

Contrôle gestuel continu de médias sonores et visuels

Frederic Bevilacqua, Bruno Zamborlin, Nicolas Rasamimanana, Norbert Schnell

Ircam - CNRS STMS
Equipe Interactions Musicales Temps Réel
1, place Igor Stravinsky
75004, Paris, France
{Frederic.Bevilacqua,Bruno.Zamborlin, Nicolas.Rasamimanana, Norbert.Schnell}@ircam.fr

RESUME

Nous présentons un système appelé "gesture follower" qui permet de reconnaître et de synchroniser des gestes continus avec des contenus numériques. Ce système peut être utilisé avec n'importe quel flux continu de données gestuelles, provenant par exemple d'une tablette graphique, de capteurs inertiels, d'analyse d'image ou de descripteurs sonores. Le "gesture follower" fournit en continu des mesures de similarité entre un geste et des modèles enregistrés, ainsi que la progression temporelle du geste en cours d'exécution. Ces paramètres permettent en particulier de reconnaître des gestes avant leur complétion, et donc d'anticiper, dans une certaine mesure, l'action de l'utilisateur. Plusieurs cas d'utilisation ont été développés dans des contextes artistiques ou de pédagogie musicale où ce système a été utilisé pour contrôler en continu des contenus sonores et visuels.

MOTS CLES : geste, interaction, musique, synchronisation, suivi

INTRODUCTION

Le rôle du contrôle gestuel dans les systèmes musicaux interactifs s'est fortement développé au cours de ces dix dernières années, à la fois à un niveau technologique et à un niveau conceptuel. D'une manière générale, le contrôle gestuel et tactile s'est popularisé grâce à la disponibilité accrue de capteurs et d'interfaces, et une communauté de chercheurs s'est cristallisée autour des problématiques musicales. En particulier de nombreux travaux sont régulièrement présentés aux conférences NIME (New Interfaces pour expressions musicales) depuis 2001. De plus, des recherches plus fondamentales réalisées sur le geste humain et l'interaction gestuelle, par exemple portant sur le couplage sensori-moteur, l'énaction ou encore "l'embodied interaction" [6], ont fortement influencé les recherches sur les interfaces musicales et le geste instrumental ([5, 9, 8, 10, 14, 7]). L'ensemble de ces développements technologiques et théoriques entraînent naturellement le développement de nouvelles approches pour le contrôle de médias numériques, où les gestes physiques occupent une place centrale.

Nous nous focalisons dans cet article sur un aspect parti-

culier, à savoir le contrôle continu à partir d'une analyse temps réel des données de captation. En particulier nous avons développé une approche basée sur l'apprentissage automatique, permettant entre autre d'effectuer de la reconnaissance de geste.

Dans la plupart des approches classiques de reconnaissance, les gestes sont caractérisés ou reconnus après leur complétion. Nous présentons ici un paradigme différent, motivé par nos applications sur le contrôle expressif de médias numériques : le système d'analyse fournit de manière continue des informations relatives au geste en cours d'exécution, en fonction d'une base de données fixée par l'utilisateur.

Précisément, deux types d'information sont calculés en temps réel. Premièrement, une mesure de similarité est estimée entre le geste et des exemples enregistrés préalablement. Ces mesures de similarité peuvent être utilisées pour une tâche de reconnaissance. Deuxièmement, la progression temporelle du geste est également évaluée, sur la base d'exemples pré-enregistrés.

Ces données d'analyse du geste sont particulièrement adaptées à la fois pour sélectionner et synchroniser la lecture de fichiers son et vidéo avec un geste "continu". Nous avons développé depuis quelques années ce type d'applications dans divers contextes artistiques et de pédagogie musicale, en utilisant soit un geste libre soit des actions sur des interfaces tangibles (entre autre dans le cadre du projet ANR Interlude). Cet article décrit succinctement le système d'analyse et un tel paradigme d'utilisation.

SUIVI ET RECONNAISSANCE DE GESTES

Nous considérons ici des gestes "continus" qui peuvent être représentés par des profils temporels échantillonnés à une fréquence constante. Le système que nous avons développé, appelé "gesture follower" [3, 1] permet précisément l'analyse de profils temporels multidimensionnels en temps réel. Aucune hypothèse n'est faite sur la nature des données, rendant le "gesture follower" compatible avec tout système de captation fournissant des données continues qu'elles soient issues de capteurs, d'analyse d'image, de tablette ou même de descripteurs sonores.

Comme de nombreux systèmes de reconnaissance de geste [11], l'algorithme est basé sur des Modèles de Markov Cachés (HMM) [13]. Cependant, contrairement aux approches classiques, nous avons établi une procédure simplifiée d'apprentissage, permettant aux utilisateurs de créer facilement une base de données à partir d'un nombre limité d'exemples (ou même d'un exemple unique dans certain cas).

Afin de permettre un calcul incrémental en temps réel, le décodage est basé sur la procédure "forward" [13], de manière similaire aux algorithmes développés pour le suivi de partition [18]. Comme illustré sur la Figure 1 le "gesture follower" permet d'aligner en temps réel un profil en cours d'exécution avec un profil de référence.

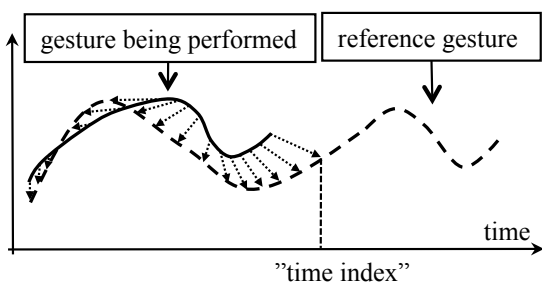


FIGURE 1 : Alignement temporel entre deux gestes, effectués en temps réel par le "gesture follower"

Cet alignement permet donc de fournir les deux types de quantités mentionnées dans l'introduction, à savoir :

- *time index*, correspondant à l'estimation de la progression temporelle du geste
- *likelihood* correspondant à des mesures de similarité avec les gestes enregistrés

Ces données permettent donc à tout moment d'inférer quel geste est en train d'être exécuté, ainsi que son degré de complétion (i.e. *time index*). L'évolution de ces inférences au cours du temps fournissent la possibilité d'anticiper sur l'évolution du geste.

IMPLEMENTATIONS

Le "gesture follower" est développé comme une bibliothèque C++ et il peut donc être mis en œuvre dans divers environnements. Pour nos applications, nous avons implémenté l'algorithme comme un objet externe dans l'environnement MaxMSP, appelé *gf*. L'intégration avec le langage de script Lua a également été réalisée.

De plus, nous avons développé d'autres bibliothèques au cours de ces dernières années pour manipuler en temps réel des données multimodales et contrôler de la synthèse sonore basée sur des enregistrements (FTM&Co [16, 12, 2], MuBu [17]). L'ensemble de ces modules nous permet de réaliser des prototypes complets de systèmes d'interactions musicales.

EXEMPLE : CONTROLE GESTUEL DE SONS ENREGISTRÉS

Plusieurs paradigmes d'interactions peuvent être utilisés avec le "gesture follower". Nous décrivons ici un paradigme possible que nous avons utilisé dans des contextes variés (performance musicale, danse, installation, pédagogie). L'idée générale est de synchroniser la lecture d'un fichier son ou vidéo avec la progression temporelle du geste. Le rendu sonore est généralement effectué avec un vocodeur de phase ou de la synthèse granulaire, afin de pouvoir varier la vitesse de lecture d'un enregistrement sonore sans en altérer la hauteur. La Figure 2 illustre ce principe où une phase d'apprentissage permet à la fois de définir le geste et la synchronisation.

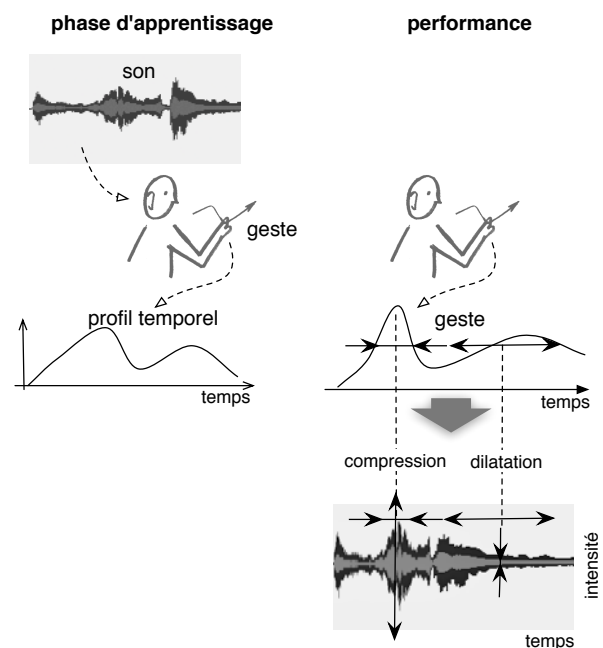


FIGURE 2 : Synchronisation d'un geste avec un son, suite à une phase d'enregistrement

1. Mode d'enregistrement : Enregistrement d'un geste en écoutant un fichier son. Cette étape fournit un geste de référence qui est synchronisé avec le fichier son. Le geste peut être librement choisi par l'utilisateur, en fonction de l'exemple sonore. A noter que le son peut également provenir d'une captation sonore, par exemple la voix de l'utilisateur ou un instrument de musique.
2. Mode de performance/contrôle : L'utilisateur répète le même geste, avec une certaine variation de vitesse et d'amplitude. Grâce au "gesture follower", le son synthétisé varie en fonction des variations temporelles et d'amplitude dans l'exécution du geste. Par exemple, le *time index* affecte la vitesse de lecture du son et la mesure de similarité (*likelihood*) est utilisé pour varier l'intensité du son (faible similarité = faible énergie sonore).

Cette procédure peut être répétée plusieurs fois, en utilisant des fichiers sons différents. Précisément, chaque geste peut être associé à un son particulier. La fonction de reconnaissance peut être utilisée pour sélectionner un son unique correspondant au geste le plus probable. Un autre choix consiste à jouer les différents sons simultanément en les mixant en fonction des valeurs de similarités.

Ce paradigme d'interaction peut être utilisé pour simuler la direction d'orchestre ou le jeu virtuel d'instrument de musique, ce que nous avons évalué dans le cadre d'applications de pédagogie musicale [1, 15]. Comparativement à des applications similaires [4] notre approche se distingue sur plusieurs points. Tout d'abord, le geste est considéré ici comme un processus continu. En particulier, aucune détection de "beat" n'est nécessaire. Deuxièmement, le choix du geste est totalement ouvert et peut être choisi avec une procédure simple. Comme mentionné précédemment, un seul enregistrement d'un geste est suffisant pour utiliser le système.

CONCLUSION ET PERSPECTIVE

Nous avons décrit dans cet article une approche basée sur l'apprentissage automatique pour caractériser en temps réel des gestes continus. Nous poursuivons actuellement une évaluation plus systématique de notre système et nous envisageons d'élargir les champs d'applications. Il semble en effet que cette approche peut se révéler intéressante pour d'autres cas d'utilisation dans le cadre d'Interface Homme Machine.

REMERCIEMENTS

La recherche présentée dans cet article a été en partie soutenue par les projets suivants : EU-ICT project : SAME, projets ANR : Interlude, EarToy et SampleOrchestrator.

BIBLIOGRAPHIE

1. Bevilacqua, F., Guedy, F., Fléty, E., Leroy, N., and Schnell, N. Wireless sensor interface and gesture-follower for music pedagogy. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2007.
2. Bevilacqua, F., Muller, R., and Schnell, N. Mnm : a max/msp mapping toolbox. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 85–88, 2005.
3. Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guedy, F., and Rasamimanana, N. Continuous realtime gesture following and recognition. In *LNAI 5934*, pages 73–84. Springer Verlag, 2010.
4. Borchers, J., Hadjakos, A., and Muhlhauser, M. Micon : A music stand for interactive conducting. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 254–259, 2006.
5. Cadoz, C., and Wanderley, M. *Trends in Gestural Control of Music*, chapter Gesture - Music. Ircam - Centre Pompidou, 2000.
6. Dourish, P. *Where the action is : the foundations of embodied interaction*. MIT Press, Cambridge, MA, USA, 2001.
7. Godøy, R., and Leman, M., editors. *Musical Gestures : Sound, Movement and Meaning*. Routledge, 2009.
8. Godøy, R.I., Haga, E., and Jensenius, A. R. Exploring music-related gestures by sound-tracing - a preliminary study. In *2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems*, Leeds, UK, 2006.
9. Jensenius, A. R. *Action-sound : Developing methods and tools to study music-related body movement*. PhD thesis, University of Oslo, Department of Musicology, Oslo, Norway, 2007.
10. Leman, Marc. *Embodied Music Cognition and Mediation Technology*. Massachusetts Institute of Technology Press, Cambridge, USA, 2008.
11. Mitra, S., Acharya, T., Member, S., and Member, S. Gesture recognition : A survey. *IEEE Transactions on Systems, Man and Cybernetics - Part C*, 37 :311–324, 2007.
12. N. Schnell et al. Gabor, Multi-Representation Real-Time Analysis/Synthesis. In *DAFx*, Septembre 2005.
13. Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.
14. Rasamimanana, N. *Geste instrumentale du violoniste en situation de jeu : analyse et modélisation*. PhD thesis, Université Paris 6 - IRCAM UMR STMS, 2008.
15. Rasamimanana, N., Guedy, F., Schnell, N., Lambert, J.-P., and Bevilacqua, F. Three pedagogical scenarios using the sound and gesture lab. In *Proceedings of the 4th i-Maestro Workshop on Technology Enhanced Music Education*, 2008.
16. Schnell, N., Borghesi, R., Schwarz, D., Bevilacqua, F., and Müller, R. Ftm - complex data structures for max. In *Proceedings of the International Computer Music Conference (ICMC)*, 2005.
17. Schnell, N., Röbel, A., Schwarz, D., Peeters, G., and Borghesi, R. Mubu and friends : Assembling tools for content based real-time interactive audio processing in max/msp. In *Proceedings of the International Computer Music Conference (ICMC)*, Montreal, Août 2009.
18. Schwarz, D., Orio, N., and Schnell, N. Robust polyphonic midi score following with hidden markov models. In *Proceedings of the International Computer Music Conference (ICMC)*, 2004.