



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 16971

The contribution was presented at JIM 2016 :
<http://jim2016.gmea.net/>

To cite this version : Guyot, Patrice and Pellegrini, Thomas *Vers la transcription automatique de gestes du soundpainting pour l'analyse de performances interactives*. (2016) In: Journées d'Informatique Musicale (JIM 2016), 31 March 2016 - 2 April 2016 (Albi, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

VERS LA TRANSCRIPTION AUTOMATIQUE DE GESTES DU SOUNDPAINING POUR L'ANALYSE DE PERFORMANCES INTERACTIVES

Patrice Guyot

Université de Toulouse, France ; IRIT
guyot.patrice@gmail.com

Thomas Pellegrini

Université de Toulouse, France ; IRIT
thomas.pellegrini@irit.fr

RÉSUMÉ

L'analyse objective et la documentation de performances interactives est souvent délicate car extrêmement complexe. Le Soundpainting, langage gestuel dédié à l'improvisation guidée de musiciens, d'acteurs, ou de danseurs, peut constituer un terrain privilégié pour cette analyse. Des gestes prédéfinis sont produits pour indiquer aux improvisateurs le type de matériel souhaité. La transcription des gestes en vue de la documentation de performances semble tout à fait réalisable mais très fastidieuse. Dans cet article, nous présentons un outil de reconnaissance automatique de gestes dédié à l'annotation d'une performance de soundpainting. Un premier prototype a été développé pour reconnaître les gestes filmé par une caméra de type Kinect. La transcription automatique de gestes pourrait ainsi mener à diverses applications, notamment l'analyse de la pratique du soundpainting en général, mais également la compréhension et la modélisation de performances musicales interactives.

1. INTRODUCTION

La musique, sous de nombreuses formes, comporte des parties plus ou moins improvisées (musiques traditionnelles, basse chiffrée dans la musique baroque, jazz, etc). La documentation de ces improvisations est un enjeu majeur en musicologie, et trouve de nombreuses applications en éducation musicale.

La documentation de la musique improvisée se heurte bien souvent à des difficultés de compréhension et de conceptualisation du contexte de jeu et des interactions entre acteurs de la performance musicale. Par exemple, en ethnomusicologie, « l'étude du contexte culturel et social de la musique est indissociable de l'analyse musicale proprement dite » [3]. Au sein même du cadre restreint du concert, de nombreuses informations peuvent circuler entre les acteurs de la performance et le public lors de l'improvisation, qu'il semble parfois difficile d'écrire ou de documenter [2].

En dehors du jeu instrumental, certaines informations sont transmises par des gestes. Certains de ces gestes sont liés à l'interprétation de la pièce, comme dans le cas des gestes du chef d'orchestre classique. Des signes précis peuvent par ailleurs être utilisés pour indiquer l'exécution

d'une partie instrumentale prédéterminée. Par exemple, Frank Zappa pouvait donner des indications précises à ces musiciens lors d'une performance, associant ainsi une part importante d'imprévu et d'interaction à des morceaux écrits [10].

Ces deux types de gestes relèvent de natures spécifiques. Si certains gestes de chef d'orchestre classique s'appuient sur un code largement partagé (par exemple au niveau du tempo et des nuances), d'autres gestes sont une forme d'expression plus spontanée et personnelle, qu'il est difficile de transcrire ou de documenter. Les gestes indiquant des parties prédéterminées, au contraire, peuvent être répertoriés et annotés. Néanmoins, leur usage et leur compréhension se limite la plupart du temps au travail interne d'un ensemble musical particulier et sont difficilement compris par d'autres musiciens.

Depuis sa création il y a une trentaine d'années, le soundpainting prend de l'ampleur dans le paysage de la musique improvisée. Dépassant le contexte interne à un orchestre spécifique, ce dispositif a l'ambition de proposer un code universel pour composer une improvisation. Les possibilités de ce code s'étendent aujourd'hui hors des frontières du monde musical et touchent d'autres champs de création, par exemple le théâtre et la danse.

Le soundpainting est un langage de composition gestuel. Il interroge ainsi la séparation du rôle de compositeur et de chef d'orchestre (effective depuis la période romantique, malgré de nombreuses exceptions [6]). Ces différences font écho aux différences entre un système établi (notation musicale) et un travail d'interprétation plus spontané et plus difficilement documentable. Il semblerait que les gestes du soundpainting appartiendraient plutôt à la catégorie de la composition à partir d'un système établi, avec des gestes souvent effectués de la manière la plus neutre possible.

La standardisation des gestes de soundpainting en fait un terrain favorisé pour la documentation de performances improvisées. Toutefois, l'annotation des gestes représente un travail fastidieux qui dans les faits semble rarement effectué. L'utilisation de l'ordinateur, par ailleurs, offre aujourd'hui de nouvelles possibilités pour l'écriture ou l'annotation des performances interactives. Il peut par exemple être utilisé dans le cas de l'écriture de musique mixte, qui associe musiciens instrumentistes et processus électroniques lors d'une performance [7]. Il est aussi

employé dans le cadre d'improvisations pour donner des signes à des musiciens en temps réel [4].

Dans ce contexte, la reconnaissance automatique de gestes trouve également des applications musicales [8]. Jusqu'à très récemment trop difficile pour être largement utilisée, la reconnaissance de gestes définis pourrait très prochainement se démocratiser, comme l'est aujourd'hui la transcription automatique de la parole sur un téléphone portable. Dans cet article, nous proposons une nouvelle application de la reconnaissance de gestes dans le cadre de l'annotation automatique de performances interactives. Dans une publication précédente [14], nous avons présenté en détails un premier prototype. Dans le présent article, nous nous intéressons plus particulièrement à sa possible utilisation pour l'analyse de performances.

La partie 2 décrit notre terrain d'étude, le soundpainting. La transcription des performances de soundpainting est discutée en partie 3 et la reconnaissance automatique de gestes est présentée en partie 4. Nous présentons alors notre prototype dans la partie 5 et ses applications dans la partie 6.

2. SOUNDPAINTING

Le soundpainting est un langage gestuel développé dans les années 70 par le compositeur Walter Thompson. Il est destiné à l'élaboration spontanée de performances interactives dans lesquelles peuvent participer musiciens, danseurs, acteurs et plasticiens [16]. Le chef d'orchestre, appelé « soundpainteur », indique aux exécutants (ou « performeurs ») le type de matériel souhaité. Dans le cadre de la musique par exemple, le soundpainteur pourra indiquer à un ensemble de contrebasses de jouer des notes longues et graves. Ainsi le soundpainting se distingue de la direction d'orchestre de part les possibilités qu'il offre dans la composition d'une improvisation.

De nos jours, le soundpainting s'impose dans de nombreux types de performances, des performances totalement improvisées à celles utilisant de plus ou moins nombreuses parties écrites. De nombreux festivals dédiés ont ainsi vu le jour. La France occupe une place privilégiée dans le domaine et a vu la création d'une vingtaine d'ensembles professionnels au cours de ces dernières décennies¹. Par ailleurs, le soundpainting offre également des possibilités simples d'inclure des temps d'improvisation dans des performances écrites. Au final, ce langage semble donc affirmer une place conséquente dans les domaines du jazz, de la musique contemporaine, et de l'enseignement musical [5].

Dans la pratique, le soundpainteur réalise un ensemble de commandes gestuelles en s'appuyant sur la syntaxe «*qui, quoi, comment, quand*». Les gestes dits fonctionnels (*qui, quand*) permettent d'une part d'identifier un ensemble d'acteurs, d'autre part d'indiquer l'instant où les consignes doivent être exécutées. Les gestes de la catégorie « sculpture » (*quoi, comment*) permettent quant

à eux de définir le matériel qui sera utilisé par les improvisateurs, la précision du *comment* étant facultative. La séquence suivante, composée des trois gestes : **Whole Group - Long note - Volume Fader : Piano - Play** indique donc à l'ensemble des exécutants de jouer des notes longues, à une nuance piano, en commençant au moment précis où le geste **Play** est exécuté.

Le processus d'improvisation est structuré à travers l'acceptation commune d'une grande variété de matériel et d'interprétation pouvant résulter d'une même séquence de gestes. Les gestes du soundpainting ne sont pas effectués de manière isolée, car le soundpainteur puise son inspiration des performeurs. La réalisation d'une performance associe donc le travail de création à celui d'interprétation des gestes et du contexte [11].

L'enseignement professionnel du soundpainting passe par la certification des soundpailleurs. Les candidats doivent maîtriser l'aspect théorique, physique, et compositionnel d'un ensemble de gestes de base.

3. TRANSCRIPTION D'UNE PERFORMANCE

L'analyse de performance d'une musique non écrite est une question récurrente abordée par de nombreux musicologues. Elle nécessite de pouvoir documenter chacune des réalisations de manière objective et pérenne. L'analyse de performance peut être facilitée via la comparaison de différentes réalisations. La captation audio et vidéo semble aujourd'hui à la base de la documentation de performances. Toutefois, la vidéo, (comme pourrait l'être dans une moindre mesure l'audio), est soumise au cadre de la prise de vue, et par ce fait élimine forcément une partie des informations échangées lors des improvisations. Par exemple, dans le cadre du soundpainting, il semble difficile de filmer d'un seul plan le soundpainteur et les performeurs.

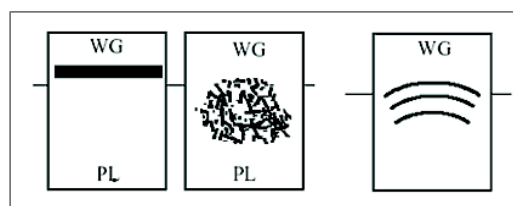


Figure 1. Extrait d'une notation d'une composition de soundpainting par Walter Thompson. Issu de <http://tinyurl.com/jec1pd5>

De manière générale, la captation audiovisuelle reste soumise à un déroulé temporel lors de son exécution, ce qui peut limiter son analyse. Ainsi, selon Gilbert Rouget, « toute réflexion tant soit peu rigoureuse sur ce genre de musique exige d'abord qu'on la transforme de chose immatérielle, invisible, se défaisant sans cesse à mesure qu'elle se fait, en objet permanent et repérable [...]. L'opération revient donc à donner de la dimension temporelle de la musique un équivalent spatial. » [15]. Par ailleurs, la transcription graphique de la musique offre

1. <https://fr.wikipedia.org/wiki/Soundpainting> consulté le 15 mars 2016

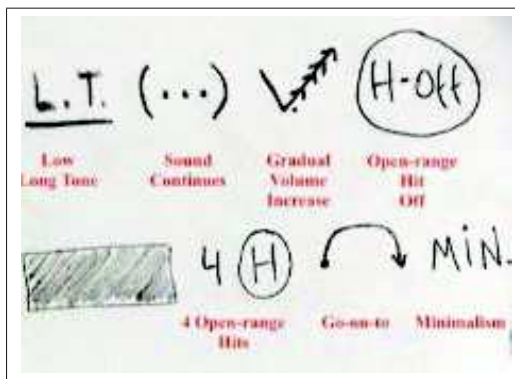


Figure 2. Exemples de notation de signes de soundpainting sous la forme d’esquisses [9].

d’autres intérêts, comme la possibilité de souligner un élément musical pour attirer l’attention du lecteur [4].

A travers l’usage d’un code strict de gestes de référence, les performances de soundpainting présentent des facilités de documentation. Dans le cadre de la composition, le créateur du soundpainting, Walter Thompson, décrit ainsi le processus d’écriture des gestes du soundpainting pour la préparation d’une performance². La figure 1 illustre un extrait de cette composition. On y trouve trois séquences de gestes successives, représentées par des rectangles. Dans ces trois séquences, le compositeur s’adresse à l’ensemble des musiciens (*Whole Group, WG*). Les dessins au centre des carrés représentent les gestes référant respectivement à *notes longues*, *pointillisme* et *scanner*, ce dernier geste invitant le musicien désigné lors d’un balayage du bras à jouer (parfois pour une durée très courte). Les deux premières séquences sont exécutées lorsque le geste *Play (PL)* est réalisé. Dans la dernière séquence, le processus de *scanner* s’applique immédiatement et prend effet dès lors qu’un *performer* est désigné.

Le travail de notation de performance de soundpainting est également évoqué dans [9]. L’auteur réalise des esquisses de chaque geste et justifie cette pratique par le fait qu’elle l’aide à travailler le soundpainting de manière autonome et quotidienne, en sortant du cadre plus ponctuel des performances et répétitions collectives. La figure 2 illustre quelques exemples de gestes représentés par ces esquisses.

A ce jour, au delà de ces exemples, il semblerait que la notation des gestes de soundpainting ne soit pas une pratique très développée. La pratique s’enseigne au cours de stages au cours desquels la transmission orale est privilégiée. Ainsi, malgré l’engouement que crée le soundpainting, on recense à notre connaissance peu d’analyses systémiques de pratiques et de performances. En effet, la transcription manuelle des gestes est fastidieuse et doit s’appuyer sur la vidéo pour être effectuée. Dans ce contexte, tout comme la transcription automatique ou le piano-roll [4] peuvent se révéler de formidables outils de documentation musicale, la reconnaissance automatique

2. <http://tinyurl.com/jec1pd5> consulté le 15 mars 2016

appliquée au soundpainting pourrait permettre de documenter facilement un grand nombre de performances, et ainsi faciliter l’analyse de cette pratique.

4. RECONNAISSANCE DE GESTES

La reconnaissance de gestes désigne le fait d’identifier de manière automatique les gestes d’un utilisateur filmé par une caméra ou équipé de capteurs de mouvement. Ce domaine ouvre de nouvelles perspectives dans les interactions homme-machine et en création artistique, comme en théâtre, en danse contemporaine, dans le cinéma et dans la musique vivante. Autrefois réservé à des productions professionnelles en raison des coûts élevés des technologies de capture de gestes, un réel essor du domaine a été favorisé par l’arrivée de dispositifs peu onéreux et performants comme les caméras de type Kinect, introduite en 2010³.

Les vidéos obtenues ont 4 canaux : les 3 canaux de couleur standard (RVB) ainsi qu’un canal de profondeur (obtenu par une caméra infrarouge) qui facilite l’interprétation des gestes en 3D.

La plupart des approches de reconnaissance de gestes décrites dans la littérature font appel à des techniques d’apprentissage automatique similaires à celles utilisées en reconnaissance de la parole. L’article [12] en recense un certain nombre, des algorithmes de type filtres à particules aux populaires chaînes de Markov cachées (HMM). Depuis quelques années, le paradigme appelé *deep learning*, a permis de grandes avancées, en particulier dans le domaine de la vision par ordinateur. Ce paradigme fait appel à des réseaux de neurones dits “profonds” qui apprennent leur propre représentation des données pour profiter le plus possible de leurs capacités de discrimination [1].

Des compétitions de reconnaissance de gestes sont organisées régulièrement, comme *CHALEARN Looking at People*. En 2014, par exemple, l’une des tâches de cette compétition consistait à identifier une vingtaine de gestes typiques de locuteurs italiens filmés à l’aide d’une Kinect. Le meilleur système, fondé sur une architecture de type réseaux profonds, avait obtenu une précision de 85% [13]. Toutefois, ces approches peuvent être difficiles à mettre en œuvre en raison des très grandes quantités de données nécessaires pour entraîner ce type de modèles.

Ainsi, comme il n’existe pas de grand corpus de gestes de Soundpainting disponible, nous avons opté pour la réalisation d’un premier prototype basé sur des chaînes de Markov cachées.

5. PROTOTYPE

Pour pouvoir mettre en place un prototype de reconnaissance de gestes de soundpainting, nous avons filmé à l’aide d’une Kinect des gestes isolés et des séquences de gestes réalisés par Christophe Mangou, chef d’orchestre

3. <https://fr.wikipedia.org/wiki/Kinect> consulté le 15 mars 2016

classique et soundpainteur professionnel, qui dirige notamment l'ensemble de soundpainting *Amalgammes*⁴.

Le corpus pilote comporte une vingtaine de gestes répétés cinq fois chacun, en faisant varier la vitesse d'exécution du geste, l'amplitude des bras, etc. Un jeu de séquences de gestes typiques de performances réelles ont également été collectées. La figure 3 montre une séquence basique de quatre gestes. Il est important de collecter des séquences de plusieurs gestes car la réalisation d'un geste isolé est différente de celle du même geste effectué dans un enchaînement (similaire avec les phénomènes de co-articulation en traitement de la parole).

Le principe du prototype est le suivant. A mesure que les images RVB et de profondeur sont capturées par la caméra, nous récupérons les coordonnées spatiales des articulations du soundpainteur, qui constituent les paramètres d'entrée du système. Seules les coordonnées de six joints sont utilisées : coudes, poignets, et mains. La reconnaissance de gestes est ensuite effectuée sur des fenêtres temporelles glissantes de taille variable.

Une étape de normalisation des gestes est réalisée avant la reconnaissance proprement dite. Les coordonnées des joints sont ré-échantillonnées par interpolation à 32 valeurs à chaque nouvelle fenêtre temporelle pour avoir toujours la même taille de données en entrée du classifieur. Sont ensuite effectuées une rotation par rapport aux joints des épaules du soundpainteur ainsi qu'une translation pour centrer l'acquisition par rapport à un point de référence. Ces normalisations correspondent à une généralisation à 3 dimensions de l'algorithme "1 dollar" [17]. Enfin, une mise à l'échelle d'un cube de côté 1 est faite. Toutes ces étapes de normalisation visent à rendre le système robuste aux différences d'exécution de gestes d'un soundpainteur donné, et permettent de reconnaître les gestes de n'importe quel soundpainteur, à partir d'un corpus d'exemples collectés auprès d'un unique soundpainteur.

Chaque geste est modélisé par une chaîne de Markov cachée, qui est une approche adaptée pour la modélisation de séquences. Cette approche est similaire à celle d'un module développé à l'IRCAM [8], à la différence que notre outil utilise un nombre d'état par geste limité à 2 ou 3 états, au lieu de modéliser chaque point par un état (i.e. 32 états par fenêtre de 32 points). Les observations sont modélisées à l'aide de distributions Gaussiennes multivariées. La reconnaissance s'appuie sur l'algorithme de Viterbi.

Les performances de notre prototype n'a pas été validé quantitativement. Toutefois nous avons pu effectuer une validation informelle sur quelques gestes réalisés par différentes personnes en direct. Les quelques gestes testés sont effectivement reconnus lorsqu'ils sont correctement exécutés. En revanche, le système reconnaît plusieurs fois de suite un même geste au cours de son exécution. Il est donc nécessaire de filtrer les étiquettes de gestes générées par le système. Pour une description technique

4. <http://www.ensemble-amalgammes.fr> consulté le 15 mars 2016

plus détaillée du prototype, nous invitons le lecteur à se référer à [14].

Enfin, une vidéo démo du prototype peut être vue sur la page Web suivante : <https://vimeo.com/112062148>. Dans cette démonstration, un message du système de reconnaissance de gestes est envoyé au logiciel Max-MSP à l'aide du protocole UDP pour déclencher une note de piano ou la lecture d'un fichier audio.

D'autres limites du système actuel concernent le panel de gestes potentiellement reconnaissables. Nous avons ainsi privilégié l'étude des gestes utilisant les bras plutôt que les doigts, dont les mouvements ne sont pas détectés par notre dispositif. De plus, notre système ne peut à l'heure actuelle reconnaître des paramètres intrinsèques à certains gestes. Par exemple, le geste *Volume Fader* (VF), grâce auquel le soundpainteur demande une modification du volume sonore en utilisant la position de sa main par rapport à celle de son bras, est identifié globalement par une étiquette "VF" mais la variation de nuance demandée n'est pas détectée (cela nécessiterait l'information de la position relative de la main vis-à-vis de l'avant-bras). Par ailleurs, notre corpus ne comporte pas de gestes se référant à la position des membres de l'orchestre, comme par exemple le geste *scanner*.

6. APPLICATIONS

6.1. Annotation automatique

Utilisé lors d'une performance de soundpainting, notre système vise à obtenir une indexation automatique des gestes produits par le soundpainteur. Cette approche offre de nombreuses applications, que ce soit au niveau de la reconnaissance d'un geste simple, ou de celui de la transcription de performances entières.

Lors d'une improvisation, le soundpainteur doit réaliser des signes établis de la manière la plus précise possible. Utiliser un outil d'analyse automatique de gestes fondé sur la modélisation du squelette humain peut donc permettre de comparer le même signe effectué par plusieurs soundpainteurs. Notre outil pourrait donc être utilisé dans un but pédagogique dans le cadre de l'apprentissage des gestes du soundpainting, ou pour évaluer, en temps réel, la réalisation de geste chez des praticiens expérimentés.

En parallèle de son utilisation en temps réel pour détecter les gestes du soundpainteur, notre prototype peut être utilisé pour enregistrer le déroulé d'un ensemble de gestes. Ce système pourrait produire à chaque utilisation un fichier xml⁵, organisé hiérarchiquement par séquences de gestes, et conservant l'information temporelle de chaque évènement. Ce fichier pourra par la suite être représenté graphiquement sous la forme de séquence de gestes successives (voir figure 1).

Au niveau de la performance, l'annotation de tous les signes utilisés pourrait compléter les captations audio et vidéos utilisées lors de l'analyse d'une séquence de

5. Extensible Markup Language

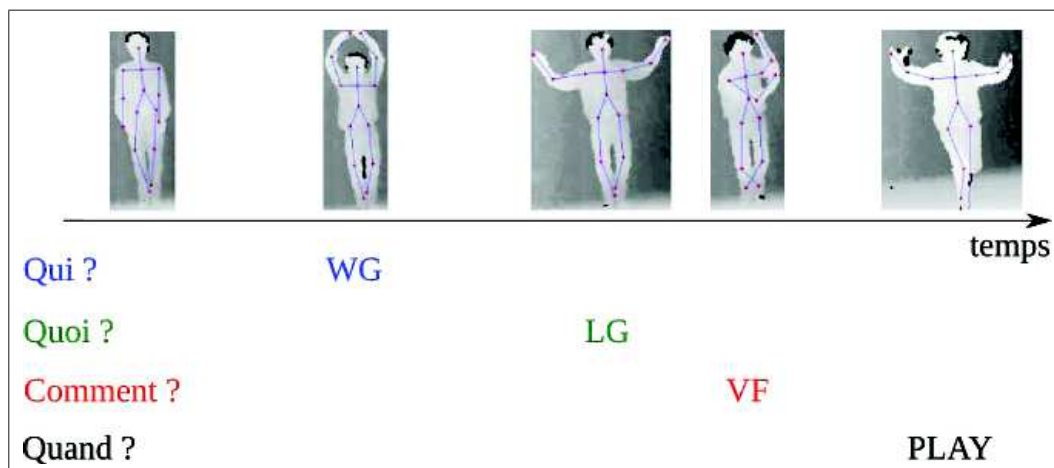


Figure 3. Illustration du déroulement d'une séquence **Whole Group - Long Note - Volume Fader (VF) - Play** avec capture du squelette du soundpainteur. Le temps s'écoule de gauche à droite de la figure. Le soundpainteur est au repos lorsque ses bras sont baissés le long du corps (image la plus à gauche).

soundpainting. La visualisation *a posteriori* d'une performance accompagnée de signes effectués par le soundpainteur permettrait donc, telle une partition que l'on suit en écoutant une œuvre, de donner un nouvel accès à la performance en dissociant clairement les consignes données de l'interprétation des artistes. La possibilité d'obtenir facilement la partition d'une performance offre en outre la possibilité de la rejouer, ce qui peut permettre, en comparant plusieurs interprétations, d'appréhender de manière objective l'analyse d'improvisations dirigées.

Au delà de l'analyse d'une simple performance, la possibilité d'enregistrer les gestes d'un ou plusieurs soundpainteurs offre la possibilité d'analyser la pratique à plus grande échelle. Un soundpainteur chevronné pourrait ainsi analyser son travail, et en suivre l'évolution. A travers un ensemble de soundpainteurs, la notation peut permettre d'analyser les récurrences de gestes ou de séquences de gestes. Nous pourrions ainsi identifier une syntaxe récurrente propre à certains soundpainteurs, en utilisant par exemple des outils de traitement automatique des langues.

6.2. Autres applications

6.2.1. Ordinateur musicien

Au delà de l'annotation d'une performance, la reconnaissance des gestes du soundpainteur offre de nombreuses possibilités d'expérimentation artistique. Une application immédiate du temps réel consiste à générer des sons à partir des gestes reconnus⁶. Le dispositif de création de sons pourrait être dans un premier temps uniquement piloté par la reconnaissance de signes. Dans un deuxième temps, notre ordinateur musicien pourrait, en plus des gestes, utiliser également son environnement sonore et les réalisations des autres musiciens. Le développement d'un tel outil offrirait, à travers la

réalisation de performances de musique mixte [7], un outil formidable pour analyser et formaliser les concepts liés à l'improvisation.

6.2.2. Composition sonore

Des travaux récents de Renaud Félix explorent les possibilités offertes par un outil informatique de composition basé sur le soundpainting, qui simule les réponses sonores d'un orchestre en fonction des indications données par les utilisateurs⁷. Ces expériences prometteuses s'appuient à l'heure actuelle sur une reconnaissance des consignes par transcription vocale. L'utilisation d'outils de reconnaissance de gestes, qui contrôlerait des modules de synthèse sonore (par exemple au format VST⁸), pourrait permettre la mise à disposition de nouveaux outils de composition sonore s'appuyant sur la syntaxe du soundpainting.

Par ailleurs, un intérêt indéniable du soundpainting réside dans son potentiel pédagogique. Il permet par exemple, sans partitions, d'obtenir de musiciens n'ayant jamais joué ensemble des résultats musicaux intéressants. Les musiciens amateurs pourraient, sous réserve d'une certaine maîtrise des gestes du soundpainting, s'essayer à la composition en dirigeant un ensemble.

7. DISCUSSION

Le soundpainting constitue un terrain très favorable à l'analyse de performances musicales interactives, puisque cette discipline associe la rigidité d'un code fixe à une interprétation très libre des performeurs. Traditionnellement plutôt fondée sur une transmission orale, la pratique du soundpainting peut être formalisée à l'aide d'un support écrit. Certaines approches d'annotation manuelle,

7. <http://www.virtual-soundpainting.com/> consulté le 15 mars 2016

8. Steinberg, A. G. (2005). "Virtual studio technology. <http://ygrabit.steinberg.de> consulté le 15 mars 2016

6. <https://vimeo.com/112062148>

comme celles présentées brièvement ici, pourraient inspirer la manière dont un système automatique devrait réaliser cette tâche.

Nous avons présenté dans cet article un prototype de transcription automatique de performances de soundpainting sous la forme d'un outil de reconnaissance de gestes. Ce premier prototype, fondé sur des modèles de Markov cachés, a été développé pour reconnaître les gestes d'un soundpainteur filmé par une caméra RVB / profondeur de type Kinect. Un corpus pilote de vidéos de gestes réalisés par un soundpainteur professionnel a été constitué pour entraîner les modèles, avec un modèle par geste cible.

L'identification de simples étiquettes des gestes est une première étape qui devra être complétée pour caractériser les gestes avec une granularité plus fine, indispensable dans la mesure où l'annotation de certains gestes requiert des informations sur la position relative des membres du soundpainteur (par exemple le geste *Volume Fader*). Une autre limite du prototype actuel concerne les gestes qui requièrent une détection fine des doigts de la main, comme les gestes servant à désigner un nombre de musiciens donné par exemple, ce qui constitue un problème de recherche actuel en vision par ordinateur. Par ailleurs, certains gestes utilisent des informations spatiales de disposition des musiciens qu'il n'est pas aisé de retranscrire (par exemple pointer un musicien sur la scène). Enfin, certains gestes sont très relatifs à l'exécution des musiciens, comme par exemple l'augmentation du volume sonore. D'un point de vue général, il semble clair que l'annotation des gestes ne représente qu'une partie de la transcription d'une performance, dans laquelle les musiciens improvisent en fonction des gestes, et par ce procédé, influencent le soundpainteur.

Au final, la transcription des gestes, accompagnée éventuellement d'audio et de vidéo, présente un grand intérêt pour l'analyse de pratique de soundpainting. Au delà de la simple transcription, l'analyse de la pratique du soundpainting peut permettre de mieux comprendre et documenter les échanges entre les acteurs d'une performance interactive. L'analyse des gestes pourrait aussi trouver des applications dépassant l'analyse de pratique, avec son utilisation dans le cadre de la musique mixte.

8. REFERENCES

- [1] I. Arel, D. Rose, and T. Karnowski. Deep machine learning—a new frontier in artificial intelligence research [research frontier]. *Computational Intelligence Magazine, IEEE*, 5(4):13–18, 2010.
- [2] J. Bonnerave. Improviser ensemble. de l'interaction à l'écologie sonore. *Tracés. Revue de Sciences humaines*, (18):87–103, 2010.
- [3] M. Chemillier. Editorial de musimédiane (numéro 3). <http://www.musimediane.com>, 2008. [en ligne ; consulté le 13 janvier 2016].
- [4] M. Chemillier. L'improvisation musicale et l'ordinateur. transcrire la musique à l'ère de l'image animée. *Terrain. Revue d'ethnologie de l'Europe*, (53):66–83, 2009.
- [5] S. De Laubier, H. Genevois, L. Feugere, G. Bertrand, V. Goudard, S. Le Beux, B. Doval, and C. D'Alessandro. Orjo et la méta-mallette 4.0. In *Actes des Journées d'Informatique Musicale (JIM)*, page 227, Mons, 2012.
- [6] M. Duby. *Soundpainting as a system for the collaborative creation of music in performance*. PhD thesis, University of Pretoria, 2006.
- [7] J.-M. Echeveste. *Un langage de programmation pour composer l'interaction musicale : la gestion du temps et des événements dans Antescofo*. PhD thesis, Thèse de doctorat. Université Pierre et Marie Curie-Paris VI, 2015.
- [8] F. F. Bevilacqua, F. Baschet, and S. Lemouton. The augmented string quartet : experiments and gesture following. *Journal of New Music Research*, 41(1):103–119, 2012.
- [9] B. Faria. Exploring points of (im)balance through artistic research transactions. In *Perspectives in Music and Music Education, Beyond Methods*, pages 49–64. Malmö, 2015.
- [10] H. J. Minors. Music and movement in dialogue : exploring gesture in soundpainting. *Les Cahiers de la Société québécoise de recherche en musique*, 13(1-2):87–96, 2012.
- [11] H. J. Minors. Music and movement in dialogue : exploring gesture in soundpainting. *Les Cahiers de la Société québécoise de recherche en musique*, 13:1-2:87–96, 2012.
- [12] S. Mitra and T. Acharya. Gesture Recognition : A survey. *IEEE Transactions on systems, man, and Cybernetics - Part C : applications and reviews*, 37:3:311–324, 2007.
- [13] N. Neverova, C. Wolf, W. Graham, F. Taylor, and F. Nebout. Moddrop : adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.
- [14] T. Pellegrini, P. Guyot, B. Angles, C. Mollaret, and C. Mangou. Towards soundpainting gesture recognition. In *Proceedings of the 9th Audio Mostly : A Conference on Interaction With Sound*, Aalborg, 2014.
- [15] G. Rouget et al. *Un roi africain et sa musique de cour : chants et danses du palais à Porto-Novo sous le règne de Gbèfa (1948-1976)*. CNRS éditions, 1996.
- [16] W. Thompson. *Soundpainting : the art of live composition. Workbook I*. Walter Thompson, 2006.
- [17] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training : a \$1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 159–168. ACM, 2007.