.

4.1.2 Résultats

La figure 1 présente un histogramme de l'erreur relative entre la fréquence mesurée et la fréquence de référence. La moyenne de l'erreur est d'environ .12 % et la deviation standard environ .74 %.

La figure 2 montre l'erreur en fonction de la hauteur. Tout d'abord nous observons que l'erreur croit avec la fréquence. Pour les notes de fréquence fondamentale inférieur à 1000Hz (C6), l'erreur reste a peu près constante inférieur à 1Hz , au delà elle croit fortement pour atteindre entre 20 et 30 Hz autour de 3000Hz (G7). Notons que les valeurs d'erreur pour des fréquence élevée sont doivent être analysé avec précaution étant donné le faible nombre d'échantillons étudier pour ces fréquences. Cependant ces observations sont en accord avec les études sur la discrimination fréquentielle de l'oreille [Moo03] qui montrent que celle-ci décroît avec la fréquence.

4.1.3 Conclusion

Pour notre problème le résultat principal est qu'en dessous de 1000Hz les erreur sont inférieur à 1Hz , et peuvent donc donner lieu à des battements lents. Ainsi, les résultats que nous obtenons ne sont qu'une borne supérieure de l'erreur. En ce qui concerne les hautes fréquences, nous avons vu que les erreur pouvait être significativement plus grande. Ces écart

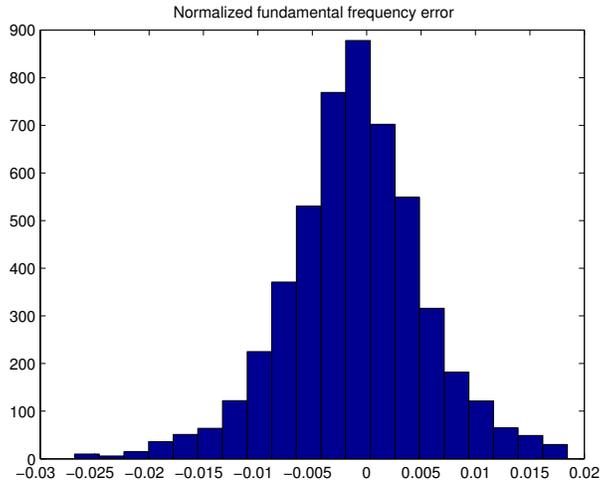


FIG. 1 – Erreur normalisée de F0

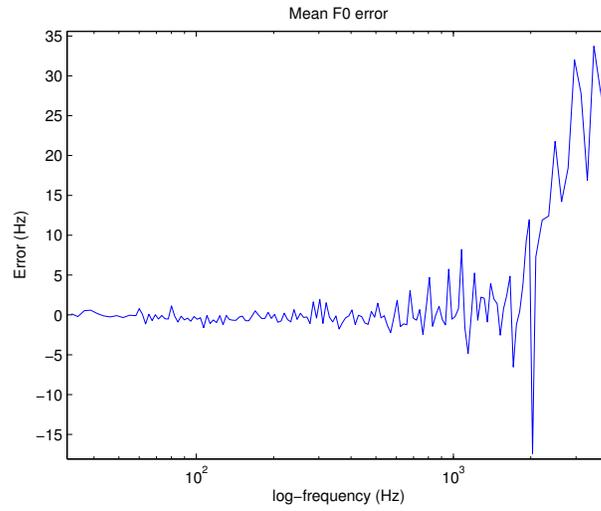


FIG. 2 – Erreur de F0 en fonction de la hauteur

donne lieu à des battements rapides qui ne se traduisent pas en un niveau de rugosité élevé car les bandes critiques en haute fréquence sont très larges.

5 Perception de sons simultanés

5.1 Sons purs

L'émission simultanée de deux sons purs résulte donc, en terme acoustique, en la production de battements. La perception de ce phénomène a été largement étudiée (voir [Pre98]), à commencer par Helmholtz. Celui-ci distingue trois zones différentes selon l'écart fréquentiel des sons. Si l'écart est faible, les battements sont entendus clairement, lorsque l'écart augmente, leur perception s'estompe, il devient par exemple impossible de les compter, et apparaît une sensation que Helmholtz appelle la rugosité. Enfin au delà d'un certain écart les deux sons peuvent être perçus séparément.

L'étude de ces phénomènes a depuis donné lieu à plusieurs interprétations (press) en particulier les interprétations dites spectrales et temporelles. Nous nous intéresserons pour le moment à l'interprétation spectrale. Celle-ci postule que la rugosité peut être prédite par la composition spectrale des sons, que pour deux sons purs, la rugosité disparaît lorsque ceux-ci sont résolus par l'oreille interne, c'est à dire lorsque l'écart fréquentiel est supérieur à une bande critique à la fréquence moyenne considérée, et enfin que le maximum de rugosité est atteint pour un écart fréquentiel approximativement égal à une fraction constante de la largeur de la bande critique à la fréquence moyenne considérée.

Ces études nous permettent d'isoler trois zones et d'en préciser les limites :

- $\Delta f < 3 - 4Hz$, les battements sont perçus.
- $3 - 4Hz < \Delta f < 110\%$ bande critique, perception de rugosité, avec un maximum de rugosité à 45% de la bande critique.
- $\Delta f > 110\%$ de la bande critique, les deux sons sont entendus séparément.

5.2 Sons complexes

Pour des sons complexes, les théories et les résultats divergent beaucoup plus, sans doute du fait de la difficulté de mise en place d'expérience et d'analyse des résultats. Cependant, les résultats s'accordent généralement sur le fait que si deux sons modulés en amplitude sont émis simultanément à des fréquences différentes, l'incohérence entre les deux modulations amène à une diminution de la détection des modulations. Ainsi Pressnitzer et McAdams ont aussi montré que la combinaison de deux sons modulés en amplitude de manière à produire de la rugosité était moins rugueuse lorsque les deux modulations étaient incohérentes.

Tout d'abord, en supposant que les deux sons sont stationnaires et harmoniques et présentent un écart de fréquence fondamentale Δf faible, l'écart au partiel de rang k sera de $k\Delta f$. Ainsi les battements pour chaque partiel n'auront pas la même fréquence, nous ne les percevons pas. En réalité, les sons possèdent des microvariations de fréquence et d'amplitude

6 Le modèle de sommation

Références

- [DR93] B. Doval and Xavier Rodet, *Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and hmms*, Proc. IEEE-ICASSP, 1993, pp. 221–224.
- [Moo03] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed., Academic Press Limited, 2003.
- [Pre98] Daniel Pressnitzer, *Perception de rugosité psychoacoustique : D'un attribut élémentaire de l'audition à l'écoute musicale*, Thèse de doctorat, Université Paris VI/IRCAM, Paris, 1998.

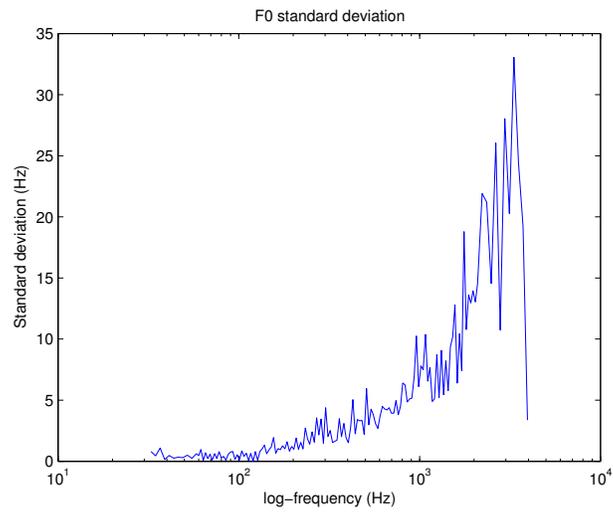


FIG. 3 – Deviation standard de la F0 en fonction de la hauteur