

Sujet de projet : Séparation informée de la voix chantée

1 Introduction

Le problème séparer les différentes sources qui composent un signal de musique dans le cas où l'on dispose uniquement d'un signal monophonique est un sujet de recherche qui suscite un grand intérêt depuis plusieurs années. Ce problème difficile a de nombreuses applications telles que la transcription automatique des morceaux de musique [1], la reconnaissance d'instruments [2], ou l'estimation de la mélodie [3].

De nombreux éléments contribuent à faire de la séparation de sources un problème difficile. On peut citer parmi eux la complexité des sources musicales qui interfèrent dans les mélanges polyphoniques, les conditions variables d'enregistrement, le bruit, les effets digitaux appliqués pour les enregistrements en studio, etc. Dans ce projet, on s'intéresse en particulier au problème de séparer la voix chantée de l'accompagnement d'un morceau de musique.

La voix chantée est un attribut complexe et fondamental des signaux de musique qui a attiré l'attention de nombreux chercheurs dans la communauté *Music Information Retrieval* (MIR). La séparation de la voix est une étape essentielle pour plusieurs applications telles que l'identification de chanteur [4], la transcription de la mélodie [5], la recherche par fredonnement [6]. On trouvera dans [7] une revue récente des méthodes pour la séparation de la voix chantée.

Récemment, des approches qui tirent parti de la répétition du signal ont vu le jour. Ces approches supposent que l'accompagnement de fond a une structure musicale répétitive, à la différence du signal vocal dont la répétition, le cas échéant, a lieu à une échelle de temps beaucoup plus grande [8, 9, 10]. L'article [8] propose une méthode simple pour séparer la voix de l'accompagnement qui est basée sur l'extraction de la structure musicale répétitive sous-jacente à l'aide de masquage temps-fréquence binaire (algorithme REPET). Cette méthode fait l'hypothèse qu'il n'y a pas de variation significative dans l'accompagnement et est donc limitée à de courts extraits de musique. La méthode est généralisée dans [9] pour permettre le traitement de morceaux de musique entiers, en s'appuyant sur une hypothèse de périodicité spectrale locale.

2 Travail demandé

Le but de ce projet est d'implémenter la méthode REPET proposée dans [8]. L'idée de base est de trouver les motifs qui se répètent (repeating pattern) dans le mélange de sources *voix + accompagnement* (mixture), d'en tirer les modèles répétitifs sous-jacents, et d'extraire l'accompagnement (répétitif) en comparant les modèles au mélange.

La méthode se décompose en 3 étapes principales, ainsi que décrit sur la figure 1.

- **Étape 1** Calcul du spectrogramme de beat b et estimation de la période de répétition p .
- **Étape 2** Segmentation du spectrogramme de mélange V et calcul du segment répétitif S .
- **Étape 3** Calcul du spectrogramme répétitif W et dérivation du masque temps-fréquence M .

3 Morceaux de test

On pourra tester l'algorithme sur les morceaux donnés avec l'énoncé (fichiers voix séparée / accompagnement / mélange). On trouvera une base de données de morceaux plus complète à l'adresse suivante : <https://sites.google.com/site/unvoicedsoundseparation/mir-1k>

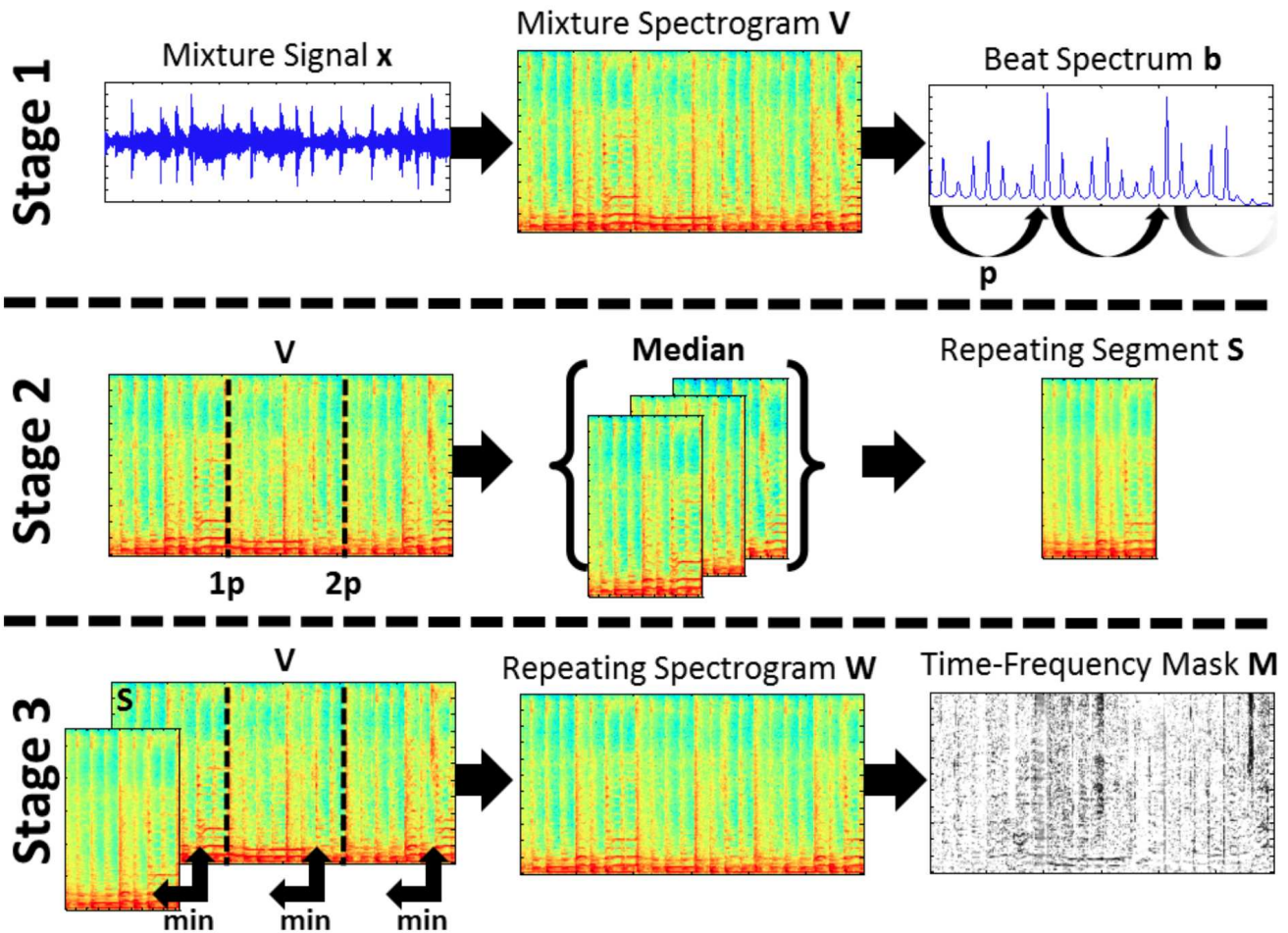


Figure 1: Vue d'ensemble de la méthode REPET.

References

- [1] Plumbley, M.D. and Abdallah, S.A. and Bello, J.P. and Davies, M.E. and Monti, G. and Sandler, M.B. “Automatic music transcription and audio source separation”. In *Cybernetics and Systems*, 33(6), pp 603–627, (2002).
- [2] Heittola, T. and Klapuri, A.P. and Virtanen, T. “Musical instrument recognition in polyphonic audio using source-filter model for sound separation”. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp 327–332, Kobe, Japan, 2009.
- [3] Durrieu, J.L. and Richard, G. and David, B. and F evotte, C. “Source/filter model for unsupervised main melody extraction from polyphonic audio signals”. *IEEE Trans. Audio, Speech, Language Process.*, 18(3):564–575, 2010.
- [4] W. Cai, Q. Li, and X. Guan. “Automatic singer identification based on auditory features”. 2011.
- [5] J. Salamon, E. G omez, D.P.W. Ellis, and G. Richard. “Melody extraction from polyphonic music signals: Approaches, applications and challenges”. In *IEEE Signal Process. Mag.*, 2013.
- [6] R.B. Dannenberg, W.P. Birmingham, B. Pardo, N. Hu, C. Meek, and G. Tzanetakis. “A comparative evaluation of search techniques for query-by-humming using the musart testbed”. In *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 5, pp. 687–701, 2007.
- [7] B. Zhu, W. Li, R. Li, and X. Xue, “Multi-stage non-negative matrix factorization for monaural singing voice separation”. In *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 10, pp. 2096–2107, 2013.

- [8] Z. Rafii and B. Pardo. “A simple music/voice separation method based on the extraction of the repeating musical structure”. In ICASSP, 2011.
- [9] A. Liutkus, Z. Rafii, R. Badeau, B. Pardo, and G. Richard. “Adaptive filtering for music/voice separation exploiting the repeating musical structure”. In ICASSP, 2012.
- [10] D. FitzGerald. “Vocal separation using nearest neighbours and median filtering”; In ISSC, 2012.