Mixage automatique

Adrien Llave
novembre 2023

Sommaire

Objectifs

- Prendre connaissance des dernières évolutions de la recherche
- Esquisser les applications en développement
- Qu'est-ce qu'une question scientifique ?

D'où je parle

- 2010 - 2012 : IUT GEII
- 2012 - 2015 : Master métiers du son à l'ENS Louis-Lumière
- 2015 - 2016 : Master traitement du signal à Grenoble-INP Phelma
- 2017 - 2022 : Doctorat traitement du signal à CentraleSupélec
- 2022 - 2023 : Chercheur à Orange
- 2023 : Enseignement à l'IUT, recherche au LTSI
- 2024 - : Chercheur à Orange

Les objectifs du mixage

Les objectifs du mixage

- équilibre entre les sources
- démasquer les sources
- maitriser la dynamique (court et long terme)
- composer une scène sonore
- transmettre de l'émotion / soutenir le propos de la musique

voir [Owsinski, 2013]

Disclaimer : charge de valeur (vertu)

Dans la review de [De Man et al., 2017b] :

"Innovation has traditionally been met with resistance and scepticism, in particular from professional users who **fear** seeing their roles disrupted or made **obsolete**. Music production technology may be especially susceptible to this kind of opposition, as it is characterised by a tendency towards **nostalgia**, skeuomorphisms and analogue workflows, and it is concerned with **aesthetic value** in addition to technical excellence and efficiency."

"These advancements have changed the nature of the sound engineering profession from primarily **technical** to increasingly **expressive**."

"There is economic, technological and artistic **merit** in exploiting the immense computing power and flexibility that today's digital technology affords, to venture away from the rigid structure of the traditional music production toolset."

Exemple du plug-in Gear Compressor de Redstair

Objectifs

Objectifs

- Mixer des enregistrements pour des musicien.nes fauché.es
- Mixer de concert en bar
- Mixer du jeu vidéo [Schmidt, 2003]
- Effectuer les tâches ingrates pour soulager les ingés son

Frontière flou entre outils d'assistance au mixage et mixage automatique total

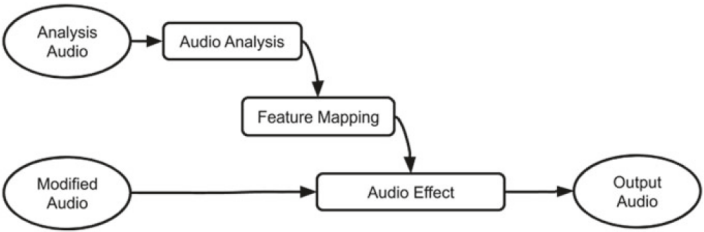
Les outils du mixage

- gain
- panning
- égalisation (EQ)
- compression (de dynamique)
- correction de délai
- reverb
- distortion (compliqué !)

Mixage automatique : historique de la recherche

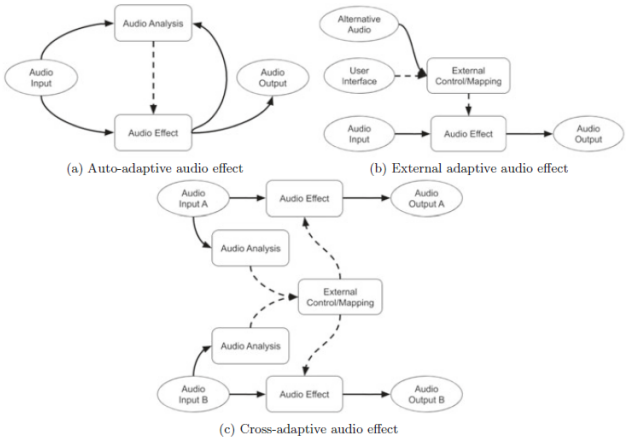
- historique :
  - origine 2007 - 2010 : gain, pan, EQ
  - depuis 2012 : compression
  - depuis 2016 : arrivée de l'apprentissage profond (deep learning)
  - depuis 2016 : reverb automatique
- voir [De Man et al., 2017b] pour une revue entre 2007 et 2017
- voir [Miranda, 2021] pour une revue moins détaillée mais plus récente

Principe général



A faire en boucle jusqu'à "convergence"

Différents niveaux de complexité



(a) Auto-adaptive audio effect

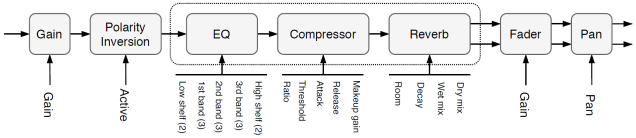
(b) External adaptive audio effect

(c) Cross-adaptive audio effect

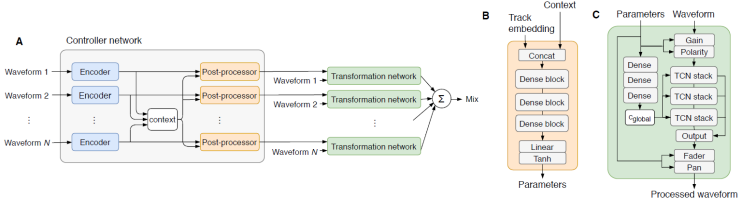
Revue détaillée des *cross-adaptive effects* [Reiss and Brandtsegg, 2018]

Traitement direct ou via effets communs ?

- Problème de l'ordre des traitements par exemple
- Transformation directe du signal
  - plus souple (ordre des traitements)
  - plus créative

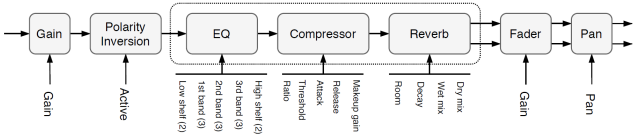


Chaine de traitement classique

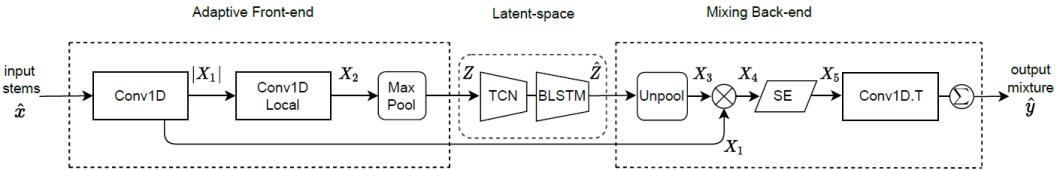


Réseau de neurones de [Steinmetz et al., 2020]

Traitement direct ou via effets communs ?



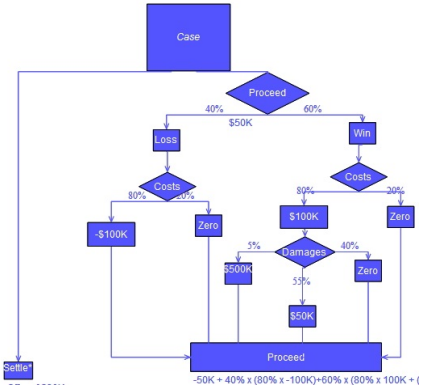
Chaine de traitements classique



Réseau de neurones de [Martínez-Ramírez et al., 2022]

Principe général

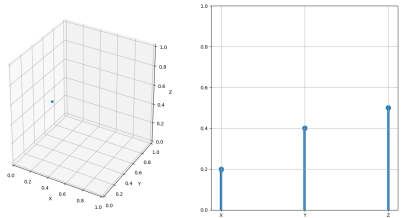
- système basé sur les méthodes humaines
- système basé sur des critères objectifs
- modélisation du processus de décision humain
- cascade de règles *si-alors-sinon*



Arbre de décision

Notion d'espace d'optimisation

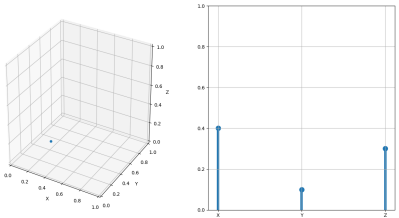
Imaginez, un espace...



Un vecteur/point de l'espace = un signal

Notion d'espace d'optimisation

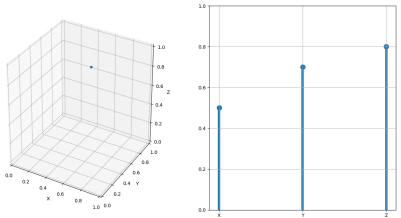
Imaginez, un espace...



Un vecteur/point de l'espace = un signal

Notion d'espace d'optimisation

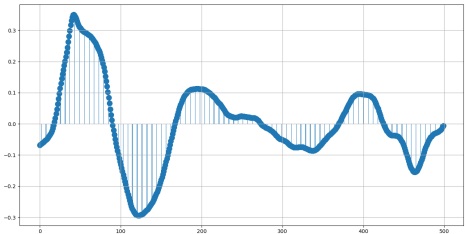
Imaginez, un espace...



Un vecteur/point de l'espace = un signal

Notion d'espace d'optimisation

Imaginez, un espace à N dimensions...



- notion importante : le produit scalaire
  - ressemblance entre deux vecteurs (signaux)
- "Malédiction des grandes dimensions" :
  - 44.1 kHz $\times$ 60 sec. $\times$ stereo = 5,3 millions de dim. !

Produit scalaire

Soit 2 vecteurs $\boldsymbol{x}$ et $\boldsymbol{y}$ en 3D :

$$\boldsymbol{x} = [x_1, x_2, x_3]$$
$$\boldsymbol{y} = [y_1, y_2, y_3]$$

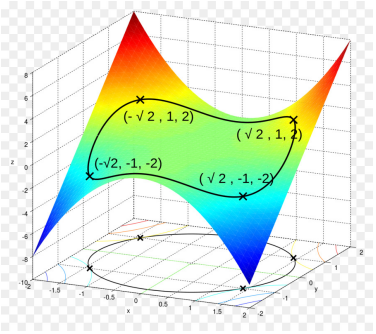
Le prod. scalaire entre $\boldsymbol{x}$ et $\boldsymbol{y}$:

$$< \boldsymbol{x}, \boldsymbol{y} > = x_1.y_1 + x_2.y_2 + x_3.y_3$$
$$= \sum_{i=1}^{3} x_i.y_i$$
$$= ||\boldsymbol{x}||\ ||\boldsymbol{y}||\ \cos(\widehat{\boldsymbol{x}, \boldsymbol{y}})$$

Quelques bases d'optimisation

problème d'optimisation sous contraintes :

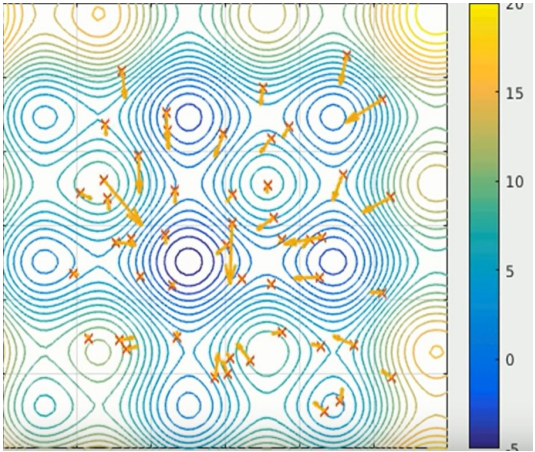
- des buts sont définis
- des contraintes doivent être respectées
- le système doit trouver une solution optimale (au sens d'un critère)



Exemple d'optimisation sous contraintes

Problème sans solution optimale mais plutôt différents mix. appropriés selon le contexte (esthétique, technique) [Jillings and Stables, 2017]

Compromis exploration/exploitation



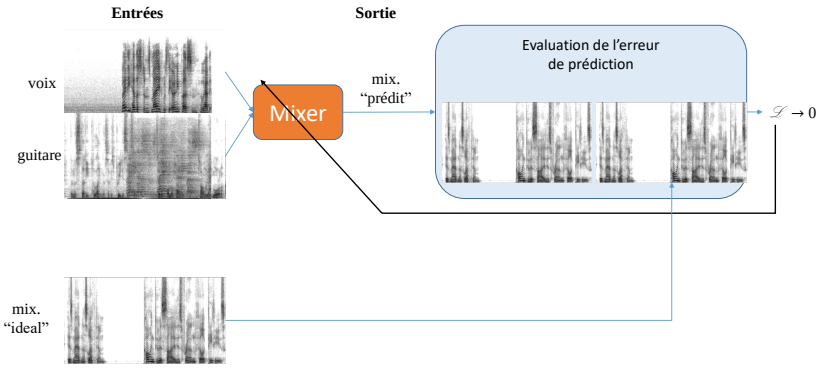
Exemple d'optimisation par essaims particulaires

voir : `https://youtu.be/8xycqWWqz50?t=67`

Différentes méthodes d'optimisation

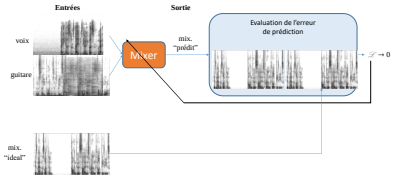
- approche des moindres carrés
  - puissance de l'erreur
- algorithme génétique [Jillings and Stables, 2017]
  - modélisation de la sélection naturelle
- optimisation par essaims particulaires (modélisation de vol d'oiseau)

Systèmes basés sur les données : principe général (i)



Systèmes basés sur les données : principe général (ii)

- Algorithme : séquence d'instructions déterministe
- Exemple : recette de cuisine
  - mélanger 100 g. de sucre et 3 oeufs
  - étaler, cuire...
- Avec apprentissage : recette à trous
  - mélanger tant d'ingrédient 1 et tant d'ingrédient 2
  - étaler, cuire...
- Processus d'apprentissage :
  - essai-erreur
  - feedback : chaud/froid



Perceptron (i)

Modèle de neurone de grenouille (Rosenblatt, 1958) :

$$y = f(w_0 x_0 + ... + w_{n-1} x_{n-1} + b)$$

$$y = f\left(\sum_{i=0}^{n-1} w_i x_i + b\right)$$

analogie : "séparer" l'espace des "x" en deux

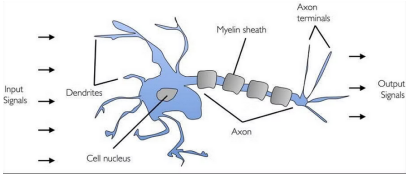
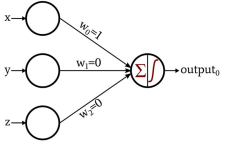


Illustration d'un neurone



Modèle simpliste de neurone

Perceptron - (ii)

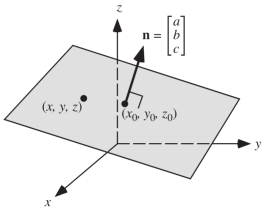
Modèle de neurone de grenouille (Rosenblatt, 1958) :

$$y = f(w_0 x_0 + ... + w_{n-1} x_{n-1} + b)$$

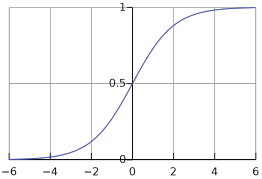
$$y = f\left(\sum_{i=0}^{n-1} w_i x_i + b\right)$$

analogie : "séparer" l'espace des x_i en deux

voir `https://deeperplayground.org`



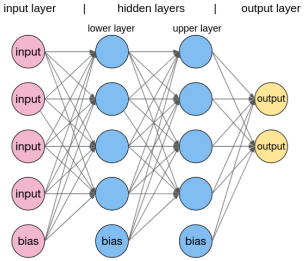
Plan défini par le vecteur n



Fonction d'activation sigmoïde

Architectures de réseau de neurones - perceptron multi-couches

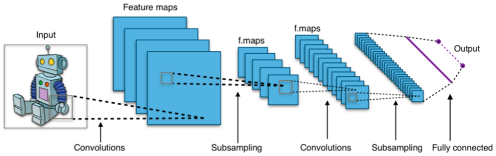
- empiler les perceptrons
  - très lourd
  - l'apprentissage ne converge pas !
- suffisant en théorie (théorème d'approximation universelle)
- il faut ajouter des *a priori* dans le système
  - exemples : dépendance dans le temps, structure harmonique...



Perceptron multi-couches

Architectures de réseau de neurones - convolutionnel

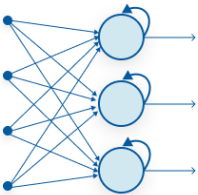
- Réseau convolutionnel [Gu et al., 2018]
  - léger
  - permet de considérer le contexte "local" (ex : temps/fréquence)
- hypothèse : invariance dans le temps et les fréquences
- exemple : recherche de contours, structures...



Réseau convolutionnel

Architectures de réseau de neurones - récurrent

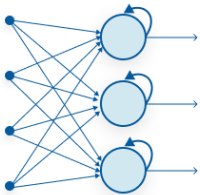
- réseau de neurones récurrents
  - permet de considérer le contexte à moyen terme
  - "plus l'info est loin dans le temps, moins elle est importante"
  - peu couteux
  - pas évident à entrainer



Réseau récurrent

Architectures de réseau de neurones - récurrent

- réseau de neurones récurrents
  - permet de considérer le contexte à moyen terme
  - "plus l'info est loin dans le temps, moins elle est importante"
  - peu couteux
  - pas évident à entrainer
- alternativement : modèle d'"attention" [Lin et al., 2022]
  - permet de piocher de l'information à long terme
  - voir `https://www.youtube.com/watch?v=CsQNF9s78Nc`



Réseau récurrent

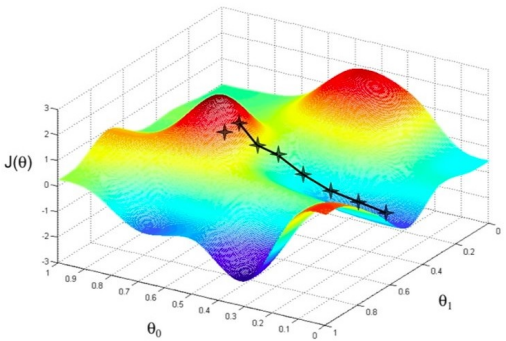
Procédure d'apprentissage

On cherche la combinaison des paramètres $\theta_0, \theta_1, ...$ qui minimise la mesure d'erreur $J(\theta_0, \theta_1, ...)$

Algorithme de la "descente de gradient"

- suivre la pente descendante de la métrique d'erreur
- (besoin d'un réseau de neurone "dérivable")
- (on rétropropage le gradient de l'erreur pour actualiser les θ_i)

En pratique, on a $\sim$10 millions de θ_i !



Descente de gradient

Procédure d'apprentissage

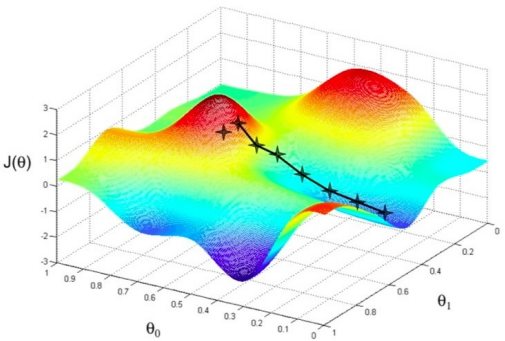
On cherche la combinaison des paramètres $\theta_0, \theta_1, ...$ qui minimise la mesure d'erreur $J(\theta_0, \theta_1, ...)$

Algorithme de la "descente de gradient"

- suivre la pente descendante de la métrique d'erreur
- (besoin d'un réseau de neurone "dérivable")
- (on rétropropage le gradient de l'erreur pour actualiser les θ_i)

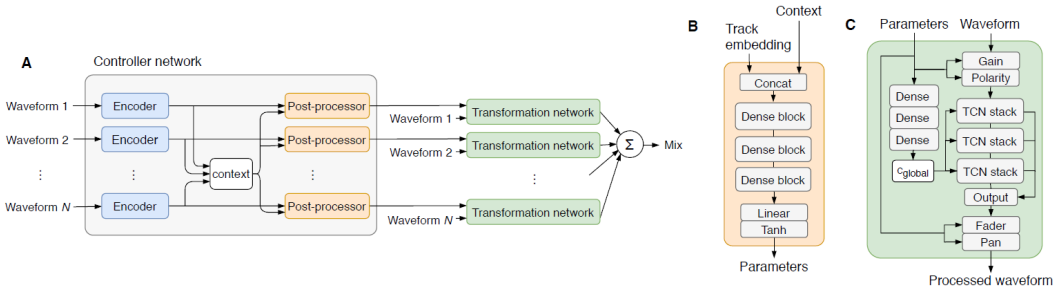
En pratique, on a $\sim$10 millions de θ_i !
Challenges :

- trouver une mesure d'erreur $J(\theta_0, \theta_1, ...)$ pertinente ET "dérivable"
- avoir beaucoup de données



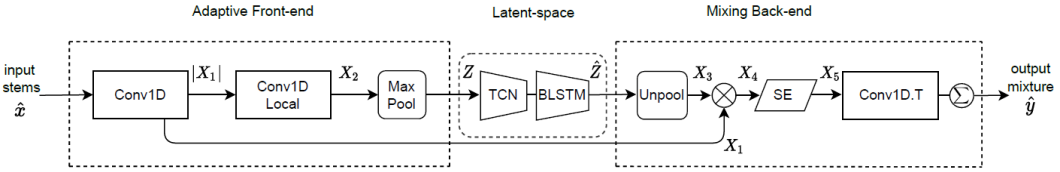
Descente de gradient

Console de mixage "différentiable"



Architecture de DNN de mixage automatique de [Steinmetz et al., 2020]

écouter `https://csteinmetz1.github.io/dmc-icassp2021/`



Architecture de DNN de mixage automatique de [Martínez-Ramírez et al., 2022]

écouter https://marco-martinez-sony.github.io/FxNorm-automix/AUDIO_SAMPLES.html

Base de données

- peu de données de session multi-pistes et mixage pro associé
  - ENST-Drums : 3h de batterie (8 pistes)
  - MedleyDB : 196 chansons, $\sim$7h [Bittner et al., 2014]
  - CoSoD : 331 chansons [Duguay et al., 2023]
- réemploi d'autres base de données (séparation de sources musicales)
  - MUSDB18 (voir [Martínez-Ramírez et al., 2022])
- représentation inégales des instruments [Bittner et al., 2014]
  - voix plus présente que la clarinette basse

Différents niveaux d'interactions

Différentes applications[a] :

- aide à la décision/préprocessing vs. full automatic (personne devant la console)
- amateur vs. pro

a. enquête socio des usages [Vanka et al., 2023]

Différents niveaux d'interaction :

- Automatique
- Indépendant
- Recommandation
- Découverte (exploration)

Automatique

- production amateur
- petit concert
- mix. étalon
- jeu vidéo [Schmidt, 2003]

Indépendant

Supervisé par un.e opérateurice

- réduction de diaphonie (repisse micro)
- synchronisation entre micro (couple + appoints)
- Démasquage fréquentiel [Wichern et al., 2015]
- pour amateur et pro.

Izotope Neutron, démasque auto.

Analogue aux manœuvres automatiques pour la voiture autonome

Recommandation

Suggestion de traitements

- annotation automatique d'instrument
- annotation automatique de structure de morceau [Hargreaves et al., 2012]
- suggestion de chaine d'effets [Sauer et al., 2013]
- suggestion de paramètres d'effets
- organisation des *stems*

Izotope Neutron, tone matching

Descripteurs vers paramètres d'effets audio (voir "prompt to picture")
Avantage : interprétable et adaptable

Découverte (exploration)

Outils d'analyse du mix

- visualisation de mix. [Ford et al., 2015]
- comparaison avec un mix. de référence
- description qualitative (textuelle) du mix.
- description quantitative (numérique) du mix.
- analyse du niveau de reverb. et de son impact sur chaque piste dans le mix. [De Man et al., 2017a]
- analyse de mix. (graphe d'effets) [Lee et al., 2023]

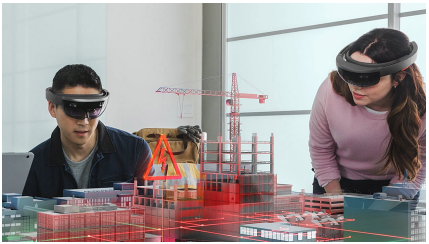
Résumé

- 15 ans de recherche (depuis 2007)
- essor du deep learning depuis 5 ans (depuis 2016)
- un problème fondamentalement mal posé ?
  - pas de "vérité terrain" [Birtchnell and Elliott, 2018]

"42" - H2G2

Perspectives

- aller vers le mix. 5.1, orienté-objet...
- considérer différentes esthétiques de mix.
- vers une meilleure connaissance de l'essence de la musique
- réduire les tâches laborieuses en montage
  - suggestion de sonothèque
  - bruitage de pas, de "présence"
  - nettoyage parole (bruit de bouche, etc.)
- recommandation d'arrangement musical
  - instrument complémentaire pour enrichir le timbre
- jeu vidéo et réalité augmentée
  - mixage ajouté au son local !



Bibliographie I

Birtchnell, T. and Elliott, A. (2018).
Automating the black art : Creative places for artificial intelligence in audio mastering.
Geoforum, 96 :77–86.

Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. (2014).
MedleyDB : A Multitrack Dataset for Annotation-Intensive MIR Research.
In *15th International Society for Music Information Retrieval Conference*, page 6.

De Man, B., McNally, K., and Reiss, J. D. (2017a).
Perceptual evaluation and analysis of reverberation in multitrack music production.
Journal of the Audio Engineering Society, 65(1) :108–116.

De Man, B., Reiss, J. D., and Stables, R. (2017b).
Ten Years of Automatic Mixing.
In *Proceedings of the 3rd Workshop on Intelligent Music Production*, page 5, Salford, UK.

Duguay, M., Mancey, K., and Devaney, J. (2023).
Collaborative Song Dataset (CoSoD) : An annotated dataset of multi-artist collaborations in popular music.
arXiv :2307.05588 [cs, eess].

Ford, J., Cartwright, M., and Pardo, B. (2015).
Mixviz : A tool to visualize masking in audio mixes.
In *Audio Engineering Society Convention 139*, volume 139. Audio Engineering Society.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G., Cai, J., and Chen, T. (2018).
Recent Advances in Convolutional Neural Networks.
Pattern recognition, 77 :354–377.
arXiv :1512.07108 [cs].

Bibliographie II

Hargreaves, S., Klapuri, A., and Sandler, M. (2012).

Structural segmentation of multitrack audio.
IEEE Transactions on Audio, Speech, and Language Processing, 20(10) :2637–2647.

Jillings, N. and Stables, R. (2017).

Automatic masking reduction in balance mixes using evolutionary computing.
In *Audio Engineering Society Convention 143*. Audio Engineering Society.

Lee, S., Park, J., Paik, S., and Lee, K. (2023).

Blind Estimation of Audio Processing Graph.
In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, Rhodes Island, Greece. IEEE.

Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022).

A survey of transformers.
AI Open, 3 :111–132.

Martínez-Ramírez, M. A., Liao, W.-H., Fabbro, G., Uhlich, S., Nagashima, C., and Mitsufuji, Y. (2022).

Automatic music mixing with deep learning and out-of-domain data.
In *23rd International Society for Music Information Retrieval Conference (ISMIR)*,. arXiv.
arXiv :2208.11428 [cs, eess].

Miranda, E. R., editor (2021).

Handbook of Artificial Intelligence for Music : Foundations, Advanced Approaches, and Developments for Creativity.
Springer International Publishing, Cham, springer edition.

Owsinski, B. (2013).

The mixing engineer's handbook.
Nelson Education.

Bibliographie III

Reiss, J. D. and Brandtsegg, y. (2018).

Applications of Cross-Adaptive Audio Effects : Automatic Mixing, Live Performance and Everything in Between.

Frontiers in Digital Humanities, 5 :17.

Sauer, C., Roth-Berghofer, T., Auricchio, N., and Proctor, S. (2013).

Recommending audio mixing workflows.

In *International Conference on Case-Based Reasoning*, pages 299–313. Springer.

Schmidt, B. (2003).

Interactive Mixing of Game Audio.

In *Audio Engineering Society Convention 115*, volume 115. Audio Engineering Society.

Steinmetz, C. J., Pons, J., Pascual, S., and Serrà, J. (2020).

Automatic multitrack mixing with a differentiable mixing console of neural audio effects.

arXiv :2010.10291 [cs, eess].

Vanka, S. S., Safi, M., Rolland, J.-B., and Fazekas, G. (2023).

Adoption of AI Technology in the Music Mixing Workflow : An Investigation.

arXiv :2304.03407 [cs, eess].

Wichern, G., Wishnick, A., Lukin, A., and Robertson, H. (2015).

Comparison of loudness features for automatic level adjustment in mixing.

In *Audio Engineering Society Convention 139*, volume 139. Audio Engineering Society.