

Iana~ : INSTRUMENT D'ANALYSE EN TEMPS REEL SELON DES CRITERES PSYCHO-ACOUSTIQUES

Todor Todoroff

Art, Recherche, Technologie et Musique (ARTEM)

artem@skynet.be

1. Introduction

Un certain nombre d'expériences psycho-acoustiques ont permis de mieux comprendre le phénomène de la perception auditive. Pierre Schaeffer, fondateur de la musique concrète, en avait eu l'intuition dès 1948 : la façon dont nous percevons les sons n'est pas décrite convenablement par les appareils de mesure de laboratoire. Notre perception de l'amplitude, de la fréquence et du timbre diffère parfois considérablement des valeurs mesurées.

On sait ainsi qu'une composante fréquentielle de forte amplitude a tendance à masquer les composantes voisines, c'est-à-dire à en atténuer la perception ou à les rendre inaudibles. On démontre également que nous sommes capable de percevoir la fréquence d'un son dont on a éliminé la fondamentale. Nous faisons cette expérience chaque jour au téléphone : nous suivons parfaitement les intonations vocales bien que la transmission téléphonique supprime les composantes de fréquence inférieure à environ 350 Hz. D'autre part, la fréquence perçue d'un son peut être différente de la fréquence mesurée : elle dépend de sa structure spectrale.

Jean-Claude Risset, un des pionniers de l'informatique musicale a très tôt révélé certains paradoxes de la perception auditive [Risset, 71, 76 et 81]. Le psycho-acousticien allemand Ernst Terhardt a proposé un modèle de la perception auditive qui tient compte des connaissances actuelles en matière d'effet de masque et de perception de la fréquence ainsi qu'une méthode de calcul que nous utilisons dans iana~ [Terhardt, Stoll & Seewann, 82a et 82b]. D'autres références sont disponibles dans [Terhardt, 73], [Zwicker & Terhardt, 80] et [Zwicker, 80]. Il est ainsi possible d'analyser un son, d'en extraire toutes les composantes, de déterminer celles qui sont audibles, l'intensité avec laquelle elles sont perçues et comment elles interviennent dans la perception de la fréquence fondamentale. C'est cette analyse que le module iana~ présenté ici permet de réaliser en temps réel dans l'environnement MAX/MSP.

Ce module temps réel s'inspire de l'algorithme en temps différé programmé à l'IRCAM en 1985 par Gérard Assayag et Dan Timis sous le nom de iana (instrument d'analyse). Ce programme a immédiatement suscité l'intérêt de nombreux compositeurs, notamment par la simplicité de sa mise en œuvre et par la pertinence des résultats fournis.

A l'époque, la puissance des ordinateurs ne permettait pas d'envisager une application en temps réel. Les compositeurs devaient donc échantillonner les sons et les faire analyser par le programme iana en temps différé. Il obtenait alors un fichier textuel contenant les résultats de l'analyse. Le compositeur pouvait utiliser

ces données de manière symbolique dans sa composition, ou les modifier en mode texte et les transférer vers une autre machine pour, par exemple, réaliser une synthèse. Les sons ainsi synthétisés étaient alors transférés sur bande magnétique ou sur système d'enregistrement sur disque dur et déclenchés manuellement pendant l'exécution du concert. Un des problèmes de cette méthode, en dehors de la lourdeur du processus, est le fait que les instrumentistes ne jouent jamais exactement de la même manière en studio et en concert, de sorte que les sons de synthèse ainsi réalisés ne correspondent jamais parfaitement à la réalité sonore de l'exécution en concert.

L'implantation de l'algorithme de Terhardt sous forme d'un module Max externe permet enfin de réaliser toutes ces étapes en temps réel dans un seul environnement. Le signal peut être simultanément échantillonné, analysé et utilisé pour commander la synthèse d'événements sonores. Les sons synthétisés suivent alors exactement l'évolution timbrale de l'instrument pendant le concert. Le compositeur Joshua Fineberg a été le premier à l'utiliser à l'IRCAM dans une version MAX/FTS tournant sur la Station d'Informatique Musicale (NeXT + ISPW) programmée à la Faculté Polytechnique de Mons [Todoroff, Daubresse & Fineberg, 95]. La collaboration avec l'assistant musical Éric Daubresse et le compositeur Joshua Fineberg s'est avérée particulièrement fructueuse dans la finalisation d'un module qui répond plus exactement aux attentes des compositeurs.

Iana~ permet d'envisager de nouveaux types d'interaction entre un instrument, ou un orchestre et un ordinateur. Alors que les possibilités de détection se limitaient jusqu'à présent à l'extraction de la fréquence fondamentale et de l'enveloppe d'amplitude d'un instrument monophonique, iana~ fournit les listes des composantes fréquentielles présentes dans le spectre de l'instrument ou du groupe d'instruments analysé. Celles-ci peuvent être classées en fonction de critères perceptifs de manière à fournir aux compositeurs une information pertinente sur leur importance respective. Le contenu spectral des sons de l'orchestre ainsi analysé en direct peut donc commander en temps réel la synthèse et la transformation des sons. Le compositeur peut également préparer des ensembles de configurations de paramètres qu'il peut rappeler au vol pour adapter l'analyse à toute variation de l'orchestration au cours du déroulement du concert.

Cette nouvelle version, dans MAX/MSP, constitue une nouvelle amélioration puisque tous les paramètres, sans exception, sont à présent accessibles en temps réel. De cette manière, la phase d'ajustement des paramètres qui pouvait prendre des heures dans la version temps différé est encore raccourcie par rapport à la version ISPW.

2. Principes de fonctionnement et étapes de l'analyse

Calcul du spectre du signal par transformée de Fourier rapide (FFT).

Calcul du niveau de pression sonore (Sound Pressure Level, ou SPL).

Extraction des composantes fréquentielles par extraction des pics du spectre de niveau de pression sonore. Remarquons que rien n'impose que ces pics soient ordonnés dans des rapports harmoniques, c'est pourquoi nous les appellerons "partiels".

Interpolation par polynôme du troisième ordre afin de calculer la fréquence et le niveau de chaque partiel avec plus précision. Si le niveau du pic est supérieur au seuil d'audition (courbe empirique déduite de tests perceptifs), le partiel est retenu, sinon il est rejeté.

Calcul des effets de masque et du niveau perceptif appelé SPL excess. Il permet de vérifier selon des critères empiriques quelles composantes peuvent réellement être entendues individuellement dans le cadre d'une écoute analytique. Seules ces composantes peuvent avoir un effet sur la perception de la fréquence fondamentale. De plus, la prise en compte de ces niveaux permet de réduire le nombre de composantes nécessaires pour resynthétiser le son, puisque les composantes obtenant un niveau perceptif inférieur à zéro ne sont théoriquement pas audibles.

Calcul des poids perceptifs relatifs des différentes composantes. Il s'agit de déterminer leurs poids respectifs dans la perception de la fréquence fondamentale. Deux facteurs interviennent : l'un tient compte du fait que le poids perceptif sature au-delà de 20 dB de niveau perceptif (SPL excess) ; l'autre tient compte du fait que la fréquence perçue dépend majoritairement des partiels présents dans la bande des fréquences couvertes par la voix humaine.

Emission des données sont à la sortie de iana~ sous forme d'une suite de listes classées selon des critères perceptifs et contenant, pour chaque partiel, son rang, sa fréquence, son niveau et, au choix, son niveau perceptif ou son poids.

3. Mise en œuvre

Nous avons fait en sorte que tous les paramètres, sans exception, puissent être modifiés en temps réel. C'est en effet l'un des intérêts majeurs de cette réalisation : l'utilisateur peut à tout moment ajuster les valeurs des paramètres d'analyse et trouver rapidement celles qui répondent le mieux à ses besoins. Comme le résultat de l'analyse peut directement être utilisé pour commander des processus de synthèse ou de transformation des sons, il gagne un temps considérable et peut aisément obtenir un résultat optimal.

Nous décrivons ici en détail les différentes étapes de l'analyse et les messages qui permettent d'en modifier les paramètres.

3.1. Sous-échantillonnage (optionnel)

Le signal peut être décimé dans un rapport 2 ou 4 après filtrage par un filtre elliptique IIR. La décimation permet d'augmenter la résolution fréquentielle de la FFT pour une taille identique, tout en réduisant proportionnellement la bande passante de l'analyse. Cette opération permet de répartir la charge de calcul plus uniformément dans le temps. Etant donné la puissance de calcul actuellement disponible sur les PowerPC Macintosh, cette fonctionnalité présente moins d'intérêt que sur la NeXT + ISPW. Elle a principalement été conservée par souci de compatibilité. Le tableau suivant montre les valeurs pour une FFT de 1024 points à 44.1 kHz :

Facteur de décimation	Résolution fréquentielle	Bande passante
1	43,07 Hz	0 à 22050 Hz
2	21,53 Hz	0 à 11025 Hz
4	10,77 Hz	0 à 5512,5 Hz

L'utilisateur peut définir le facteur de sous-échantillonnage désiré à la création du module et le modifier ensuite en temps réel à l'aide du message ***set_down_factor***.

3.2. Fenêtrage

L'utilisateur a le choix entre les fenêtres suivantes: Hamming, Hanning, Blackman, triangulaire, cosinoïdale ou rectangulaire. La multiplication du signal par une fenêtre avant analyse par FFT permet de réduire les effets de bord. Chaque fenêtre présente des caractéristiques différentes dont les deux principales sont la largeur spectrale (base du lobe principal) et le rapport d'amplitude entre le lobe principal et le premier lobe secondaire [Harris, 76]. L'utilisateur peut changer de fenêtre à tout moment en envoyant le message ***set_window_type***. La fenêtre de Blackman est sélectionnée par défaut.

3.3. Analyse par FFT et calcul du niveau de pression sonore

Le spectre du signal est calculé par transformée de Fourier rapide (FFT) à 128, 256, 512, 1024, 2048, 4096, 8192 ou 16384 points. Les petites tailles de FFT ont été autorisées pour la réalisation d'effets particuliers, mais il est illusoire d'obtenir une analyse qui reflète un tant soit peu le signal d'entrée en dessous de 1024 points. L'utilisateur définit la taille maximale désirée à la création du module afin de réserver l'espace-mémoire qui ne peut être alloué dynamiquement. La taille effective peut être modifiée en temps réel dans cette limite à l'aide du message **set_fft_size**. Le tableau suivant montre la résolution fréquentielle en fonction des différentes tailles de FFT à 44.1 kHz, sans décimation :

Taille FFT [points]	128	256	512	1024	2048	4096	8192	16384
Résolution [Hz]	344,53	172,27	86,13	43,07	21,53	10,77	5,38	2,69

Le niveau de pression sonore (Sound Pressure Level) est ensuite calculé pour chaque point de la FFT :

$$\text{SPL}_i = \text{ref_level} + 10 * \log_{10} (\text{Re}_i^2 + \text{Im}_i^2)$$

Le terme `ref_level`, modifié par l'utilisateur à l'aide du message **set_ref_level**, permet de prendre en compte les conditions d'enregistrement ou d'amplification du signal. En effet, tous les calculs psycho-acoustiques sont basés sur des valeurs absolues (le seuil d'audition, les effets de saturation), de sorte que les calculs des effets de masque et des poids n'ont de sens que s'ils prennent en compte les conditions d'écoute. Pour plus de facilité, ce paramètre n'a aucune influence sur les niveaux envoyés dans les listes à la sortie de `iana~`.

3.4. Extraction des partiels - interpolation - application du seuil d'audition

La recherche des partiels s'effectue entre deux points définis à l'aide des messages **set_first_to_search** et **set_last_to_search**. Par défaut, la recherche s'effectue entre 0 et (taille de la FFT)/2.

Les maxima locaux du spectre sont détectés à l'aide d'une procédure d'extraction de pics dont le pouvoir discriminant est déterminé par les paramètres de largeur à l'aide du message **set_width** (nombre de bandes fréquentielles à prendre en considération de part et d'autre d'un candidat pic potentiel) et de seuil à l'aide du message **set_threshold** (différence minimale en décibels entre le sommet du pic et les composantes fréquentielles adjacentes).

$$\begin{aligned} L_{i-1} < L_i \geq L_{i+1} \\ L_i - L_{i+j} \geq \text{threshold} \quad \forall j = - \text{width}, \dots, + \text{width} \end{aligned}$$

Lorsqu'un pic est trouvé, on procède à une interpolation polynomiale du troisième ordre afin de déterminer avec précision son amplitude (Sound Pressure Level ou

L_μ) et sa fréquence centrale (f_μ).

Les tests ont montré, en l'absence de bruit de fond, une détection de la fréquence d'une composante unique avec une précision supérieure à 1 Hz en utilisant une fenêtre de Blackman et une FFT de 2048 points sans décimation à 44.1 kHz. Cette bonne détection malgré une résolution de la FFT de seulement 21 Hz (voir tableau ci-dessus) montre l'importance de cette phase d'interpolation. Notons que cette interpolation ne fonctionne bien que si les composantes du signal analysé sont situées dans des bandes suffisamment éloignées les unes des autres, de sorte que la taille de FFT nécessaire à une bonne analyse dépend fortement de la composition du signal à analyser.

Les partiels d'amplitude inférieure au seuil d'audition sont écartés par défaut, afin de ne pas gaspiller inutilement de la puissance de calcul avec des partiels inaudibles. Il est toutefois possible de les prendre en compte à l'aide du message **keep_peaks_under_TH**.

Le seuil d'audition (Threshold of Hearing, L_{TH}) peut être obtenu à partir de la formule empirique suivante:

$$L_{TH}(f_\mu) = \left\{ \begin{array}{l} 3.64 (f_\mu / \text{kHz})^{-0.8} \\ -6.5 \exp [-0.6 (f_\mu / \text{kHz} - 3.3)^2] \\ + 10^{-3} (f_\mu / \text{kHz})^4 \end{array} \right\} \text{ dB}$$

Elle établit une correspondance entre la fréquence et l'intensité minimale perçue par l'oreille humaine.

A la fin de cette phase d'analyse, on obtient un maximum de N partiels définis à l'aide du message **set_nmaxpeaks**.

3.5. Calcul de l'amplitude perceptive des composantes

Ernst Terhardt propose une formule permettant de calculer le Sound Pressure Level Excess (LX), estimation de l'intensité avec laquelle chacun des partiels est perçu par un auditeur moyen. Il s'agit d'une formule empirique, obtenue à la suite d'expériences de perception, qui tient compte de l'effet de masque produit sur chacune des composantes μ par l'ensemble des autres:

$$LX_\mu = L_\mu - 10 \log_{10} \left[\left(\sum_{v=1, v \neq \mu}^N 10^{L_{EV}(f_\mu) / (20 \text{ dB})} \right)^2 + I_{N\mu} + 10^{L_{TH}(f_\mu) / (10 \text{ dB})} \right] \text{ dB}$$

$$\forall v = 1, \dots, \mu-1, \mu+1, \dots, N$$

Le terme $I_{N\mu}$ correspond à l'intensité du bruit présent dans la bande critique qui contient le partiel μ afin de tenir compte de l'effet de masque produit par un spectre bruité. Le calcul de ce terme n'avait pas été implanté dans la version en temps différé programmée à l'IRCAM. Comme les compositeurs désirent obtenir exactement les mêmes résultats avec notre module, nous l'avons également ignoré. Par contre, comme une composante donnée exerce un effet de masque à plus longue distance vers les fréquences plus élevées que vers les fréquences plus basses, nous avons réduit la charge de calcul en limitant son effet sur les 15 composantes plus aiguës et sur les 5 composantes plus graves. Des comparaisons avec les résultats du programme temps différé qui calcule toutes les interactions montrent que cette simplification se justifie car elle ne produit aucune modification significative des valeurs.

Lorsque la formule fournit une valeur LX_μ négative, on en déduit que le partiel μ est inaudible et on l'élimine de la suite des calculs.

3.6. Calcul du poids des partiels dans la perception de la fréquence fondamentale

Il y a un effet de saturation aux alentours de 20 dB de SPL Excess: au-delà de cette intensité, l'influence du partiel ne croît plus dans la perception de la fréquence fondamentale. De plus, la région spectrale dominante dans la perception de la fréquence fondamentale correspond globalement à celle qui régit la perception de la parole.

Ces deux effets sont modélisés respectivement par le premier et le second facteur de l'équation suivante :

$$WS_\mu = \left[1 - \exp(-LX_\mu / 15 \text{ dB}) \right] \left[1 + 0.07 \left((f_\mu / 0.7 \text{ kHz}) - (0.7 \text{ kHz} / f_\mu) \right)^2 \right]^{-1/2} \quad \forall LX_\mu \geq 0$$

Comme une composante inaudible ne peut influencer la perception de la fréquence fondamentale, il est évident que $WS_\mu = 0$ si $LX_\mu < 0$.

3.7. Présentation des données à la sortie de iana~

Le module fournit à sa sortie principale une suite de listes contenant chacune le rang du partiel (n), sa fréquence (f) son intensité mesurée (L), et, en option, son niveau perçu (LX) ou son poids (WS). Les messages **list_nfl**, **list_nfLLX** et **list_nfLWS** définissent la composition des listes.

L'utilisateur détermine à l'aide des messages **all_peaks** et **LX_peaks** s'il désire que iana~ émette tous les partiels ou seulement ceux qui sont perçus selon la théorie, pour lesquels $LX > 0$.

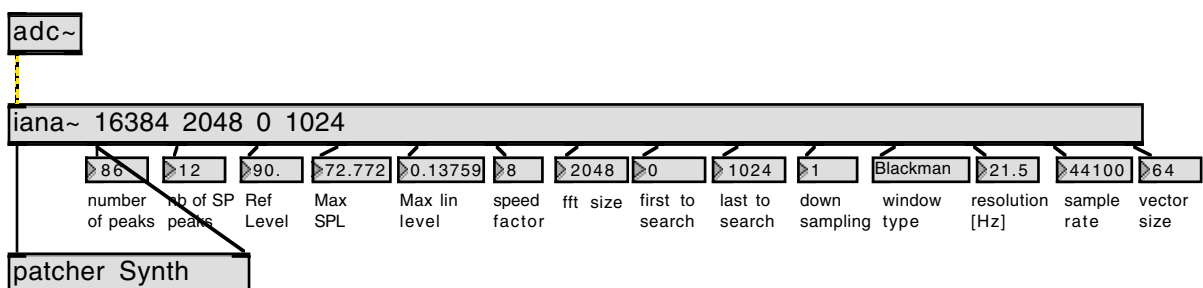
Le message **set_nmax_list_peaks** permet de limiter nombre de listes émises, de

manière à l'adapter à la capacité du synthétiseur.

Les listes sont ordonnées selon la fréquence, le niveau, le niveau perceptif ou le poids. Le choix du critère d'ordonnement se fait par envoi d'un des messages suivants à l'entrée de iana~ : **f_order** (classement par ordre de fréquences croissantes), **L_order** (classement par ordre de niveaux décroissants), **LX_order** (classement par ordre de niveaux perceptifs décroissants), **WS_L_order** (classement par ordre de poids décroissants lorsque les poids sont positifs et par ordre de niveaux décroissants pour les pics restants) et **WS_LX_order** (classement par ordre de poids décroissants lorsque les poids sont positifs et par ordre de niveaux perceptifs décroissants pour les pics restants).

Le niveau (L) de l'ensemble des partiels peut être modifié par le message **set_out_level** suivi d'un volume exprimé en décibels. Et leurs fréquences peuvent être transposées ou décalées à l'aide des messages **set_transpo** et **set_shift**. Il est très important de noter que ces 3 derniers paramètres modifient les valeurs APRES analyse et prise en compte des facteurs psycho-acoustiques. Ils ont été introduits pour faciliter une expérimentation créative en dehors du champ de validité de la théorie, sachant que ces listes seront en général utilisées pour commander des bancs d'oscillateurs.

Les autres sorties permettent de visualiser en permanence les paramètres suivants : le nombre de partiels extraits, le nombre de partiels perçus selon la théorie (LX>0), le niveau de référence, le niveau maximum détecté en dB (dépendant du niveau de référence, utile pour vérifier le domaine d'application de la théorie), le niveau maximum détecté linéaire (ne dépendant pas du niveau de référence, utile pour le réglage du synthétiseur), le facteur de vitesse de calcul, la taille de la FFT, le premier et le dernier point du spectre sur lequel s'effectue la recherche des partiels, le facteur de décimation, le type de fenêtre choisi, la résolution résultante avant interpolation, la fréquence d'échantillonnage et la taille des vecteurs. La figure suivante montre l'entrée et les sorties.



4. Récapitulation des arguments de création, des messages et des valeurs par défaut

Chaque paramètre de *iana~* reçoit une valeur par défaut lors de la création du module. Le module est donc immédiatement opérationnel, mais toutes ces valeurs peuvent être modifiées à tout moment par l'envoi de messages.

4.1. Arguments de création de iana~

Argument 1 (obligatoire) : taille maximale de la FFT, comprise entre 128 et 16384 points ; cette taille est utilisée pour la réservation d'espace de mémoire de *iana~* et ne pourra pas être dépassée par les messages subséquents ;

Argument 2 (optionnel) : taille de la FFT, comprise entre 128 points et la taille maximale définie par le premier argument, pris par défaut ;

Argument 3 (optionnel) : premier point à prendre en compte pour la recherche de partiels (0 par défaut) ;

Argument 4 (optionnel) : dernier point à prendre en compte pour la recherche de partiels (demi taille de la FFT par défaut) ;

Argument 5 (optionnel) : facteur de décimation qui peut prendre les valeurs 1, 2 ou 4 (1 par défaut).

4.2. Paramètres d'analyse

set_fft_size suivi d'un entier modifie la taille de la FFT avec pour maximum la taille maximale définie à la création du module (prise par défaut) ; les valeurs admises sont : 128, 256, 512, 1024, 2048, 4096, 8192 ou 16384 ;

set_down_sampling suivi d'un entier de valeur 1, 2 ou 4, fixe le facteur de décimation (1 par défaut) ;

set_window accepte les chaînes de caractères Hamming, Hanning, Blackman, triangle, cosine, rectangle (Blackman par défaut) ; il est également possible d'envoyer les entiers 1, 2, 3, 4, 5 ou 6 ;

set_width suivi d'un entier entre 1 et 16 permet de définir le nombre de points de la FFT qui seront examinés par l'algorithme de détection des pics à gauche et à droite du point testé (2 par défaut) ;

set_threshold suivi d'un réel entre 0 et 15 exprime, en décibels, la différence minimale de niveau requise entre le sommet du pic et tous les points adjacents pris en compte pour pouvoir accepter un candidat pic (2 par défaut) ;

set_nmax_peaks suivi d'un entier définit le nombre maximum de pics à extraire du spectre ; ce nombre doit être limité afin de réduire la charge de calcul. Le maximum est fixé à 200 (100 par défaut) ;

set_ref_level suivi d'un réel permet d'adapter l'entrée du module d'analyse aux conditions d'enregistrement ou d'amplification ; l'argument est exprimé en décibels (90 par défaut) ;

keep_peaks_under_TH suivi de 0 ou 1 permet de choisir si on ignore ou si on conserve les partiels d'amplitude inférieure au seuil d'audition (0 par défaut) ;

man_trigger fixe le mode de déclenchement manuel de l'analyse. Celle-ci est alors initiée par l'envoi d'un message *bang* à l'entrée du module (c'est le mode choisi par défaut) ;

auto_trigger fixe le mode de déclenchement automatique du module : dès qu'une analyse est terminée, l'analyse suivante débute ;

bang démarre une analyse en mode manuel.

4.3. Paramètres de définition des messages générés par le module

set_nmax_list_peaks suivi d'un entier définit le nombre maximum de partiels, limité à 200 (40 par défaut) dont les informations seront présentées à la sortie. Si l'on désire par exemple utiliser la sortie de l'algorithme pour synthétiser les sons à l'aide d'un banc de 10 oscillateurs, il est inutile de surcharger Max en envoyant plus de 10 listes par analyse ;

all_peaks ordonne l'émission de tous les partiels détectés, dans la limite du nombre maximum déterminé par le paramètre précédent (choix par défaut) ;

LX_peaks ordonne de n'émettre que les partiels non masqués ($LX > 0$), toujours dans la limite du nombre maximum ;

f_order informe le module que les partiels devront être ordonnés selon les fréquences croissantes (choix par défaut) ;

L_order informe le module que les partiels devront être ordonnés selon les intensités (SPL) décroissantes ;

LX_order informe le module que les partiels devront être ordonnés selon les intensités après calcul des effets de masques (SPL Excess) décroissantes ;

WS_L_order informe le module que les partiels devront être ordonnés selon les

poids décroissants. Les pics restants (masqués et donc de poids nuls) sont classés par ordre des intensités (SPL) décroissantes;

WS_LX_order informe le module que les partiels devront être ordonnés selon les poids décroissants. Les pics restants (masqués et donc de poids nuls) sont classés par ordre des intensités perçues (SPL Excess) décroissantes ;

list renvoie les listes correspondant à la dernière analyse, ordonnée suivant le dernier message de type **xx_order** reçu ; ce message permet de tester l'effet d'un autre critère de classement des paramètres sur le même ensemble de partiels ;

print imprime dans la fenêtre MAX les résultats de la dernière analyse ; afin d'éviter des clics, cette commande ne fonctionne pas lorsqu'une analyse est en cours ;

set_out_level suivi d'un réel définit, en décibels, un niveau d'amplification des amplitudes des listes de paramètres envoyées à la sortie du module ;

set_transpo suivi d'un réel définit un facteur de transposition, exprimé en demi-tons, des fréquences des listes de paramètres envoyées à la sortie du module ;

set_shift suivi d'un réel définit, en Hz, un facteur de décalage des fréquences des listes de paramètres envoyées à la sortie du module ;

4.3. Paramètre de vitesse de calcul

set_speed_factor suivi d'un entier de 1 à 8 spécifie un facteur de division du temps d'analyse qui permet de limiter la puissance de calcul nécessaire au détriment de la vitesse d'analyse. 8 spécifie la vitesse la plus rapide, choisie par défaut.

5. Développements futurs

En partant des partiels ainsi extraits, avec leur niveau mesuré, leur niveau perçu et leur poids, on peut également estimer la fréquence qui sera perçue. Cette dernière étape est un peu plus complexe, car cette fréquence dépendra non seulement du signal, mais également de l'intention d'écoute : analytique ou globalisante.

Deux mécanismes de perception de la fréquence entrent en effet en compétition et chacun propose un certain nombre de candidats potentiels. On définit ainsi le "Spectral Pitch Pattern" et le "Virtual Pitch Pattern". Le premier correspond à l'intensité de chacune des composantes spectrales après calcul des effets de masque. Le second tient compte de la théorie dite "des résidus", qui explique par exemple la perception de la fréquence de la voix au téléphone malgré une bande passante réduite qui ne laisse pas passer la fondamentale.

Dans le cas de sons complexes, contenant des composantes inharmoniques, il se peut que l'on puisse entendre plusieurs fréquences selon l'écoute. Le programme fournirait alors les différents candidats assortis de poids proportionnels à la probabilité avec laquelle ils seront entendus. On pourrait encore implanter les décalages fréquentiels causés par les effets de masque et réaliser un module d'extraction des fréquences fondamentales virtuelle nominale (qui ne tient pas compte des décalages fréquentiels) et virtuelle réelle (qui en tient compte) de manière à réaliser la totalité de l'algorithme de Terhardt.

Cette partie n'est pas réalisée par iana~. Nous l'implanterons peut-être s'il existe un intérêt suffisant de la part des compositeurs. Il semble en effet que ces données ne soient généralement pas utilisées. Leur calcul alourdirait alors inutilement le module iana~. N'hésitez pas à faire part de vos remarques et suggestions en envoyant un message à artem@skynet.be. Nous vous tiendrons alors au courant de toute nouvelle version.

Références

[Harris, 76] Harris Frederic J. (1976) On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform, *Proceedings of IEEE, Vol. 66, No. 1, Janvier 1978*, pp. 51-83.

[Risset, 71] Risset J. C. (1971) Paradoxe de hauteur : le concept de hauteur sonore n'est pas le même pour tout le monde, *7th ICA, Budapest 1971*, pp. 613-616.

[Risset, 76] Risset J. C. (1976) Hauteur et timbre des sons, *Festival International du Son Haute Fidélité Stéréophonique, 1976*, pp. 141-153.

[Risset, 81] Risset J. C. (1981) Fréquence et perception de la musique, *Festival International du Son Haute Fidélité Stéréophonique (Conférences des journées d'études), 1981*, pp. 23-42.

[Terhardt, 73] Terhardt Ernst (1973) Pitch, consonance, and harmony, *JASA 55(5), May 1974*, pp. 1061-1069.

[Terhardt, Stoll & Seewann, 82a] Terhardt Ernst, Stoll Gerhard & Seewann Manfred (1982) Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions, *JASA 71 (3), March 1982*, pp. 671-678.

[Terhardt, Stoll & Seewann, 82b] Terhardt Ernst, Stoll Gerhard & Seewann Manfred (1982) Algorithm for extraction of pitch and pitch saliency from complex tonal signals, *JASA 71 (3), March 1982*, pp. 679-688.

[Todoroff, Daubresse & Fineberg, 95] Todoroff, T., Daubresse Eric & Fineberg Joshua (1995) "Iana~: a real-time environment for analysis and extraction of frequency components of complex orchestral sounds and its application within a musical realization", *Proc. ICMC 1995, Banff*, pp. 292-293.

[Zwicker & Terhardt, 80] Zwicker E. & Terhardt Ernst (1980) Analytical expressions for critical-band rate and critical bandwidth as function of frequency, *JASA 68(5), November 1980*, pp. 1523-1525.

[Zwicker, 80] Zwicker E. (1980) Formulae for calculating the psychoacoustical excitation level of aural difference tones measured by the cancellation method, *JASA 69(5), May 1981*, pp. 1410-1413.