

Desenvolupament d'una Plataforma per a l'Anàlisi i Generació de Partitures Musicals a partir d'Àudio Multitrack

Pol Colomer Campoy

10 de març de 2024

Resum– TODO: Aconseguir separar pistes d'àudio donat un àudio multitrack i obtenir les notes de cada pista per posteriorment generar la partitura de la mateixa. Estudiar els resultats obtinguts per diferents algorismes amb diferents configuracions i combinacions d'àudios.

Paraules clau– Paraules clau del treball, màxim 2 línies TODO Multitrack, .

Abstract– TODO: To separate audio tracks by giving a multitrack audio and obtain the notes of each track to later generate the score of it. Study the results obtained by different algorithms with different configurations and audio combinations.

Keywords– Versió en anglès de les paraules clau.

Resum– TODO: Aconseguir separar pistes d'àudio donat un àudio multitrack i obtenir les notes de cada pista per posteriorment generar la partitura de la mateixa. Estudiar els resultats obtinguts per diferents algorismes amb diferents configuracions i combinacions d'àudios.

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

EN els darrers anys, el camp de l'enginyeria informàtica ha experimentat un notable avanç en el tractament del senyal d'àudio i en l'anàlisi musical automatitzada. El meu projecte d'enginyeria informàtica, que implica la separació d'instruments, la generació de partitures i la creació de tutorials visuals, s'ubica en la intersecció de diverses àrees de recerca i desenvolupament. La meva principal motivació és adquirir pràctica, comprensió i realitzar una investigació en l'àmbit de l'àudio, ja que durant la meva carrera, és un dels temes que han quedat menys explorats. D'aquesta manera, pretenc enriquir la meva experiència com a estudiant en aquest àmbit.

- E-mail de contacte: Pol.ColomerC@autonoma.cat
- Menció realitzada: Computació
- Treball tutoritzat per: Felipe Lumbreras (departament - CVC)
- Curs 2023/2024

2 OBJECTIUS

L'objectiu principal d'aquest projecte és aconseguir la separació de pistes d'àudio d'un arxiu multitrack i obtenir les notes musicals de cada pista per a la posterior generació de la seva partitura. A continuació, es detallen els objectius ramificats en forma d'arbre:

1. Recopilar i preparar les dades necessàries per al desenvolupament del projecte, incloent-hi toy problems, arxius MIDI i àudios de concerts.
 - (a) Elaborar i preparar toy problems representatius dels casos d'ús.
 - (b) Recopilar arxius MIDI de cançons amb els seus respectius ground truths.
 - (c) Obtenir àudios multitrack de concerts amb soroll ambient del públic per a casos més desafiant.
2. Desenvolupar algorismes i tècniques per a la separació de pistes d'àudio tant amb tècniques clàssiques com actuals
 - (a) Investigar i seleccionar algorismes clàssics com Independent Component Analysis (ICA), Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF) i algorismes de filtratge per a la separació de pistes.

- (b) Explorar les possibilitats de les tècniques de Deep Learning com Xarxes Neuronals Convolucionals (CNNs), Xarxes Neuronals Recurrents (RNNs), U-Net, Wave-U-Net i Deep Clustering per millorar la separació de pistes.

3. Obtindre les notes musicals de cada pista separada:

- (a) Desenvolupar mètodes d'extracció de característiques i anàlisi musical per obtenir les notes musicals de cada pista separada.

4. Generar la partitura musical de cada pista:

- (a) Implementar algoritmes per a la generació de partitures utilitzant la informació de les notes musicals obtingudes.

3 TASQUES

Per assolir els objectius establerts, s'han definit les següents tasques, les quals estan dissenyades per abordar de manera efectiva els diversos aspectes del projecte:

1. Recopilació de dades i preparació: (5 setmanes en total paral·lel)

- (a) Investigar i seleccionar una varietat de toy problems representatius dels casos d'ús. (1 setmana)
- (b) Recopilar una col·lecció d'arxius MIDI de diverses cançons juntament amb les seves partitures corresponents. (1 setmana)
- (c) Obtindre àudios multitrack de concerts amb soroll ambient del públic per a casos més desafiant i variats. (1 setmana)
- (d) Pre-processar i netejar les dades recopilades per garantir la seva qualitat i coherència. (2 setmanes)

2. Desenvolupament d'algoritmes i tècniques per a la separació de pistes d'àudio:

- (a) Realitzar una revisió exhaustiva per identificar i comprendre els algoritmes clàssics i tècniques de Deep Learning més adequats per a la separació de pistes d'àudio. (2 setmanes)
- (b) Implementar una gamma d'algoritmes clàssics com ara Independent Component Analysis (ICA), Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF) i algorismes de filtratge per explorar la seva eficàcia en la separació de pistes. (3 setmanes)
- (c) Investigar i experimentar amb les tècniques de deep learning, incloent-hi Xarxes Neuronals Convolucionals (CNNs), Xarxes Neuronals Recurrents (RNNs), U-Net, Wave-U-Net i Deep Clustering, per millorar la qualitat i la precisió de la separació de pistes. (4 setmanes)
- (d) Avaluar i comparar el rendiment dels diferents algoritmes i tècniques implementades utilitzant mètriques rellevants com la relació senyal-soroll (SNR), la separació de font de bescanvi (SiSdr) i la puntuació de seguiment de referència (SAR). (2 setmanes)

3. Obtenció de les notes musicals de cada pista separada:

- (a) Investigar i desenvolupar mètodes d'extracció de característiques i anàlisi musical per obtenir les notes musicals de cada pista separada. (1 setmana)
- (b) Implementar algoritmes d'anàlisi de freqüències, reconeixement de patrons i sincronització temporal per a l'extracció precisa de les notes musicals. (3 setmanes)
- (c) Provar i ajustar els mètodes d'extracció de notes en diferents escenaris i tipus de pistes d'àudio. (1 setmana)

4. Generació de partitures:

- (a) Desenvolupar algoritmes i processos per a la generació automàtica de partitures a partir de les notes musicals obtingudes. (3 setmanes)
- (b) Implementar funcionalitats addicionals com la detecció de claus i la transcripció automàtica de les partitures generades. (2 setmanes)
- (c) Validar les partitures generades mitjançant comparació amb les partitures originals i avaluació manual de la seva qualitat musical. (1 setmana)

5. Documentació del projecte:

- (a) Preparar un Jupyter Notebook exhaustiu que descriu el procés complet del projecte, des de la recopilació de dades fins a la generació de partitures, incloent-hi explicacions detallades, codi font i visualitzacions. (2 setmanes)
- (b) Crear una documentació detallada que explica els passos realitzats, els resultats obtinguts, les conclusions del projecte i les possibilitats de desenvolupament futur. (3 setmanes)
- (c) Preparar una presentació o demostració del projecte per a l'avaluació final, incloent-hi la mostra dels resultats, les comparacions i les explicacions pertinents. (2 setmanes)
- (d) Documentar i elaborar els informes per a cada entrega (2 setmanes)

4 ESTAT DE L'ART

En aquesta secció, intentaré fer un resum dels treballs rellevants en el camp de la separació de pistes d'àudio i la generació de partitures automàtiques.

La separació de pistes d'àudio és un tema amplament investigat en l'àmbit de la processament de senyals i la música computacional. En l'actualitat, s'han desenvolupat diverses tècniques per aconseguir aquest objectiu, que van des de mètodes clàssics fins a enfocaments basats en Deep Learning.

Entre els mètodes clàssics més utilitzats es troben l'Independent Component Analysis (ICA) [12], que intenta descompondre l'àudio en fonts independents [5], la Principal Component Analysis (PCA), que cerca les components principals de l'àudio [7], i la Non-negative Matrix Factorization (NMF), que divideix l'àudio en components no negatives [8]. Aquests mètodes han demostrat ser eficaços en

diverses tasques de separació de pistes d'àudio, tot i que poden tenir limitacions en situacions de senyals molt complexes.

D'altra banda, els mètodes basats en Deep Learning han guanyat popularitat en els últims anys, especialment amb l'aparició de xarxes neuronals profundes. Les Xarxes Neuronals Convolucionals (CNNs), les Xarxes Neuronals Recurrents (RNNs) i les architectures especialitzades com la U-Net i la Wave-U-Net han estat utilitzades amb èxit per a tasques de separació de pistes d'àudio [4, 3, 6, 9]. Aquests models són capaços d'aprendre representacions de nivell alt i baix de l'àudio, permetent una millor separació de les pistes.

Pel que fa a la generació de partitures automàtiques, s'han desenvolupat diversos mètodes per convertir les dades d'àudio en notació musical. Aquests mètodes inclouen la detecció de notes [11], l'anàlisi harmònica [10], el reconeixement de melodies [1], i la transcripció automàtica [2]. Els avanços recents en Deep Learning han millorat significativament la precisió d'aquests mètodes, permetent la generació de partitures amb una qualitat cada vegada més alta.

En resum, el camp de la separació de pistes d'àudio i la generació de partitures automàtiques és ric i divers, amb una varietat de mètodes i tècniques disponibles. En aquest projecte, es buscarà explorar i comparar diferents enfocaments per aconseguir els objectius establerts.

5 METODOLOGIA

En aquest apartat es descriu la metodologia seguida per dur a terme el desenvolupament del projecte, incloent-hi els objectius específics, les tasques realitzades i les eines utilitzades.

5.1 Metodologia de Treball

La metodologia seguida per dur a terme el projecte ha estat iterativa, permetent una evolució gradual del sistema. Això ha implicat la realització de reunions setmanals amb el tutor per revisar el progrés, discutir possibles problemes i establir les següents tasques a realitzar.

S'han utilitzat diverses eines durant el desenvolupament del projecte, incloent-hi GitHub per al control de versions, Jupyter Notebook per a la visualització i anàlisi dels resultats, així com Notion, Obsidian i LaTeX per a la documentació del projecte i la presa de notes.

5.2 Esquema de Desenvolupament

El desenvolupament del projecte ha seguit el següent esquema:

1. Inicialment, es van abordar problemes simples utilitzant toy problems per a la prova de concepte.
2. A mesura que es van obtenir resultats satisfactoris, es va augmentar la complexitat de les dades utilitzades per al desenvolupament del sistema, passant a utilitzar conjunts de dades més reals com arxius MIDI i àudios de concerts.
3. S'ha mantingut una iteració constant entre el desenvolupament del codi i l'estudi del problema, permetent una millora contínua del sistema.

5.3 Diagrama de Gantt

A continuació adjunto el diagrama de Gantt referent al punt 3. el qual es pot visualitzar en una imatge que ocupa les dues columnes anomenada Figura 1.

Com es pot observar, les tasques de l'apartat 5 (documentació) són més paral·leles, és a dir, realment s'aniran realitzant durant el desenvolupament del projecte. En total s'ha aconseguit ajustar en el plaç de temps donat tot i que per a aconseguir-ho s'haurà de realitzar alguna hora extra, sobretot en la part inicial del projecte.

6 DEVELOPMENT

7 CONCLUSIONS

[illegible]

AGRAÏMENTS

[illegible]

REFERÈNCIES

- [1] Samer A Abdallah and Mark D Plumbley. A fundamental frequency estimator for speech and music. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–421. IEEE, 2004.
- [2] Emmanouil Benetos and Simon Dixon. Automatic music transcription: Challenges and future directions. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, 2013.
- [3] Thomas Grill, Jan Schluter, and Hermann Ney. Two-stage singing voice separation via deep convolutional neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 121–125. IEEE, 2017.
- [4] John R Hershey, Zhuo Chen, Jonathan Le Roux, and et al. Deep clustering: Discriminative embeddings for segmentation and separation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 31–35. IEEE, 2016.
- [5] Aapo Hyvarinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

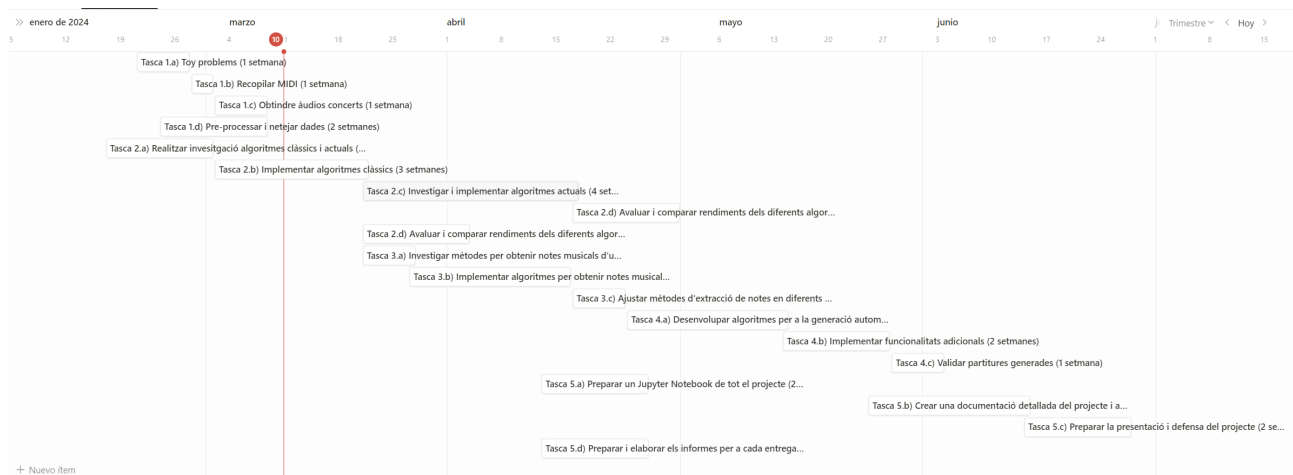


Fig. 1: Diagrama de Gantt del projecte.

- [6] Andreas Jansson, Eric J Humphrey, Nicola Montecchio, Rachel M Bittner, Anssi Kumar, and T Weyde. Singing voice separation with deep u-net convolutional networks. *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 23–27, 2017.
- [7] Ian Jolliffe. *Principal component analysis*. John Wiley & Sons, 2002.
- [8] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [9] Édouard Lva, Paul Rolland, and Mark D Plumbley. Wave-u-net: A multi-scale neural network for end-to-end audio source separation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 321–325. IEEE, 2018.
- [10] Bryan Pardo. An improved multipitch analysis model. *ICMC Proceedings*, 2:35–38, 2002.
- [11] Colin Raffel and Daniel PW Ellis. mir eval: A transparent implementation of common mir metrics. In *Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR)*, 2014.
- [12] Hiroshi Sawada, Nobutaka Ono, Hirokazu Kameoka, Daichi Kitamura, and Hiroshi Saruwatari. A review of blind source separation methods: two converging routes to ilrma originating from ica and nmf. *APSIPA Transactions on Signal and Information Processing*, 8:e12, 2019.

APÈNDIX

A.1 Secció d'Apèndix

.....

.....

.....

.....

A.2 Secció d'Apèndix

.....

.....