



**HAL**  
open science

# Préférence, évaluation subjective et verbalisation des différences entre les fragments musicaux enregistrés en **WAVE** et **MP3**

Valery N. Nosulenko, Irina V. Starikova

► **To cite this version:**

Valery N. Nosulenko, Irina V. Starikova. Préférence, évaluation subjective et verbalisation des différences entre les fragments musicaux enregistrés en WAVE et MP3. 10ème Congrès Français d'Acoustique, Apr 2010, Lyon, France. hal-00538351

**HAL Id: hal-00538351**

**<https://hal.science/hal-00538351v1>**

Submitted on 22 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

## Préférence, évaluation subjective et verbalisation des différences entre les fragments musicaux enregistrés en WAVE et MP3

Valery Nosulenko<sup>1</sup>, Irina Starikova<sup>2</sup>

<sup>1</sup>Institut de psychologie, Académie des sciences de Russie, valery.nosulenko@gmail.com

<sup>2</sup>Centre de psychologie expérimentale, Université psychologique et pédagogique de la Ville de Moscou, irina.starikova@psyexp.ru

Les différences perçues entre les sons numériques enregistrés en formats WAVE (qualité CD) et MP3 (64 kbps) ont été analysées. Neuf fragments musicaux ont été choisis pour le traitement numérique : enregistrement d'instruments naturels, orchestre avec la voix humaine, instrument synthétisés (3 fragments de chaque sorte). Trois séries d'expériences avec la participation de 35 auditeurs ont été réalisées. Chaque fragment musical enregistré en deux formats a été présenté aux auditeurs par paire (90 présentations par casque au total). La première tâche pour les auditeurs était d'évaluer la différence entre les sons de deux formats selon l'échelle 0-8. Une autre tâche était de choisir dans la paire présentée le son « le plus plaisant » (la première série d'expérience) ou « le plus naturel » (deuxième série). Dans la troisième série d'expériences, les auditeurs devaient décrire à voix haute les particularités des différences entre les sons présentés par paires et aussi expliquer leur choix de préférence. En général, les auditeurs préféraient plus souvent les sons WAVE par rapport au format comprimé MP3. Cette tendance est significativement plus marquée chez les auditeurs ayant une formation musicale. Mais la préférence ainsi que l'évaluation subjective des différences entre les sons comparés dépend considérablement du type de fragment musical. Les auditeurs ne perçoivent quasiment pas de différences entre les deux formats pour les enregistrements d'instruments synthétisés et de piano (l'un des instruments naturels). L'analyse des données a permis également de distinguer un groupe de 7 auditeurs dont les choix dépendaient de la tâche demandée : le son WAVE leur apparaissait comme « le plus plaisant » alors que le format MP3 leur semblait « le plus naturel ». L'analyse des verbalisations a permis de préciser les critères utilisés par les auditeurs pour leurs choix.

## 1 Introduction

L'objectif de cette recherche expérimentale était d'analyser les perceptions et les évaluations des sons numériques largement répandus dans la vie quotidienne de l'homme.

Il s'agit de comparer les particularités des perceptions des sons issus de différents formats de compression. Les préférences des auditeurs, les critères de choix, ainsi que les évaluations subjectives de dissemblance entre les sons, enregistrés dans des formats différents, ont été analysés. L'objectif de cette analyse est de mettre en évidence les caractéristiques des comparaisons et des préférences de ces sons en fonction de leur type, de l'expérience de leur écoute et de la tâche pour le choix préférentiel.

La recherche était organisée dans le cadre du paradigme de « qualité perçue », dont l'idée consiste à définir un ensemble de caractéristiques perceptivement pertinentes de l'objet [1, 2]. Le contenu de cet ensemble est déterminé par l'expérience passée de l'individu, son contexte socioculturel, son éducation, etc.

Le paradigme de qualité perçue comprend une méthode d'analyse des verbalisations recueillies au cours des activités des individus. Cette méthode permet de traiter les verbalisations par rapport aux paramètres « objectifs » (mesurés) des objets perçus, et par rapport aux perceptions par le sujet de l'objet donné [1, 2, 3, 4, 5]. La tâche de comparaison imposée au sujet constitue l'une des conditions dans lesquelles les données verbales peuvent être considérées comme des indicateurs pertinents de l'activité perceptive [1]. C'est ce qui a déterminé la procédure

expérimentale : les auditeurs devaient comparer et verbaliser les sons présentés par paire.

Partant de la notion de qualité perçue, nous supposons que le contenu de la qualité perçue des sons de l'environnement acoustique numérisé est lié à l'expérience d'écoute des sons par l'individu, et à sa formation (notamment, musicale). Cette liaison doit se manifester, d'une part, dans le caractère du choix préférentiel réalisé parmi des sons soumis au traitement numérique ; et d'autre part, dans les évaluations subjectives de dissemblance entre les sons comparés. Au regard de la nouveauté grandissante des sons qui peuvent être créés au moyen des technologies numériques, l'intérêt de la recherche est d'analyser les choix préférentiels sur la base du paramètre « naturel – artificiel ».

## 2 Méthode

### 2.1 Stimuli

Neuf fragments musicaux ont été choisis pour le traitement numérique : enregistrement d'instruments naturels (n°1 – guitare, n°8 – orchestre symphonique, n°9 – piano) ; orchestre d'instruments naturels avec une voix humaine (n°2 – orchestre symphonique avec une voix féminine et n°4 – avec une voix masculine, n°3 – pop orchestre avec une voix masculine) ; enregistrement d'instruments synthétisés (n°5 – ensemble d'instruments synthétisés avec le groupe rythmique, n°6 et n°7 – sans le groupe rythmique). Chaque fragment musical, d'une durée allant de 8 à 10 secondes, a été enregistré en deux formats

de compression : WAVE (qualité CD) et MP3 (64 kbps). C'est ainsi que les paires de fragments musicaux ont été construites ; l'intervalle entre les deux sons aux formats différents étant de 0,5 secondes.

## 2.2 Procédure

La procédure de comparaison par paire a été utilisée dans trois séries d'expérience.

Dans la première série, il était demandé à l'auditeur de comparer les deux sons de la paire présentée et de choisir le son « le plus plaisant ».

Dans la deuxième série d'expériences, l'auditeur devait choisir dans la paire présentée le son qui lui semblait « le plus naturel ».

Dans la troisième série, on demandait aux auditeurs de décrire à voix haute les particularités des différences perçues entre les sons présentés par paire, mais aussi d'expliquer leur choix préférentiel.

De plus, dans chacune des trois séries d'expériences, les auditeurs évaluaient les dissemblances perçues entre les sons de la paire, sur une échelle de 0 à 7 (0 signifiant très similaire et 7 très - différent).

Le protocole expérimental des deux premières séries comprenait 90 présentations (10 paires pour chaque fragment musical : 5 dans l'ordre WAVE-MP3 et 5 dans l'ordre inverse). Dans la troisième série, le nombre de présentations de chaque paire a été diminué jusqu'à deux, ce qui faisait 18 présentations au total.

Les sons ont été présentés aux auditeurs à l'aide de l'ordinateur Vaio PCG-K415S (carte son Realtec AC97), par casque Sennheiser HD 280pro.

## 2.3 Auditeurs

Au total, 35 auditeurs (18 hommes et 17 femmes) ont participé aux tests. Leur répartition était réalisée sur la base des critères suivants : le niveau de formation musicale (9 « musiciens » et 26 « non-musiciens »), l'expérience d'écoute de la musique électroacoustique (moins de 15 heures par semaine – 17 auditeurs, plus de 15 heures par semaine – 18 auditeurs) et l'âge (inférieur ou égal à 22 ans – 19 auditeurs, plus de 22 ans – 16 auditeurs).

## 2.4 Analyse des données

Pour les données préférentielles, les fréquences relatives au choix de l'un des deux formats comparés ont été calculées (nombre de choix du format  $N_i$  par rapport au nombre total de présentations  $N$ ). Ce calcul a été réalisé séparément pour chacun des types de fragments musicaux. Ensuite, étaient calculées les moyennes pour l'ensemble des présentations.

En ce qui concerne les données de dissemblance, nous avons calculé les moyennes pour chaque fragment musical et pour chaque groupe d'auditeurs.

Pour l'analyse statistique, nous avons utilisé le critère non-paramétrique (Whitney Rank Sum Test).

L'analyse des verbalisations consistait essentiellement à créer des « portraits verbaux », permettant de comparer les valeurs de présence de différentes caractéristiques perceptives, en analysant les fréquences d'utilisation des unités verbales extraites des protocoles verbaux [1, 3, 4].

# 3 Résultats

## 3.1 Première série expérimentale : choisir le son « le plus plaisant »

En général, les auditeurs considéraient plus souvent les sons WAVE comme « plus plaisants » par rapport au format comprimé MP3. Cette tendance est valable pour la plupart des fragments musicaux, et notamment pour les enregistrements de l'orchestre symphonique (fragments n°2 et n°8). Notons des exceptions concernant trois fragments où la préférence n'est pas statistiquement significative : deux sons synthétisés (n°5 et n°7) et un enregistrement de piano (n°9).

La comparaison des données de différents auditeurs montre que cette tendance est significativement plus marquée chez les auditeurs ayant une formation musicale ( $p < 0,05$ ). Comme nous pouvons le constater sur la figure 1, le choix du son WAVE chez les « musiciens » est plus fréquent par rapport aux « non musiciens ». Respectivement, ils choisissent plus rarement les sons en MP3 et le nombre de leurs réponses neutres est également inférieur par rapport aux « non musiciens ». Seul le son synthétisé n°7 ne fait pas émerger de préférence stable.

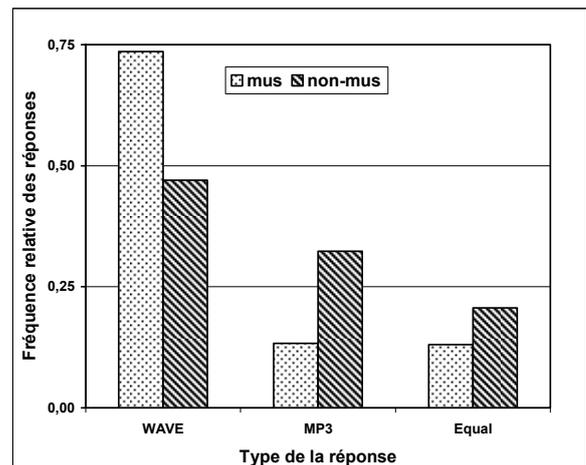


Figure 1. Le choix du son « le plus plaisant » par les auditeurs ayant une formation musicale (mus) et par les auditeurs sans formation musicale (non-mus).

Dans l'analyse des données de dissemblance entre les sons issus de différents formats de compression, on distingue deux fragments : le fragment n°7 (instrument synthétisé séparé) et le fragment n°9 (enregistrement du piano). La différence perçue entre WAVE et MP3 pour ces deux fragments est significativement plus faible que pour tous les autres fragments. Il est à noter qu'il s'agit de deux des fragments pour lesquels les préférences pour l'un ou l'autre des formats de compression ne se manifeste pas.

L'analyse des données des auditeurs regroupés selon leur expérience d'écoute n'a pas permis de trouver de différences, ni dans le choix préférentiel, ni dans les évaluations de dissemblance entre les enregistrements issus des deux formats de compression.

Pour le groupe des « jeunes » auditeurs (<22 ans), la moyenne des évaluations de dissemblance entre le son WAVE et le son MP3 est plus élevée ( $p < 0,001$ ) que pour le groupe des auditeurs plus âgés. Cependant, il n'y a pas de différences entre ces deux groupes d'auditeurs dans le caractère de leur préférence.

### 3.2 Deuxième série expérimentale : choisir le son « le plus naturel »

Les données obtenues dans la deuxième série montrent que le choix du son WAVE comme étant « le plus naturel » est moins fréquent que le choix le considérant comme étant « le plus plaisant » (dans la première série d'expériences). En comparaison avec la première série, la fréquence relative des choix du format MP3 augmente (de 0,27 à 0,35) ; respectivement, la part de réponses neutres « c'est pareil » diminue de 0,19 à 0,12. Par conséquent, dans la deuxième série, les indices de préférence sont devenus non significatifs pour différencier les sons issus de différents formats de compression. Pour nous, ce résultat confirme l'effet de changement de la tâche. Sinon, il aurait fallu attendre une stabilité plus élevée de réponses pour cause d'entraînement.

En ce qui concerne les évaluations de dissemblance, leurs valeurs et tendances sont quasiment les mêmes dans les données des deux séries d'expériences (coefficient de corrélation  $r^2=0,98$ ).

Une autre confirmation de l'effet de changement de la tâche se situe dans l'analyse plus détaillée des données individuelles des auditeurs. Cette analyse nous a permis de distinguer un groupe de 7 auditeurs dont le choix préférentiel a considérablement changé dans la deuxième série d'expériences. Si, dans la première série, ces auditeurs préféraient les sons enregistrés en WAVE (ils leur apparaissaient comme « plus plaisants »), dans la deuxième série, leur préférence portait plutôt sur le format MP3 qui leur semblait être « le plus naturel » (Figure 2).

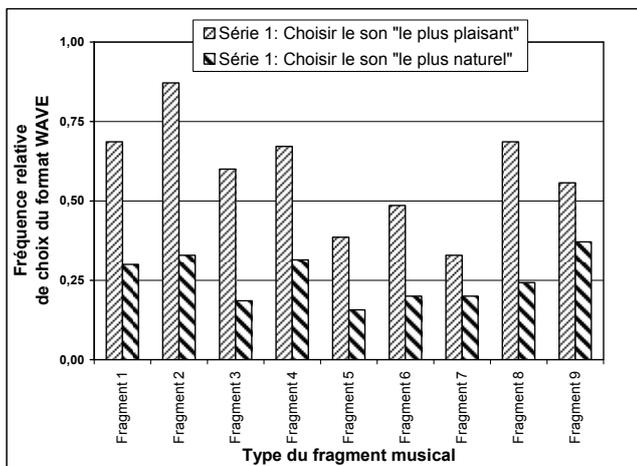


Figure 2. Comparaison des données de deux séries d'expériences pour le groupe de 7 auditeurs.

Pour ce groupe d'auditeurs, la différence entre les données des deux séries d'expériences est significative pour la plupart des fragments musicaux (pour les fragments n°1, 2, 3, 4 et 8  $p<0,0001$  ; pour le fragment n°5  $p=0,0195$  ; pour le fragment n°6  $p=0,0035$ ). Encore une fois, c'est seulement dans le cas des données obtenues avec les fragments n°7 et n°9, qu'on ne trouve pas de différences significatives avec le changement de la tâche de choix préférentiel.

Ainsi, nous pouvons constater que pour certains auditeurs, la situation de choix du son comme étant « le plus plaisant » et la situation de choix du son comme étant « le plus naturel » ne sont pas équivalentes. Il est important de noter que pour le groupe d'auditeurs qui participait aux expériences, le son « le plus naturel » est celui qui est le plus compressé, c'est-à-dire, le plus éloigné de

« l'original » par des moyens numériques, et donc le plus « artificiel ». Ici, on peut parler de déviation des « standards auditifs » pour cause d'écoute fréquente des sons compressés [6].

### 3.3 Troisième série expérimentale : décrire verbalement les particularités des différences entre les sons présentés

Les résultats d'analyse des verbalisations produites par les auditeurs nous ont permis de créer 18 groupes d'unités verbales pour différencier les sons selon leur « qualité perçue ». Dans l'ensemble de verbalisations, 75% sont intégrées dans 10 groupes de descripteurs (traduction du russe) : « naturel – artificiel », « agréable – désagréable », « net – pas net », « sonore – sourd », « haut – bas », « détaillé – non détaillé », « saturé – non saturé », « volumineux – pas volumineux », « perçant – pas perçant », « clair – pas clair ».

Néanmoins, cela ne signifie pas pour autant que 8 autres catégories ne sont pas importantes. Par exemple, la catégorie « fort – pas fort » est absente parmi les plus fréquentes dans la distribution générale. Elle devient par contre, la plus pertinente pour caractériser les fragments musicaux n°4, 5 et 7. De même, la catégorie « dynamique – pas dynamique » est la plus utilisée pour interpréter les spécificités du fragment n°6 (ensemble d'instruments synthétisés avec un groupe rythmique).

La distribution des descripteurs dépend également du format de compression du son présenté ainsi que de la « préférence » ou « non préférence » pour l'auditeur donné.

Ces résultats sont les plus marqués dans les portraits verbaux des sons comparés. Les portraits verbaux montrent non seulement la fréquence de présence d'une certaine catégorie verbale, mais également le « poids » de cette catégorie dans l'ensemble des verbalisations.

La Figure 3 donne l'exemple des portraits verbaux du même son mais dans deux situations de préférence.

Sur cette figure, on peut voir les portraits verbaux du fragment musical n°4 (voix masculine accompagnée de l'orchestre symphonique) enregistré au format WAVE. Il s'agit de deux situations : (1) la description des impressions caractérisant ce fragment quand il était perçu comme « plaisant » et (2) la description du même fragment dans les cas où la préférence portait sur le format MP3 (dont celui-ci était « déplaisant »).

On peut également constater sur la figure 3 que la composition de la « qualité perçue » est principalement différente pour ces deux situations. Les différences concernent le nombre de catégories utilisées pour la description ainsi que leur contenu.

Les auditeurs qui n'aiment pas le son de ce fragment en comparaison avec le son MP3 le décrivent comme « artificiel » et « désagréable ». Mais dans le cas de sa préférence, les auditeurs le considèrent « naturel » et « agréable ». En plus, un certain nombre de détails apparaissent : « fort », « riche », « détaillé », etc. Il est intéressant de noter que dans les deux cas, les auditeurs utilisent la catégorie « volumineux ». Mais dans le premier cas, cette catégorie entre dans l'ensemble des critères caractérisant la préférence pour le son enregistré au format WAVE. Tandis que dans le deuxième cas, il s'agit de l'ensemble des critères pour la non-préférence de ce format.

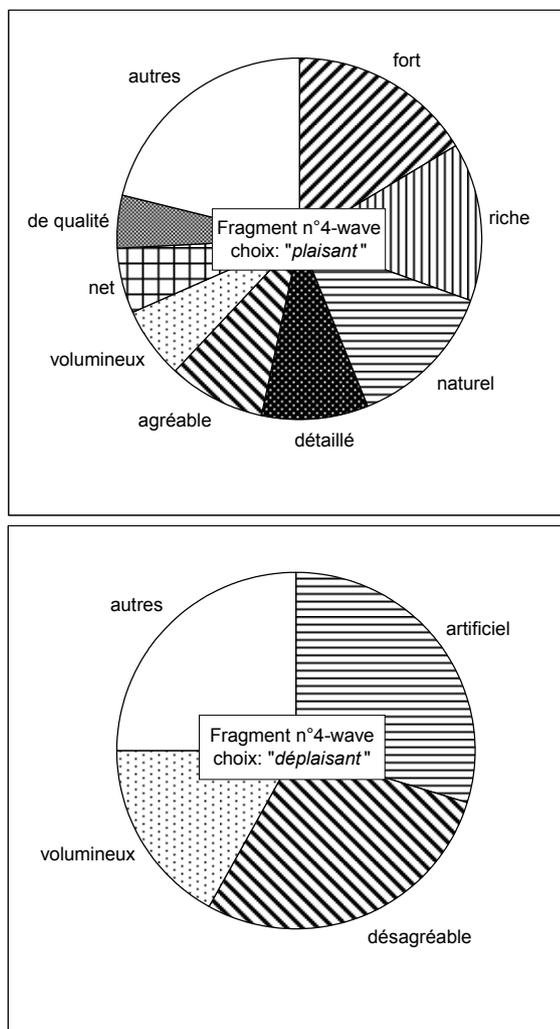


Figure 3. Portraits verbaux du fragment n°4 au format WAVE construits selon les descriptions produites dans deux situations différentes : choix de ce son comme « le plus plaisant » dans la paire (en haut) et choix du son au format MP3 comme « le plus plaisant » (« déplaisant » au format WAVE – en bas).

Ceci est un exemple de traitement des données verbales dans le cadre du paradigme de qualité perçue. La création du portrait verbal est une sorte de « mesure » des caractéristiques perceptives des événements comparés [1, 2]. Les résultats de cette analyse nous ont permis de préciser les critères que les auditeurs utilisaient dans leurs choix.

## 4 Conclusion

Les résultats obtenus permettent de conclure que la préférence de l'un des formats de compression, dans la perception des fragments musicaux, dépend du type de fragment présenté, du niveau de formation musicale de l'auditeur ainsi que de la tâche portant sur le choix préférentiel (« le plus plaisant » ou « le plus naturel »). Le type de fragment musical influence également l'évaluation subjective de dissemblance entre les sons enregistrés dans différents formats de compression. Cependant, notre supposition sur l'influence de l'expérience individuelle d'écoute n'a pas été confirmée.

Les enregistrements de grands ensembles d'instruments « naturels » (orchestre symphonique) permettent de percevoir les différences les plus marquées entre les sons

issus de deux formats. Par contre, dans les sons des instruments synthétisés, ces différences ne sont pas perçues.

Les auditeurs ayant une formation musicale sont plus sensibles, que les auditeurs sans formation musicale, à la différenciation des formats de compression du son. Pour ces auditeurs, les sons enregistrés au format WAVE sont toujours préférés par rapport au format MP3.

Cependant, les particularités de la perception, qui sont liées à l'interprétation par les auditeurs de la tâche portant sur la préférence, ont été constatées. Il s'agit notamment des situations de choix préférentiels selon « la plaisance » ou la « naturalité » du son. Pour la majorité des auditeurs, le choix du format de compression était toujours en faveur du format WAVE. Mais certains auditeurs différencient principalement ces deux situations : le son WAVE leur apparaissant comme « le plus plaisant » alors que le format MP3 leur semble « le plus naturel » et donc préférentiel. Ceci peut signifier que pour ces auditeurs, l'enregistrement MP3 (64 kbps !) est plutôt « naturel » dans leur écoute quotidienne et leur standards auditifs sont formés sur cette base.

## Remerciements

Cette recherche a été subventionnée par le projet « La perception de l'environnement acoustique » (Fondation Maison des Sciences de l'Homme et Fondation russe pour les sciences humaines, Grant n° 08-06-95072a).

## Références

- [1] Nosulenko V., Samoylenko E. "Approche systémique de l'analyse des verbalisations dans le cadre de l'étude des processus perceptifs et cognitifs", *Informations sur les Sciences Sociales*, 36(2), 223-261 (1997).
- [2] Nosulenko V. "Mesurer les activités numérisées par leur qualité perçue", *Informations sur les Sciences Sociales*, 47(3), 391-417 (2008).
- [3] Nosulenko V., Samoylenko E. "Evaluation de la qualité perçue des produits et services: approche interdisciplinaire", *International Journal of Design and Innovation Research*, vol. 2, 35-60 (2001).
- [4] Nosulenko V., Parizet E., & Samoylenko E. "La méthode d'analyse des verbalisations libres: une application à la caractérisation des bruits de véhicules", *Informations sur les Sciences Sociales*, 37(4), 593-611 (1998).
- [5] Samoylenko E., McAdams S., & Nosulenko V. "Systematic analysis of verbalizations produced in comparing musical timbres", *International Journal of Psychology*, 31(6), 255-278 (1996).
- [6] Nosulenko V. "Psychological Peculiarities and Acoustical Environment Changes", *International Journal of Psychology*, (26)5, 623-632 (1991).