

La communication et l'agentivité de Djazz

Intelligence artificielle et savoirs musicaux relevant de l'oralité

Présenté par Heni Zouari, doctorant à l'EHESS (CAMS) et affilié à l'IRCAM (REACH / REPMUS)



Structure de la présentation

1

La communication

2

L'intelligence artificielle

3

Djazz ; agent dans un context de la musique tunisienne

4

Conclusion



La communication

- Action de transmettre, c'est-à-dire donner une information à quelqu'un. On dit qu'il y a une communication à faire à quelqu'un, ou une communication de pièces
- Désigne toutes les formes de relations, qu'elles soient commerciales, téléphoniques, ou d'un autre type. On parle aussi de voies de communication, c'est-à-dire des routes, chemins de fer, voies maritimes ou aériennes qui permettent les échanges.
- Communiquer signifie rendre commun ou faire connaître quelque chose. On parle aussi de chambres qui communiquent lorsqu'elles ont une porte commune.



Les éléments de la communication

- + Le destinataire et le destinataire : celui qui émet le message et qui le reçoit respectivement
- + Le message et le code : le contenu de la communication selon un code conventionnel entre le destinataire et le destinataire il peut être une langue parlée, signes, gestes...
- + Le canal : c'est le support dans lequel la communication est mise en place il peut être un téléphone ou tout simplement par voie orale (en face)
- + L'intention : c'est la réponse du pourquoi on communique
- + L'effet. : la ou les conséquence de cette discussion chez le destinataire ou le destinataire



Les niveaux de la communication

- Niveau anthropologique ou naturel : La communication est une expérience humaine fondamentale, présente dans toutes les sociétés et dépendante de leurs cultures. Elle traduit toujours une manière de voir l'autre et de structurer les relations sociales. "l'acte de l'échange d'informations, de propriétés, d'états, s'explique par des lois et des relations de cause à effet"(Maigret,2022) : cela veut dire que le destinataire émet selon une cause qu'on l'a déjà mentionné (l'intention) et le destinataire reçoit (effet)
- Niveau technique : Ce niveau se caractérise par le passage d'un canal de communication primitif, fondé sur la rencontre directe, à des techniques modernes qui l'ont transformé en communication à distance, grâce aux médias et aux technologies ce qui a facilité l'échange à l'échelle mondiale



Les modèles du pouvoir des médias

Le niveau technique de la communication a joué un rôle indispensable dans la naissance du pouvoir des médias :

1- *Stimulus-réponse*

2- *Two-Step Flow/ Flux en deux étapes*

3- *Usage et Gratifications*

4- *Construction sociale de la réalité*

5- *Agenda setting*

6- *Spirale de la silence*

7- *Chemins de persuasion*



Stimulus-réponse

=> C'est le **premier modèle** utilisé pour analyser et comprendre l'influence des médias sur les récepteurs par Serge Tchakhotine dans son ouvrage "Viol des foules" .

+l'émetteur :

est censé pouvoir « injecter » n'importe quelle idée dans l'esprit de n'importe qui.

Son discours est un stimulus qui entraîne instantanément, chez l'auditeur, la réponse qu'il a programmée



Stimulus-réponse

+Ce modèle était en vigueur pendant la Première Guerre mondiale.

les forces armées dépendaient de la production industrielle de leurs pays ; elles avaient donc besoin du soutien des civils.

C'est là que la propagande est intervenue pour diffuser des messages de patriotisme, de responsabilité collective, d'amour de la nation et de haine de l'ennemi. Ici, les messages sont directs et influencent immédiatement les destinataires.

=> Finalement, la relation entre le public et les médias repose ici sur un modèle où le récepteur demeure passif. Il accepte les messages reçus sans y réagir, ce qui correspond au schéma de stimulus-réponse



Two-Step Flow/ Flux en deux étapes

l'ouvrage l'influence personnelle (Personal Influence), de Katz et Lazarsfeld, propose une nouvelle vision de l'effet des médias sur les individus, **différente** de l'influence directe supposée par le modèle stimulus-réponse

(une enquête auprès d'environ 800 jeunes femmes dans un comté des États-Unis afin de déterminer sur quelles bases elles faisaient leurs choix en matière de cinéma, de mode et de consommation de produits culturels.)

leaders d'opinion jouent donc un rôle crucial dans la propagation et l'interprétation des messages médiatiques, en filtrant et en transmettant l'information à leur réseau social, souvent de manière plus efficace que les médias eux-mêmes.



Two-Step Flow/ Flux en deux étapes

Média => message => leader d'opinion => public

“les leaders d'opinion sont plus fréquemment des femmes mariées dans le cas des biens de consommation, des jeunes femmes dans le cas de la mode et du cinéma, et des femmes de niveau social élevé dans le cas des affaires publiques...”

=> Cette étude souligne l'importance de l'influence sociale locale et invite à repenser la manière dont les médias affectent les comportements et les opinions individuelles



Usage et Gratifications

une continuité de flux de communication à deux étages,

propose ainsi un renversement de perspective que résume la formule de Katz :

« Il faut concentrer l'attention moins sur ce que les médias font aux gens que sur ce que les gens font des médias.»”

Usage et Gratifications

Selon une étude menée par **Matilda White Riley et John W. Riley Jr.**, l'usage des programmes télévisés par les enfants varie selon leur intégration sociale : (certains enfants font partie de peer groups (petits groupes d'amis) tandis que d'autres sont isolés).

+Les enfants intégrés utilisent les émissions, notamment romantiques, pour développer leurs relations sociales et leurs interactions avec le sexe opposé,

+ les enfants isolés s'en servent comme moyen d'évasion.

=> Les médias délivrent donc différents messages, et le public choisit comment les utiliser : l'effet devient ainsi personnalisé, variant selon chaque individu ou chaque groupe d'individus

La réalité est la qualité de ce qui est effectif, de ce qui existe : tous les objets et les événements qui se produisent autour de l'être humain sont réels

Si aucun média ne nous a transmis des informations sur l'état du pôle Antarctique, notamment sur la diminution de la glace due à la pollution et au réchauffement climatique,

nous n'aurions aucune information sur ce problème ; **il ne serait donc pas réel pour nous**

Le 26 janvier 1991 par exemple, en pleine guerre du Golfe, CBS consacre 8 secondes à un rassemblement de 50 000 pacifistes et 13 secondes à quelques centaines de partisans du président Bush. Le même jour, ABC ne consacre que 1 minute à 100 000 manifestants anti-guerre rassemblés à Washington, et plus de 2 minutes 30 à un reportage sur des ouvriers « fiers de fabriquer le missile Patriot ». Dans les deux cas, par le simple choix qu'il opère entre les événements à couvrir, et indépendamment des commentaires positifs ou négatifs qui accompagnent ces reportages, le journal télévisé tend à donner l'image d'une Amérique « virile »(GREGORY,1998)

=> Quoique, dans cet exemple, les deux avis coexistent dans les médias, le paramétrage de la durée de transmission peut déjà favoriser l'un au détriment de l'autre.

les médias ne **sont pas capables d'influencer** directement le comportement ou les opinions des individus, mais ils peuvent déterminer ce qui est mis en avant ou non dans les sujets abordés dans les réseaux d'information.

les «études d'agenda setting » : les médias nous disent non pas ce qu'il **faut penser**, mais **à quoi il faut penser**; ils jouent le rôle d'un «maître de cérémonie» ou encore d'un tableau d'affichage sur lequel viendraient s'inscrire les problèmes qui doivent faire l'objet du débat dans une société

=> un sujet devient plus important qu'un autre, non pas parce qu'il serait intrinsèquement plus intéressant pour les individus, mais parce qu'il est plus présent dans le vaste réseau d'information



Spirale de la silence

La publication de prévisions favorables en faveur d'un parti, ont poussé les électeurs à adopter le même choix. Il y a les effets de suivisme : La peur de l'isolement, en suivant l'opinion dominante...

Elisabeth Noelle-Neumann : l'un des deux camps en présence **accroît son avance** pendant que **l'autre recule**. La tendance à s'exprimer dans un cas et à garder le silence dans l'autre, engendre un processus en spirale qui **installe** graduellement une **opinion dominante**.

=> La spirale du silence montre que les médias ne force pas les individus à ce qu'il faut penser(modèle stimulus-réponse) ni à quoi faut il penser (agenda setting), mais ce qu'il ne faut pas penser, peur d'être différent de la volonté majoritaire, être marginalisé de la société

Les médias tendent surtout à fournir des **thèmes de discussion**, autour desquels les individus **débattent** ensuite dans leur environnement social, ce qui contribue à **renforcer l'importance de ces sujets pour le grand public**

=> **Kapferer affirme que la persuasion dépend à la fois du public et du type de sujet, selon qu'il implique fortement ou faiblement les individus, et il souligne que la persuasion repose moins sur le contenu des messages que sur la répétition et la pression médiatique**

=> On peut conclure que, avec l'évolution technologique qui favorise la communication de masse, de nombreux questionnements ont émergé quant au pouvoir des médias sur les individus, ce qui nous conduit à une interrogation plus profonde avec l'apparition de l'intelligence artificielle : une IA peut-elle improviser avec l'être humain, ou assistera-t-on à une persuasion technologique en faveur d'une idée, comme dans les modèles du pouvoir des médias ?



l'intelligence artificielle

Le domaine de l'intelligence artificielle, ou IA, s'intéresse non seulement à la compréhension, mais aussi à la création d'entités intelligentes, c'est-à-dire de machines capables de calculer comment agir de manière efficace dans une grande variété de situations nouvelles

le «machine learning» est une forme d'IA qui permet à un système d'apprendre à partir de données plutôt que par une programmation explicite. Cependant, ce n'est pas un processus simple, car il utilise une variété d'algorithmes qui apprennent de manière répétitive à partir des données pour s'améliorer, décrire des données et prédire des résultats qui nécessitent souvent une ingénierie manuelle des caractéristiques (feature engineering)



Deep learning / apprentissage automatique

Étant un sous-catégorie du machine learning, le deep learning repose sur l'utilisation de réseaux de neurones artificiels comportant plusieurs couches, capables d'apprendre automatiquement des représentations hiérarchiques complexes à partir des données qui

+ Apprendre directement à partir de données brutes telles que des images, du texte ou de l'audio



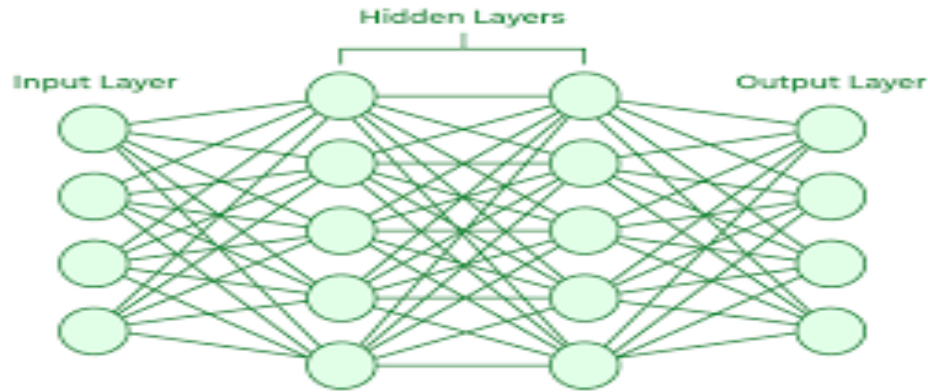
type de réseaux de neurones

Fully Connected Network (FCN)

Convolutional Neural Network (CNN)

Recurrent Neural Network (RNN)

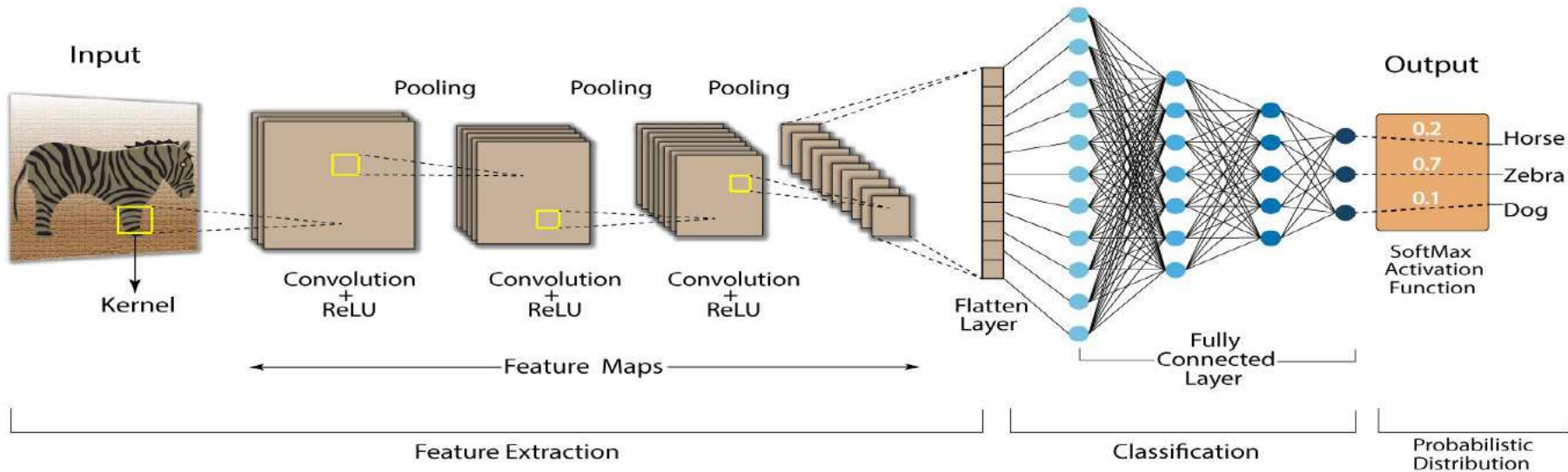
FCN



+Un FCN est un reseau de neurones ou chaque neurone d'une couche est connecté à tous les neurones de la couche suivante, ce qui permet de combiner toutes les informations de manière globale

=> Classification simple / Régression (ex : prédire un prix ou une valeur)

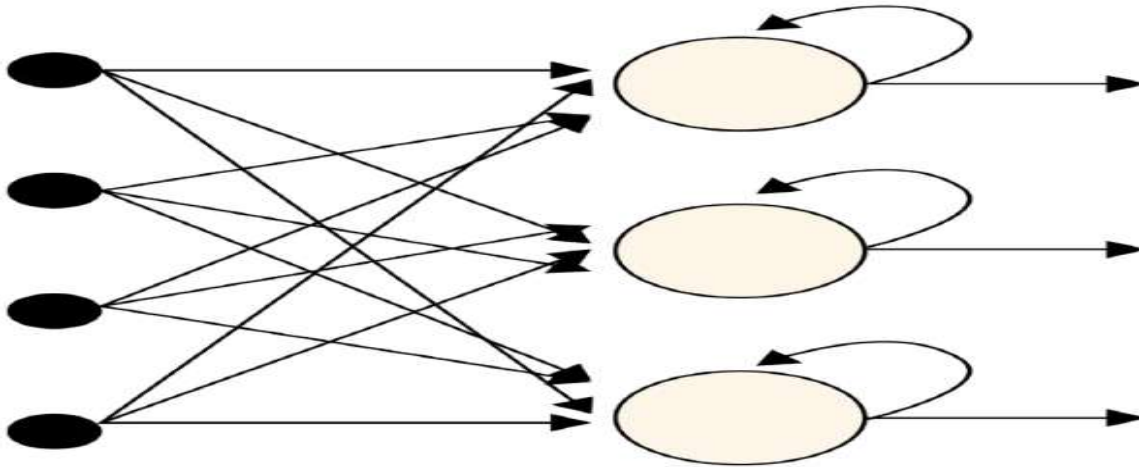
CNN



+Un CNN est un réseau de neurones conçu pour analyser des structures spatiales en détectant automatiquement des motifs locaux comme les bords, formes ou textures dans les données.

=>Images (reconnaissance, classification)/ Vidéo

RNN



+Un RNN est un réseau de neurones capable de traiter des séquences en gardant une mémoire des informations précédentes afin de comprendre le contexte dans le temps.

=> Audio / texte



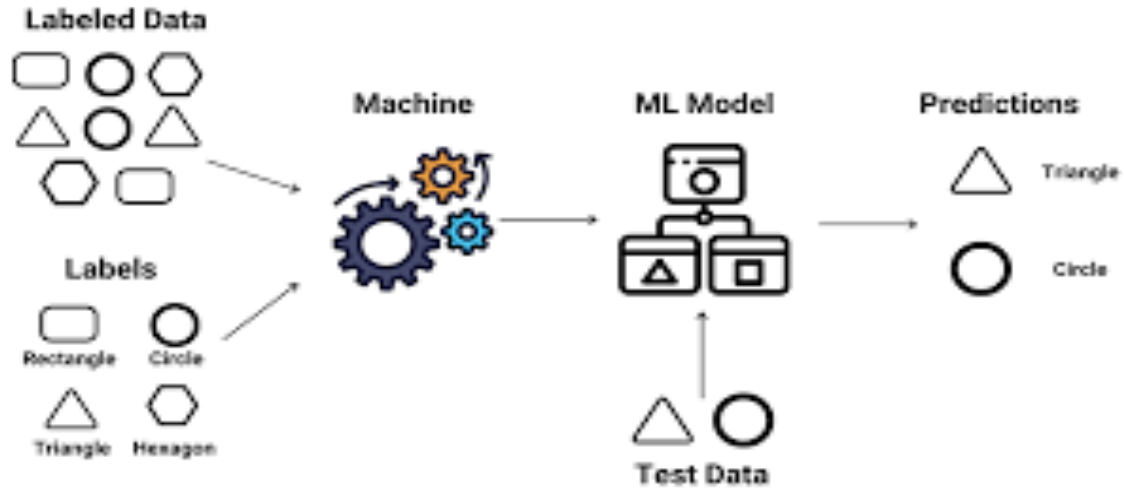
Méthodes d'apprentissage

Supervisé

Non-supervisé

Par renforcement

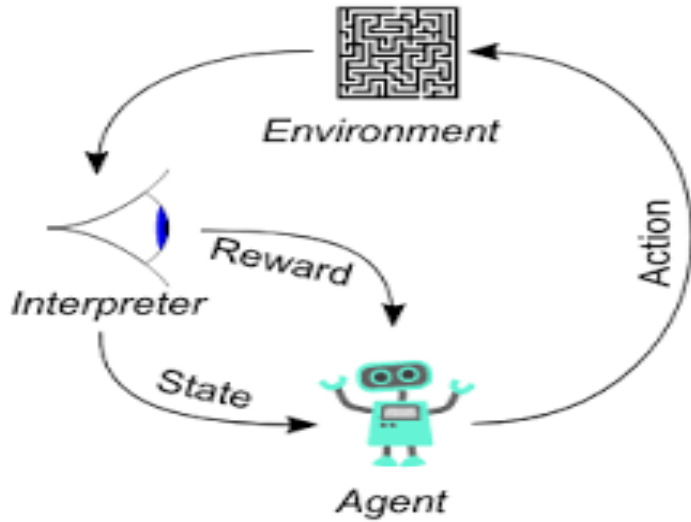
Supervisé



+les données d'entrée sont classifiées(étiquetées)



+les données d'entrée ne sont pas classifiées(étiquetées)



=> L'algorithme apprend à partir d'une série de renforcements récompenses et punitions. Par exemple, à la fin d'une partie d'échecs, on dit à l'agent s'il a gagné (récompense) ou perdu (punition).



Le transformer

Le Transformer est une nouvelle architecture de réseaux de neurones introduite dans un article en 2017 intitulé “Attention Is All You Need”. Contrairement aux RNN, il ne traite pas les données de manière séquentielle, mais utilise un mécanisme d’attention qui lui permet de considérer simultanément toutes les parties d’une séquence. Cette approche le rend particulièrement efficace pour les tâches liées au texte et à l’audio, où le contexte global est important. Grâce à sa capacité à bien se généraliser et à passer à l’échelle, cette architecture a conduit au développement des grands modèles de langage (LLM), comme GPT-2, qui a marqué une étape importante dans l’amélioration des performances et l’adoption de ces modèles par les utilisateurs.



L'un des modèles connus dans le domaine de la génération musicale repose sur l'architecture Transformer, comme par exemple Music Transformer ou encore Jukebox, qui exploitent cette approche pour générer des séquences musicales cohérentes.

Toutefois, ces modèles nécessitent généralement de très grandes bases de données ainsi qu'une puissance de calcul importante afin d'obtenir des résultats satisfaisants.

C'est pour cette raison que nous avons choisi de travailler avec Djazz, qui se comporte davantage comme un agent interactif dans un contexte d'improvisation musicale, plutôt que comme un modèle de génération classique dépendant d'un apprentissage massif sur de larges corpus.



Djazz

Le système Djazz, développé par le CAMS (EHESS) et l'IRCAM, est un « improvisateur artificiel » qui génère de la musique en s'appuyant sur une base de séquences enregistrées.

Djazz constitue une variante du logiciel ImproteK, développé par Marc Chemillier et Jérôme Nika dans le cadre de la collaboration entre deux laboratoires de recherche, le CAMS et l'IRCAM.

La nouvelle version Djazz 2.0 est développée par Daniel Brown

Dans le jazz, les musiciens se basent souvent sur une grille d'accords pour structurer l'improvisation autour de l'harmonie..

LES FEUILLES MORTES

Am	D7	G	C	F#Ø	B7	Em	E7
Am	D7	G	C	F#Ø	B7	Em	%
F#Ø	B7	Em	E7	Am	D7	G	C
F#Ø	B7	Em Eb7	Dm Db7	CM7	B7	Em	%

Apprendre le
Jazz Manouche...

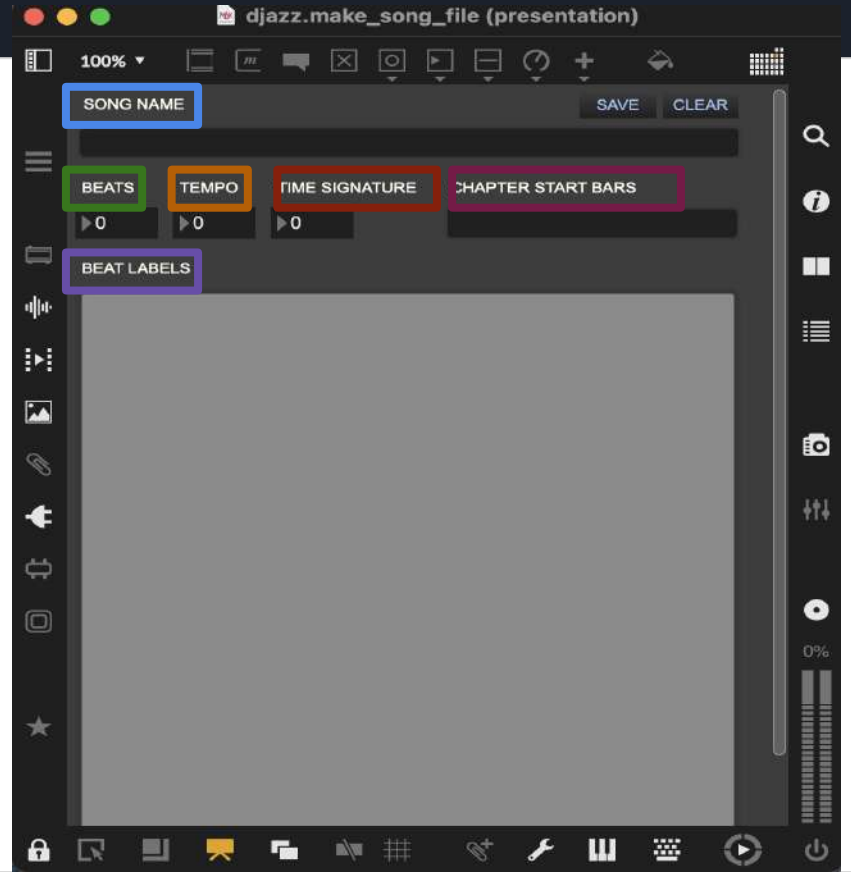
Création de la chanson

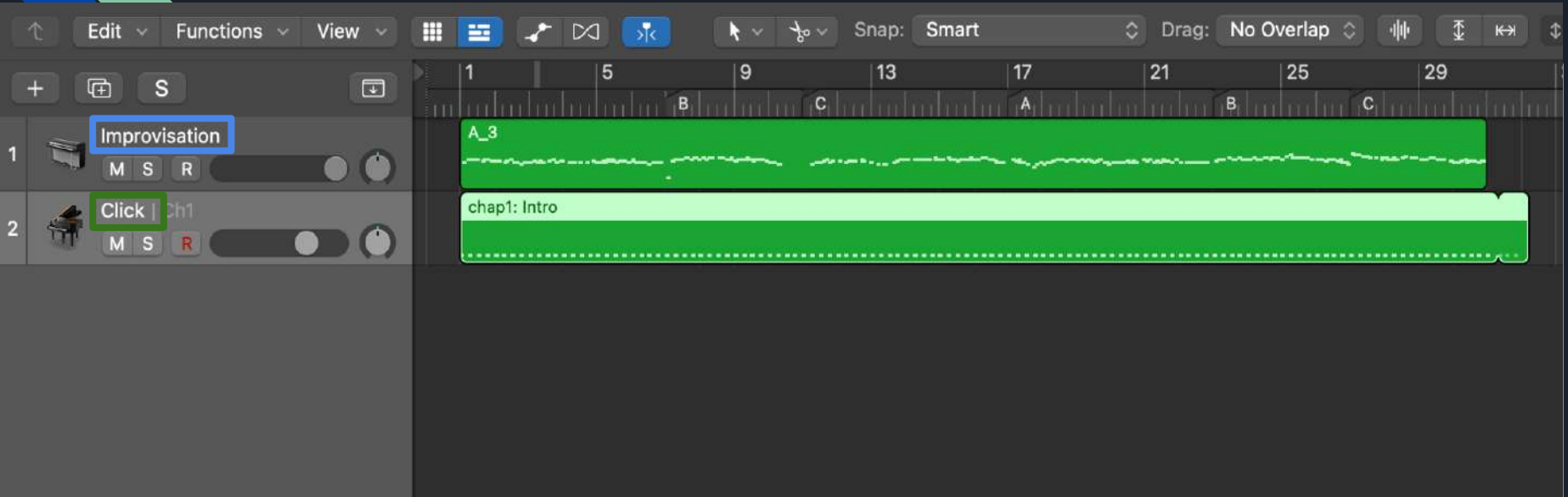
- + **Song Name** : le nom de la chanson
- + **Beats** : Nombre total de pulsations
- + **Tempo** : tempo sera chargé dans le métronome de Djazz lorsque le morceau sera chargé
- + **Time signature** : Nombre de pulsations par mesure
- + **Chapter** :

si un morceau contient quatre chapitres de 16, 16, 32 et 64 mesures, dans cet ordre, les débuts de chapitres seront 1, 17, 33 et 65. Entrez ces nombres sans virgules.

+ **Beat labels**

Une étiquette d'accord comporte trois parties : la fondamentale, la qualité et le nombre de pulsations de l'accord. La manière de saisir les étiquettes est la suivante : (A maj7 4) (G 7 4)...

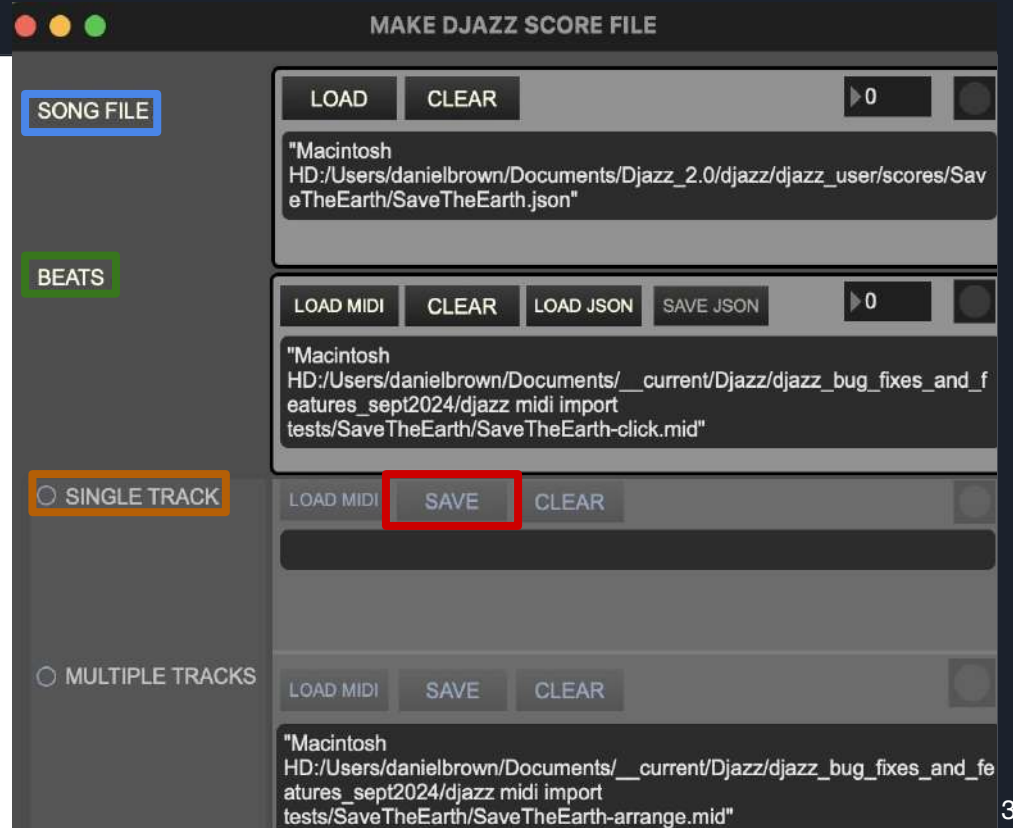




- + **Improvisation** : La piste où il ya a notre improvisation en midi
- + **Click** : La piste où il ya les pulsations en midi

Création de la partition

- + **Song file** : La chanson qu'on a créé
- + **Beats** : La piste midi des pulsations
- + **Single track** : la piste midi de l'improvisation
- + **Save** : on doit sauvegarder ce fichier



Djazz Midi

+ **Song file** : La chanson qu'on a créé

The screenshot shows the DJAZZ software interface. At the top, there's a header with the 'diaz digital jazz' logo and 'MIDI' and 'AUDIO' tabs. Below this, there are controls for 'TAP' (metronome, 200), 'PITCH' (108), and 'CHAN' (16). A 'browse' button is next to a text field containing 'Free4'. Below that, a 'CHAPTER' section is highlighted with a blue oval, showing a list of bars (1, 2, 3, 4) and a 'CHAPTER SONG' button. At the bottom, there are 'AUDIO' and 'MIDI INPUT' sections with various sliders and buttons.

The screenshot shows the DJAZZ software interface. At the top, there's a header with the 'diaz digital jazz' logo and 'MIDI' and 'AUDIO' tabs. Below this, there are controls for 'TAP' (metronome, 60), 'PITCH' (108), and 'CHAN' (16). A 'browse' button is next to a text field containing 'select', which is highlighted with a blue box. Below that, there are 'BEATS' (1, 0), 'SECTION' (1, 1), 'TEMPO' (60), and 'BEAT' (0) controls. At the bottom, there are 'AUDIO' and 'MIDI INPUT' sections with various sliders and buttons.

OUT GLOBAL PORT NONE

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

CHANNEL 1 CHANNEL 2 CHANNEL 3 CHANNEL 4 CHANNEL 5 CHANNEL 6 CHANNEL 7 CHANNEL 8 CHANNEL 9 CHANNEL 10 CHANNEL 11 CHANNEL 12 CHANNEL 13 CHANNEL 14 CHANNEL 15

AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S... AU DLS S...

TRACKS

15	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
14	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
13	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
12	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
11	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
10	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
9	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
8	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
7	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
6	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
5	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
4	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse
3	↑	▼	▼	0	▲	▼	▼	0	▲	►	0	⌚	1/2	3/2	2	4	⌛	1	2	3	4	5	6	7	8	🔊	🔊	>255	MAX	↕	►	0	📁	🔊	🎵	Browse

On choisit une piste pour télécharger la partition qu'on a créé avec l'outil "Make Score"

En s'inspirant de la théorie de Mead, qui distingue trois dimensions dans la formation du soi — la société, le “moi” et le “je” —, on propose de transposer ce modèle à l'analyse de l'improvisation musicale. Ainsi, on peut décomposer l'improvisation en trois niveaux :

A : les clichés essentiels, qui représentent l'héritage collectif et la dimension sociale de la musique ;

B : le cheminement maqamique, reflétant une appropriation personnelle mais encore guidée par les cadres existants ;

C : la créativité et l'apport personnel, correspondant à la part spontanée et innovante de l'improvisateur.

Essaie sur une improvisation de la musique Tunisienne

Une improvisation qui suit le cheminement suivant :

A B C A B C etc...

⇒ On peut créer une chanson en remplaçant les accords par ces chiffres et en toute en suivant le processus mentionné



Impro originale



Impro en midi

Edit Functions View
 A_3 on Track Riginal
 Time Quantize (classic): 1/16 Note Q
 Strength 100
 Swing 0
 Scale Quantize Off N Q
 Velocity 32

Inst 1
on Track 95 DJazz

Time Quantize (classic)

1/16 Note Q

Strength 100

Swing 0

Scale Quantize

Off N Q

Velocity 32

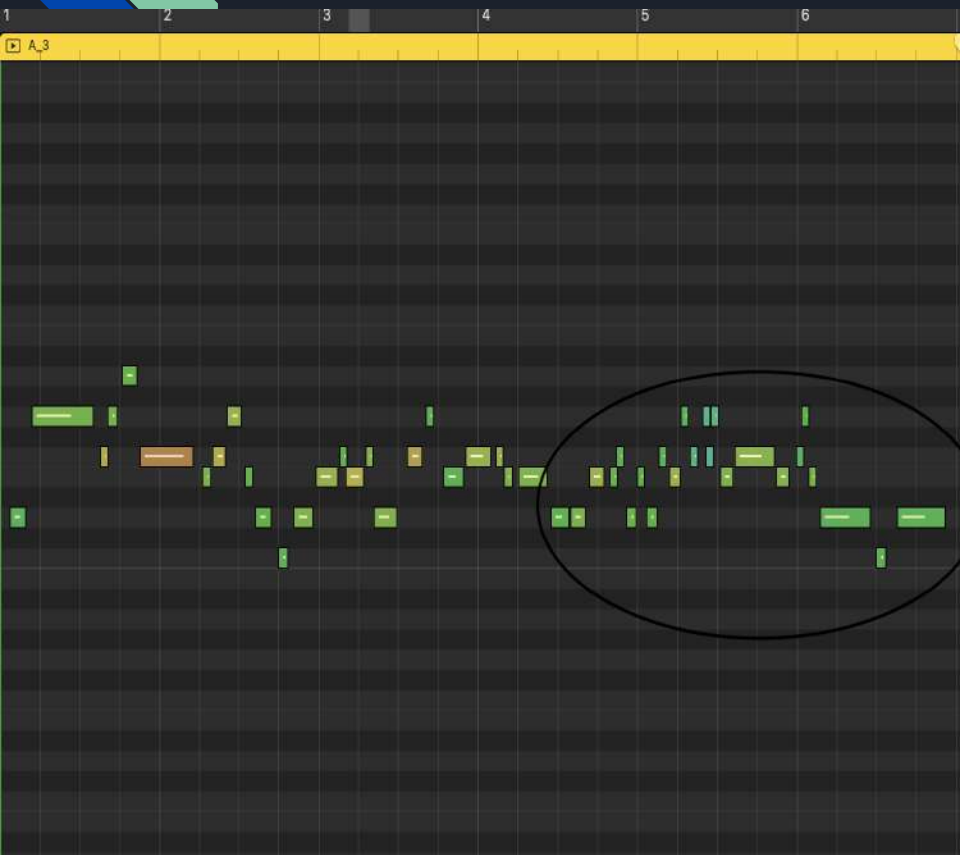


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

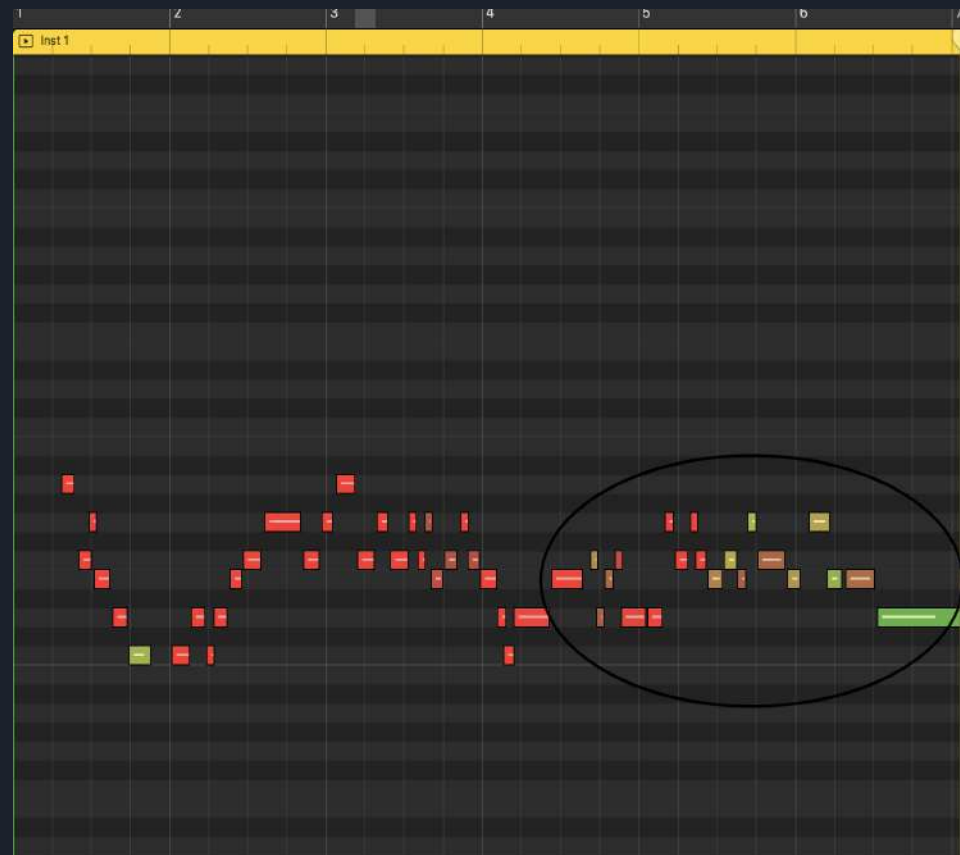
Inst 1 Inst 1



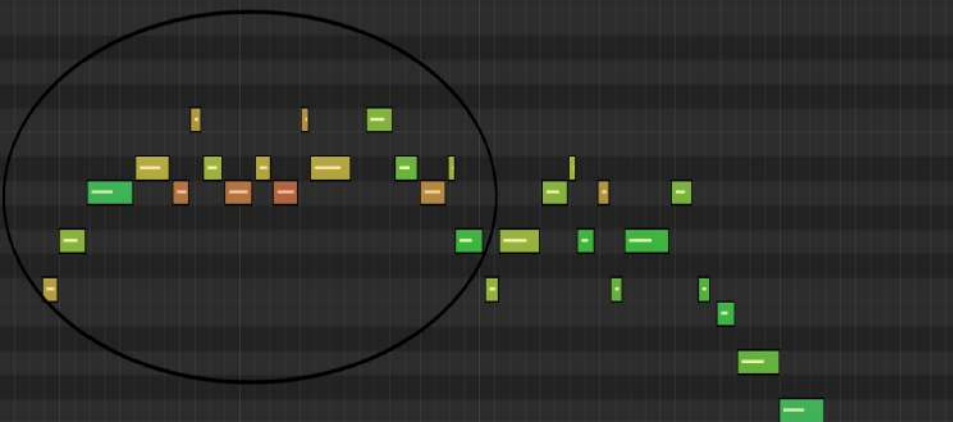
Phrase 1 Originale



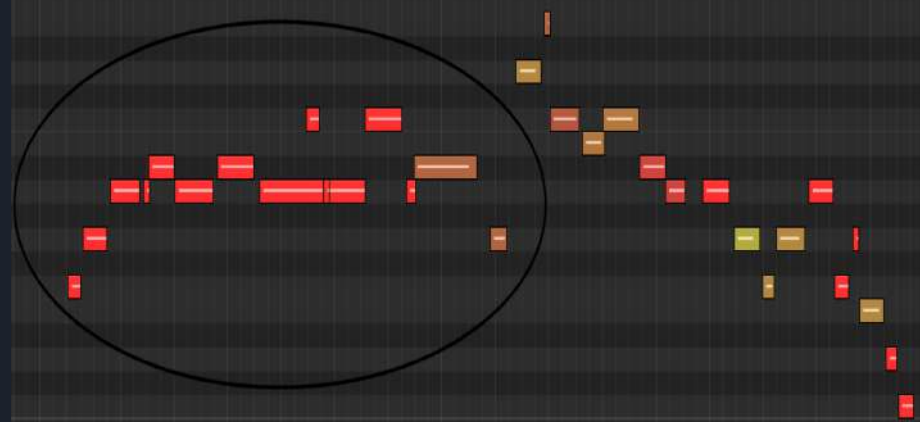
Djazz



Phrase 2 Originale



Djazz





Conclusion

C'est pourquoi, dans notre thèse, nous cherchons à partager la musique tunisienne avec le monde entier via l'application Djazz, tout en générant une musique qui correspond aux spécificités socioculturelles, puisqu'il ne s'agit pas d'un modèle nécessitant une large base de données, mais plutôt d'un agent qui écoute et reconstitue.

Merci !

