HAL open science

Notation et reconnaissance des actions scéniques par ordinateur

Rémi Ronfard

To cite this version:

Rémi Ronfard. Notation et reconnaissance des actions scéniques par ordinateur. La notation du travail théâtral : du manuscrit au numérique, Monique Martinez, Sophie Proust, Matthieu Pouget, Apr 2012, Toulouse, France. hal-00768893

HAL Id: hal-00768893

https://inria.hal.science/hal-00768893v1

Submitted on 26 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Notation et reconnaissance des actions scéniques par ordinateur

Rémi Ronfard (INRIA, Université de Grenoble)

Résumé : Dans ce chapitre, nous abordons la question de la notation de la mise en scène sous l'angle de la théorie de la notation de Goodman, qui fixe des contraintes syntaxiques et sémantiques beaucoup plus strictes qu'il n'est de coutume dans le domaine des études théâtrales. Ceci nous amène à proposer un schéma de notation restreint à un petit nombre d'actions scéniques reconnaissables visuellement, pouvant faire l'objet d'une transcription automatique par ordinateur à partir d'enregistrements vidéo. Nous terminons ce chapitre en décrivant deux projets de recherche en cours visant à démontrer expérimentalement la possibilité d'une telle transcription et son utilité pour analyser et documenter le travail des répétitions théâtrales.

Introduction

La question de la notation de la mise en scène est soulevée régulièrement dans le monde du théâtre. En particulier, Patrice Pavis s'est interrogé avec brio sur les différents usages et les différentes problématiques ouvertes par cette question complexe[1], et a introduit une distinction importante entre la sémiologie du théâtre, qui est un discours sur le théâtre, et la sémiotique du théâtre, qui en est une théorie formelle, qu'il appelait de ses voeux. Cependant, Pavis ne fait pas référence à la théorie de la notation de Goodman, qui pourtant apporte un éclairage utile à cette question, et permet d'éviter les deux écueils de la sémiologie et de la sémiotique, pour aborder directement la question qui me paraît centrale, celle de la partition scénique.

Trente ans après l'étude classique de Pavis, la question de la notation théâtrale peut être revisitée à la lumière de plusieurs évolutions récentes. : (1) la démocratisation des caméras vidéo numériques, qui permettent de réaliser des enregistrements de spectacles de grande qualité technique et à moindre coût ; (2) l'informatisation des systèmes de régie son et lumière, qui représentent le spectacle sous la forme d'une partition électronique exécutable.

Ces évolutions expliquent l'intérêt suscité par cette question chez les informaticiens et les chercheurs[2]. En effet, si les représentations peuvent être numérisées, il devient possible d'analyser, indexer et comparer ces enregistrements pour en faire le support d'une analyse critique, comme cela commence à se faire pour le cinéma[3] [4]. Et si les sons et les lumières peuvent être inscrits dans un programme exécutable, il paraît également possible d'imaginer une partition scénique plus générale

qui permettrait de retranscrire et exécuter toute la mise en scène.

Ce n'est donc pas un hasard si nous traduisons « notation » par « partition » dans ce chapitre. Nous le faisons en nous appuyant sur la théorie de la notation de Goodman, et dans l'objectif de démontrer (1) qu'il est possible d'établir la partition d'un spectacle en adoptant le point de vue et le vocabulaire du metteur en scène ; (2) que les techniques numériques d'enregistrement et d'analyse audio et vidéo permettent d'aborder cette question selon une approche expérimentale et scientifique, permettant la comparaison, la validation et la réfutation des systèmes de notation; (3) qu'une telle notation est souhaitable pour favoriser l'intégration des nouvelles technologies dans le théâtre, et du théâtre dans les nouvelles technologies.

Historiquement, l'enjeu principal de la notation théâtrale a d'abord été de conserver une trace de la mise en scène. Aujourd'hui, les captations vidéo jouent ce rôle. Peut-on parler pour autant d'une « notation vidéo » du travail théâtral ? A notre avis, ce serait une erreur, même si l'enregistrement numérique des spectacles peut servir de support à la notation informatique. La vidéo n'est d'ailleurs qu'une des nombreuses techniques de numérisation possibles – les autres étant la capture de mouvement, la localisation des acteurs par ondes radio ou sonore, etc. La pratique de l'enregistrement vidéo libère la notation de la tentation de « tout noter » puisqu'on dispose en dernier recours de l'enregistrement. En revanche, elle pose la question de « quoi noter », de quel point de vue et sous quelle forme.

Qu'il s'agisse de numériser ou de noter le théâtre, il est important de choisir un point de vue - celui du public ou des acteurs - et une syntaxe - la langue naturelle ou un schéma formel de notation. Autrement dit : où placer la caméra et comment décrire ce que l'on voit . Ces deux problématiques concernent généralement tous les arts vivants. Dans le domaine de la danse, différents systèmes de notation se partagent les deux points de vue du danseur (Laban[5], Benech[6], Eskhol-Wachman[7]) ou du spectateur (Sutton[8]). Dans le domaine du théâtre, il nous faut également choisir entre les points de vue symétriques de l'acteur et du spectateur.

Un autre choix important concerne la syntaxe de la notation. La notation chorégraphique des mouvements d'acteurs peut être étendue au théâtre, mais est-elle adaptée à un art qui se définit principalement par rapport au texte? Même dans le domaine de la danse, l'annotation vidéo en langue naturelle offre un complément utile à la notation chorégraphique. Par exemple, des chercheurs de l'université de Stanford ont ainsi proposé un carnet vidéo d'annotation destiné aux chorégraphes[9]. Cette annotation en langue naturelle peut-elle être considérée comme une notation ? Pas si l'on se réfère à la théorie classique de la notation de Nelson Goodman, vers laquelle nous nous tournons maintenant.

Théorie de la notation

La théorie de la notation de Goodman offre une perspective utile pour comprendre les enjeux et les limites de la notation théâtrale. Cette théorie est exposée dans les chapitres 4 et 5 de Langages de l'art [10]. Nous en rappelons ici quelques éléments.

La théorie de la notation de Goodman permet de distinguer trois catégories d'objets esthétiques, qui entretiennent des rapports différents à la notion d'œuvre d'art – la partition, l'esquisse et le script. Goodman souligne ainsi les différences entre ces trois notions : « Une partition musicale fonctionne dans une notation et définit une œuvre ; une esquisse ou image ne fonctionne pas dans une notation, mais elle est elle-même une œuvre ; et un script littéraire à la fois fonctionne dans une notation et est lui-même une œuvre ». Concernant la notation de la mise en scène, la question est donc posée de savoir si on souhaite en faire une partition, qui définirait la mise en scène comme une œuvre à part entière, reconnaissable dans chacune de ses représentations ; ou une esquisse, qui serait elle-même une œuvre originale, autographique et non reproductible ; ou un script, qui serait lui-même une œuvre allographique et reproductible, mais indépendante de ses représentations.

Notre point de vue dans ce chapitre consiste à favoriser la première interprétation, et à nous interroger sur la possibilité de constituer une partition scénique définissant une œuvre de mise en scène, selon un schéma de notation au sens de Goodman. En effet, une notation qui ne constitue qu'une esquisse ou un script ne pose guère de problème théorique. Dans chaque cas, elle constitue un nouvel objet d'étude, plutôt qu'un outil d'analyse de « l'œuvre de mise en scène ».

Goodman propose dans sa théorie de la notation cinq conditions nécessaires pour qu'un langage constitue une partition. Les deux premières conditions portent sur la syntaxe du langage, les trois dernières portent sur la sémantique du langage.

Sur le plan syntaxique, le langage doit être composé de caractères disjoints et articulés. Ensemble, ces deux conditions permettent que chaque copie de la partition puisse recevoir la même interprétation. Deux caractères sont disjoints si l'ensemble de leurs inscriptions ont une intersection vide, c'est à dire qu'il n'existe pas d'inscription appartenant aux deux caractères simultanément. Pour un langage constitué d'un nombre fini de caractères discrets, il est possible d'attribuer chaque inscription à un caractère unique, et donc de « déchiffrer la partition » de façon non ambiguë, en examinant toutes les interprétations possibles. Pour un langage constitué d'un nombre infini de caractères disjoints, Goodman impose une seconde condition, qui est l'articulation, ou différenciation finie. Pour chaque inscription, il doit être possible de trancher entre deux interprétations possibles. En théorie, cette seconde condition permet de « déchiffrer » la

partition même si elle utilise un vocabulaire infini de signes.

Sur le plan sémantique, Goodman introduit la notion de concordance, qui définit la relation entre le schéma symbolique utilisé comme notation et le domaine de référence qui peut être noté. A chaque caractère de la notation correspond une *classe de concordance* qui est l'ensemble de toutes ses exécutions possibles. Les conditions sémantiques sont celles qui permettent de *transcrire la partition* à partir d'une seule exécution de l'oeuvre. Plus précisément, un système de notation ne peut constituer une partition que si les classes de concordance de ses caractères sont invariantes, disjointes et articulées. Les propriétés d'invariance et de disjointure permettent d'exclure la possibilité d'exécutions ambigues, comme par exemple une note de musique qui serait parfois un do et parfois un fa (invariance) ou en même temps un do et un fa (disjointure). La propriété d'articulation (ou différenciation finie) permet de distinguer les différents caractères de la partition selon une séquence finie d'opérations, par exemple en comparant la note entendue avec un do et un fa, de façon à rejeter catégoriquement au moins l'une de ces deux interprétations.

Dans la terminologie de Goodman, les conditions syntaxiques définissent un système notationnel, qui garantit seulement que la partition peut être reproduite en conservant ses qualités notationnelles – ainsi, chaque copie de la partition peut être déchiffrée en conservant la même « interprétation ». Si les conditions sémantiques sont également respectées, on obtient un schéma notationnel, qui garantit également que chaque exécution de l'œuvre peut être utilisée pour reconstituer la partition – en d'autres termes, chaque exécution permet de « reconnaître l'œuvre ».

Est-il possible de constituer une notation de la mise en scène vérifiant les conditions syntaxiques et sémantiques suggérées par Goodman, et qui permettrait - au moins en théorie - de transcrire la représentation théâtrale et d'en déchiffrer la partition ? Avant d'esquisser notre réponse dans la suite de ce chapitre, nous pouvons déjà exclure un certain nombre de techniques de notation qui ne vérifient pas ces contraintes.

La langue naturelle, qui est polysémique et ambiguë, ne peut certainement pas servir de schéma de notation. En particulier, les mots de la langue naturelle permettent de nommer des classes d'objets et d'événements à plusieurs niveaux d'abstraction, du plus général au plus particulier. Cela n'est pas possible dans un schéma de notation, qui doit nécessairement faire le choix d'un niveau de description et s'y tenir. De même, la notation musicale doit-elle choisir entre la basse chiffrée (qui indique les accords) ou la notation moderne (qui indique chaque note). La partition ne peut alterner entre les deux niveaux car elle ne pourrait pas être reconstituée à partir d'une seule exécution de l'oeuvre.

L'esquisse d'un tableau n'est pas une notation du tableau car il n'est pas possible en général de reconstituer l'esquisse à partir du tableau, ni le tableau à partir de l'esquisse. De la même façon, le storyboard d'un film, qui peut être assimilé à une série d'esquisses, ne constitue pas non plus une notation du film. Les deux films intitulés « Psychose » et réalisés à 40 ans d'intervalle par Alfred Hitchcock et Gus Van Dam sont basé sur le même scénario et le même storyboard. Pourtant, personne ne considérerait qu'il s'agit du même film. Pour cette raison, les arts plastiques et le cinéma sont des arts autographiques.

L'exemple type de la notation au sens de Goodman est la partition musicale classique élaborée au 12ème siècle par Guido d'Arezzo. Celle-ci marque un tournant dans l'histoire de la musique puisque elle permet l'éclosion de la polyphonie. La partition musicale définit l'œuvre musicale, et chaque exécution de l'œuvre en est une interprétation. C'est ce qui permet de distinguer compositeur et interprète. Le compositeur signe la partition, tandis que l 'interprète exécute la partition. C'est ce qui fait de la musique classique occidentale un art allographique.

Les systèmes de notation de la danse, qui s'apparentent pour certains d'entre eux à la notation musicale, sont également des schémas de notation, au sens de Goodman, même si leur utilisation est moins systématique. L'important est que la notation de la danse soit possible en théorie, ce qui fait également de la danse classique occidentale un art allographique.

Notation du théâtre

Nous en arrivons à la question de la notation théâtrale. Quelles sont les « notes » de la mise en scène, qui permettraient d'en établir une partition ? Pour Goodman, la réponse est simple : le texte théâtral est composé d'une partie notationnelle - les dialogues - et d'une partie non-notationnelle - les didascalies, qu'il appelle actions.

Le raisonnement de Goodman est le suivant. Les dialogues sont destinés à être dits au cours de chaque représentation, et peuvent être retranscrits phonétiquement à chaque représentation. Chaque mot des dialogues est donc utilisé comme un symbole phonétique et il est facile de vérifier qu'il vérifie les postulats syntaxiques et sémantiques rappelés ci-dessus. Les dialogues constituent donc bien une partition de l'œuvre théâtrale.

Les didascalies jouent un rôle différent, puisqu'elles indiquent en langue naturelle, et de façon ambiguë, comment le texte doit être dit et quelles actions les acteurs doivent exécuter sur scène. Ces indications héritent de toutes les difficultés de la langue naturelle : elles sont en particulier

sémantiquement non disjointes et non différenciées. Les didascalies ne peuvent donc faire partie de la partition théâtrale.

La conséquence du raisonnement de Goodman est de consacrer la pièce de théâtre comme une œuvre artistique, dont chaque représentation est une interprétation. Dans cette perspective, deux interprétations d'une même pièce sont donc des exécutions de la même œuvre, quelque soit leur mise en scène. Il suffit même de prononcer les dialogues dans le bon ordre pour constituer une interprétation de la pièce, que ce soit à la radio, à la télévision, au cinéma ou sur la scène. Selon le même raisonnement, deux films réalisés à partir du même scénario, comme par exemple les deux versions de « L'homme qui en savait trop » de Hitchcock, sont des exécutions d'une même œuvre.

En revanche, Goodman ne s'attarde pas sur la mise en scène et ne se pose pas la question d'une partition scénique qui pourrait se substituer aux didascalies, en les précisant et qui permettrait de constituer la mise en scène elle-même comme œuvre artistique. C'est cette hypothèse que nous voudrions explorer dans ce chapitre.

La théorie de Goodman choisit une définition restrictive de ce qui constitue une œuvre dramatique (théâtre ou cinéma). Deux mises en scène d'un même scénario constituent simplement des interprétations différentes de « la même œuvre ». Dans le cas du film, on pourrait se contenter de répondre que chaque film constitue une œuvre autographique et non reproductible. Dans le cas du théâtre, les choses sont plus compliquées puisque chaque mise en scène donne lieu à un grand nombre de représentations différentes. A notre avis, la seule solution consiste à inventer une notation de la mise en scène.

La partition scénique

La partition scénique est composée d'une succession d'arrangement scéniques (disposition des décors et des acteurs sur la scène) et d'actions scéniques qui modifient ces arrangements au cours du temps. On peut inclure le cas limite d'une mise en scène entièrement statique, composée d'un unique arrangement scénique ou « tableau vivant » et ne comportant aucune action. Mais dans la majorité des cas, c'est l'alternance entre arrangements et actions qui compose la partition scénique.

Quelle forme peut prendre cette partition ? Historiquement, les régisseurs, metteurs en scène et assistants de mise en scène ont recours aux annotations du texte en langue naturelle, sous la forme d'un cahier de mise en scène, qui est peut être illustré également de dessins, de plans en élévation et de photographies. L'organisation spatiale et temporelle des actions scéniques est l'aspect le plus

délicat de cette annotation informelle, qui n'est pas sans rappeler le storyboard de cinéma. L'article de Pavis déjà cité contient une excellente introduction aux techniques de notation utilisées pour constituer le cahier de régie, notamment par Brecht.

Dans le domaine des sciences sociales, Poggi note les gestes et autres comportements non verbaux sous forme d'une partition musicale avec plusieurs portées par acteur[11]. Cette partition permet de représenter sur un même axe temporel les différentes composantes de l'action: verbale, prosodique, gestuelle, faciale et corporelle. Ce formalisme est particulièrement adapté à la notation de monologues et de dialogues entre deux acteurs. Sa généralisation à un plus grand nombre d'acteurs en mouvement pose des problèmes de complexité et de cohérence spatiale et temporelle qui mériteraient d'être étudiés plus en détails.

La notation hiéroglyphique, proposée par Artaud[12], a fait l'objet d'une tentative de formalisation pour l'annotation de vidéo à l'aide d'icones[13]. Cependant, la complexité de la notation proposée semble disproportionnée par rapport aux possibilités et aux besoins de la mise en scène de théâtre. De plus, les caractères de cette notation ne sont ni disjoints, ni différenciés sémantiquement.

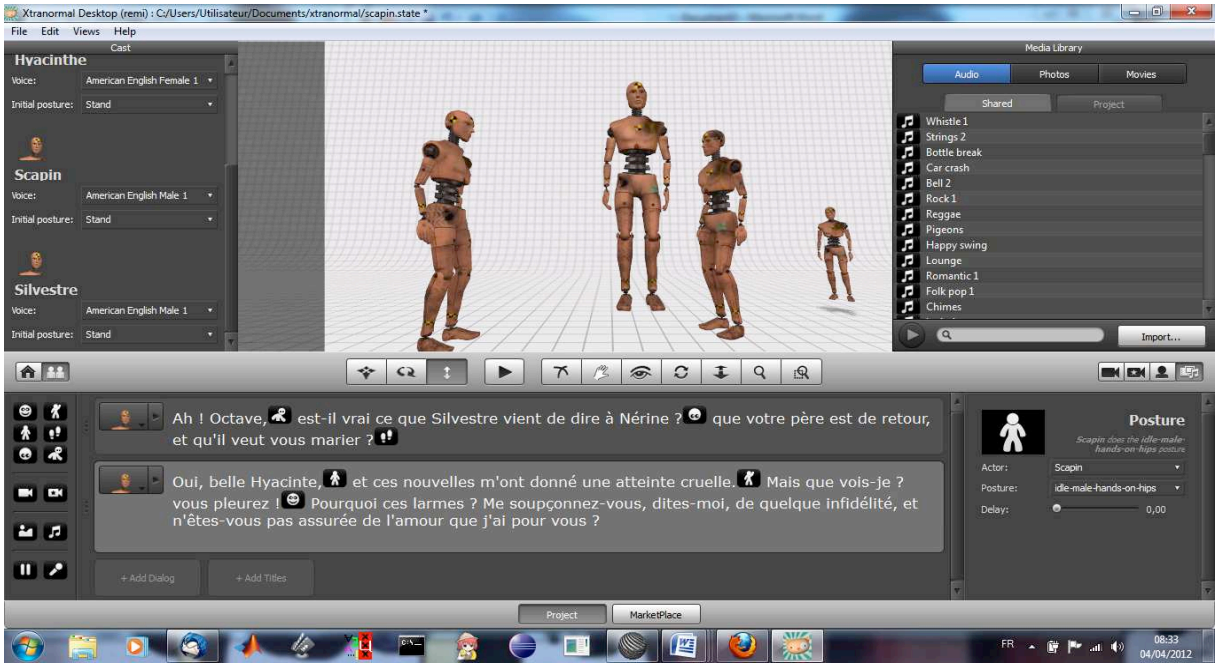

Figure 1: Partition scénique du système text-to-scene de Xtranormal. Le texte de la pièce est enrichi d'icones qui représentent les actions scéniques.

Une notation icônique est également utilisée par le site d'animation « text-to-scene » de la société américaine Xtranormal[1]. Dans ce cas, la notation fonctionne réellement comme une partition puisque les indications de mise en scène sont exécutées sous la forme d'un film d'animation 3D, comme l'illustre la Figure 1 sur l'exemple de quelques dialogues des Fourberies de Scapin de Molière joués par quatre « acteurs » virtuels. La partition des mouvements et expressions des acteurs virtuels est créée à partir du texte des dialogues et de son annotation, mais n'est pas directement manipulable. Son vocabulaire expressif est limité mais suffisant pour contrôler la disposition des acteurs, leurs mouvements, leurs expressions faciales, et leurs regards.

Dans un autre registre, le système LabanDancer développé par Calvert à l'Université Simon Fraser de Vancouver, est un synthétiseur de mouvements de danse basé sur la notation Laban[14]. Une telle partition, beaucoup plus expressive, s'avère souvent difficile à écrire et exécuter, parce qu'elle décompose le mouvement en primitives motrices, plutot que visuelles ou verbales. Si cette approche paraît adaptée à la danse, son application au théâtre paraît donc problématique. Entre ces deux extrêmes, une notation symbolique de la mise en scène théâtrale devrait concilier la référence au texte théâtral et l'expressivité de la notation du mouvement.

Les actions scéniques

Une autre difficulté à résoudre pour la notation des actions scéniques, c'est que leur répertoire (le « thésaurus » de l'acteur[15]) est de grande taille, bien au delà des sept notes de la gamme tempérée utilisée en musique. On peut envisager deux approches pour affronter cette difficulté. La première approche consiste à couvrir l'ensemble des actions scéniques avec un petit nombre de termes très généraux. La seconde approche consiste à limiter au contraire l'expressivité du langage à un petit nombre d'actions très spécifiques, à l'exclusion de toutes les autres, qui deviennent des éléments d'interprétation non-notationnels.

Comme exemple de la première approche, Pinhanez[16] propose un dictionnaire d'actions scéniques basé sur les primitives de Schank[17], dont les principaux termes sont ATTEND (l'acteur dirige son attention sur un objet); PTRANS (l'acteur se déplace) ; SPEAK (l'acteur produit des mots, sous forme de sons ou d'images); PROPEL (l'acteur met un objet en mouvement) ; GRASP (l'acteur prend un objet); MOVE (l'acteur bouge une partie de son corps); INGEST (l'acteur fait entrer un objet dans son corps) et EXPEL (l'acteur fait sortir un objet de son corps). Pinhanez obtient ainsi

[1] www.xtranormal.com

[2] http://4drepository.inrialpes.fr/public/viewgroup/6

une notation compacte des actions scéniques, mais dans des termes génériques qui ne permettent pas de noter les qualités sonores, visuelles ou rythmiques de la mise en scène. Comme exemple de la seconde approche, on peut mentionner les notations de mise en place (blocking) utilisés par les assistants de mise en scène, qui limitent la notation aux seuls mouvements d'acteurs, mais avec une plus grande précision et une meilleure prise en compte de leurs qualités sonores, visuelles et rythmiques.

Reconnaissance visuelle des actions scéniques

Parmi l'ensemble dense des actions scéniques possibles, nous avons souligné l'importance de noter en priorité les actions scéniques qui font l'objet, à la fois de dénominations non ambiguës, par exemple sous la forme de commandes de direction d'acteurs, et d'apparences visuelles reconnaissables par le spectateur. La première contrainte a pour objet de noter seulement les actions scéniques voulues par le metteur en scène, plutôt que les « accidents » de chaque représentation. La seconde contrainte est nécessaire pour obtenir un système formel de notation, au sens de Goodman.

Dans le domaine de la vision par ordinateur, la reconnaissance d'actions se donne précisément le même objectif, de distinguer si une séquence vidéo est la représentation d'une catégorie d'action donnée ou d'une autre. Ce domaine en plein essor donne lieu à un grand nombre de méthodes, qui s'appliquent à un petit nombre de catégories d'actions. On peut se référer à un état de l'art que nous avons publié récemment[18] pour recenser les méthodes, les problématiques et les perspectives de ce domaine.

Nous en arrivons au problème de la reconnaissance visuelle des actions scéniques. La question fondamentale est de choisir des actions scéniques dont les classes de concordance sont disjointes et articulées, c'est-à-dire reconnaissables. Dans le domaine de la vision par ordinateur, les capacités de reconnaissance visuelle des classes d'objets, ont connu une évolution rapide, passant de 2 classes d'objets seulement en 1992 (les visages et les voitures), à 101 classes[19] à partir de 2004, et près de 10,000 classes[20] en 2010. Dans le domaine de la reconnaissance d'actions, l'évolution est plus récente et plus lente, passant de 2 classes d'actions en 2002 (marcher et courir) à 12 classes aujourd'hui, mais on peut anticiper qu'une centaine d'actions pourront être reconnues automatiquement dans quelques années. Parmi les actions reconnaissables visuellement, on trouve en particulier les déplacements d'acteurs qui composent la mise en place des acteurs, que nous examinons maintenant.

Figure 5. Camera setup and extracted silhouettes: (Top) the action "watch clock" from the 5 different camera views. (Middle and bottom) sample actions: "cross arms", "scratch head", "sit down", "get up", "turn", "walk", "wave", "punch", "kick", and "pick up". Volumetric exemplars are mapped onto the estimated interest regions indicated by blue box.

Figure 2: Exemple de 11 actions simples reconnaissables par ordinateur [21]. INRIA Xmas Motion Acquisition Sequences (IXMAS) [2]: multiview dataset for view-invariant human action recognition.

Vocabulaire des actions scéniques

Pendant les répétitions, c'est une pratique courante de noter la mise en place (blocking) des positions et déplacements d'acteurs, sous la forme de diagrammes et de notes. Cette notation peut faire référence temporellement au texte, et spatialement à la scène. Dans les pays anglo-saxons, la notation est réalisée (à la main) par le « stage manager ». En France, la notation est réalisée par l'assistant de mise en scène. Cette pratique donne lieu a un grand nombre de variantes, mais les principes de base sont relativement stables. Parce qu'il s'agit d'une pratique éprouvée et établie, cette notation est naturellement candidate au statut de partition scénique. Mais forme-t-elle un schéma de notation valide ? Pour répondre à cette question, et en l'absence de ressources disponibles en langue française, nous nous sommes penchés sur deux propositions décrites en détails dans deux manuels anglo-saxons destinés aux « stage managers ».

L'ouvrage de McCoy propose une notation minimaliste des mouvements d'acteurs destinée au travail quotidien de l'assistant de mise en scène[22]. Le vocabulaire des actions scéniques est limitée à 8 actions : entrer, sortir, se déplacer, s'asseoir, se lever, passer derrière, passer devant et s'arrêter. Le système de notation proposée est un langage contrôlé, limité aux noms de personnages, aux 8 verbes d'actions (Enter, Exit, Cross, Sit, Stand, Circle up, Circle down, Pause) et aux 9 régions de la scène (Upstage Left/Center/Right, Middlestage Left/Center/Right et Downstage Left/Center/right).

[2] http://4drepository.inrialpes.fr/public/viewgroup/6

En dépit de son apparente simplicité, cette notation vérifie les contraintes de Goodman, et permet donc d'établir une partition (simpliste) de la mise en scène, qui peut être aisément transcrite et déchiffrée.

Dans le même registre, l'ouvrage de Schneider propose une nomenclature étendue des actions scéniques susceptibles d'être notées par l'assistant à la mise en scène (stage manager)[23]. La notation utilise un système iconique[24] dont nous avons représenté les 40 symboles d'actions les plus importants dans la figure ci-dessous.

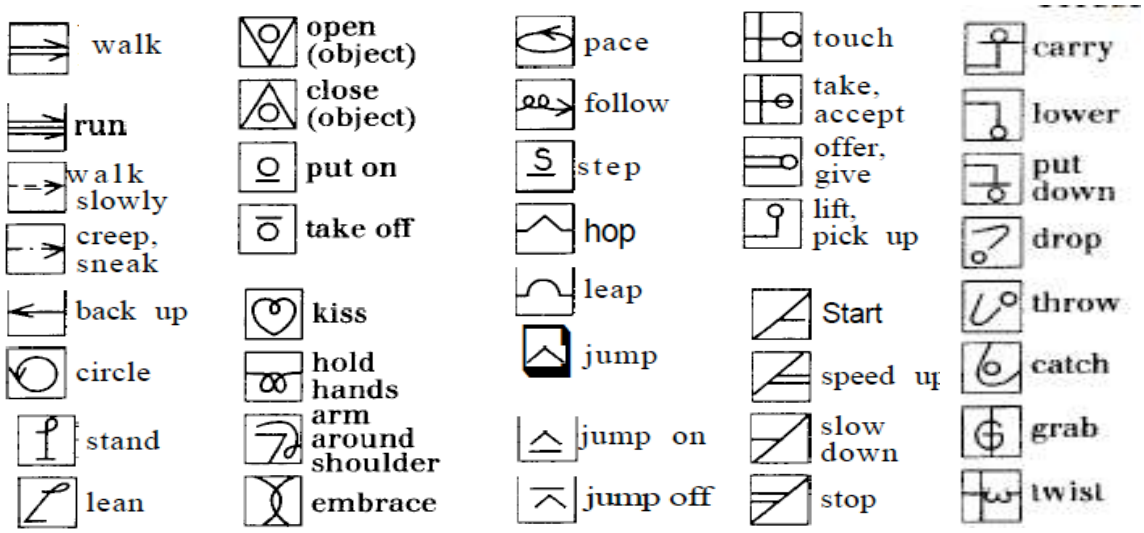

Figure 3: actions scéniques dans la notation de Schneider.

S'agit-il d'un schéma de notation au sens strict ? D'un point de vue syntaxique, cela ne semble pas faire de doute. Le déchiffrage des différents symboles est relativement aisé, et les contraintes de disjointure et d'articulation sont facilement vérifiées. D'un point de vue sémantique, cette notation ne remplit pas les contraintes de Goodman, faute de définitions précises pour chaque terme – ainsi, par exemple, les différences entre les actions walk, walk slowly et creep ne sont a priori ni disjointes, ni articulées. Cependant, on peut imaginer que le nombre de pas par minute, ou la durée des pas, soient utilisés pour fournir un critère de différenciation. De même pour les autres termes voisins (hop/leap/jump, circle/pace) pour lesquels on peut imaginer des définitions précises permettant de rendre leurs classes de concordance disjointes et articulées.

Comme dans le cas précédent, les symboles des actions sont destinés à composer des phrases simple de type sujet-verbe-objet, qui peuvent être synchronisée avec le texte de la pièce, à l'aide de systèmes de marques et de renvois, et repérés dans l'espace de la scène par l'utilisation de systèmes conventionnels de coordonnées.

Par rapport aux pratiques de la mise en scène, de telles notations apparaissent cependant rigides et contraignantes, et les assistants de mise en scène leur préfèrent souvent des versions personnalisées, simplifiées et adaptées à la scène ou la production en cours. Pour remédier à cette situation, il nous paraît utile de proposer des outils de transcription automatique. Des outils permettant de transcrire les éléments « standard » de la partition scénique lors des répétitions seraient précieux pour les assistants, et leur permettraient de se consacrer aux éléments plus spécifiques de la mise en scène.

Cependant, la transcription automatique d'un grand nombre d'actions exécutées par un grand nombre d'acteurs est un problème difficile, qui nécessite la mise au point de méthodes spécifiques. Si les classes d'actions sont bien choisies et constituent une notation rigoureuse, le problème a le mérite d'être bien posé. Les tâches qui relèvent de la vision par ordinateur concernent (1) le suivi des mouvements d'acteurs ; (2) l'identification des acteurs ; (3) le découpage temporel des actions scéniques de chaque acteur ; (4) l'identification des actions scéniques faisant partie du vocabulaire de notation choisi, à l'aide d'implémentations pratiques de la méthode de différentiation finie prévue théoriquement par Goodman.

Les techniques modernes de vision par ordinateur abordent cette question sous l'angle de l'apprentissage statistique. A partir d'un nombre suffisant d'exemples et contre-exemples de chaque classe d'actions, ces méthodes déterminent une fonction permettant de prédire la probabilité qu'un nouvel exemple appartienne à l'une ou l'autre classe. Cette fonction fournit donc une solution numérique au problème de la différentiation finie entre les actions, puisqu'elle permet de choisir entre deux classes et de rejeter celle dont la probabilité est la plus faible. Toutefois, il faut noter que l'apprentissage d'une action est compliqué par le fait qu'il doit être combiné avec l'apprentissage des apparences des acteurs qui exécutent l'action et des objets qui participent à l'action. En pratique, cela nécessite de constituer d'importantes collections d'exemples et contre-exemples d'actions exécutées par un grand nombre d'acteurs avec un grand nombre d'objets. C'est un travail que nous entreprenons dans le cadre de notre collaboration avec le théâtre des Célestins, que nous décrivons dans la dernière partie de ce chapitre.

Notation et montage des repetitions

Le projet de recherche « Scénoptique » est une collaboration entre le centre de recherche INRIA de Grenoble et le théâtre des Célestins de Lyon. Pour ce projet, nous développons un dispositif de prises de vues vidéo dédié aux répétitions de théâtre, comprenant trois caméras haute définition, de

façon à enregistrer une image de très haute résolution, du point de vue du « meilleur siège » de la salle. La problématique est celle de l'aide au montage des centaines d'heures de répétitions qui précèdent chaque spectacle. Cette problématique est importante pour un metteur en scène comme Ariane Mnouchkine, qui a souhaité montrer le « travail d'accouchement » de sa mise en scène de Tartuffe sous la forme d'un film[25] au prix d'un travail de prise de vues et de montage titanesque.

Dans le projet « Scénoptique », nous utilisons la notation de la mise en scène pour automatiser le dérushage et le montage des répétitions. La collaboration avec le théâtre des Célestins nous permet de disposer des indications de mise en scène, qui nous localisons à l'écran, afin de proposer automatiquement au metteur en scène des solutions de cadrage et de montage montrant les actions scéniques et leur enchainement. Le metteur en scène peut alors choisir parmi les solutions proposées et composer le film des répétitions, pour le diffuser par exemple sur le site internet du théâtre, ou auprès de ses abonnés.

La question de l'enregistrement cinématographique des spectacles de théâtre a été très bien formulée par Jean-Luc Godard[26] : *« Pourquoi les gens de théâtre n'ont-ils jamais envie de filmer leurs spectacles pour les garder comme archives ? (..) Ce serait très simple : la caméra au milieu de l'orchestre avec un objectif moyen – mais pas le zoom, qui donnerait déjà lieu à une interprétation ».* Nous avons adopté ce parti-pris de filmer au milieu de l'orchestre avec un objectif moyen, et avons découvert que cela revient à placer la caméra exactement à l'endroit que choisit le metteur en scène pendant les répétitions. Ce « point de vue du metteur en scène » présente plusieurs avantages théoriques et pratiques. D'une part, c'est le point de vue le plus adapté à la projection des films des spectacles dans la même salle. D'autre part, c'est le point de vue le plus proche des acteurs, et le seul qui permette de saisir le jeu des regards entre les acteurs et la salle. Enfin, c'est également le point de vue depuis lequel les éclairages sont mis au point, ce qui nous permet de filmer sans modifier ces éclairages. Ce dernier point reste difficile à réaliser en pratique, surtout lorsque les éclairages varient brusquement, mais cela nous paraît important de respecter les choix de mise en scène, ce qui n'est possible que de ce seul point de vue.

Suivi de la partition scénique

Pour faciliter le montage des répétitions, nous faisons l'hypothèse que les annotations fournies par les assistants de mise en scène forment une partition de la mise en scène, et nous développons les algorithmes de « suivi de partition » qui permettent de reconnaître les actions scéniques prévues par la partition. Ce projet nous permet donc de tester et valider nos hypothèses sur la partition scénique,

dans un cadre applicatif où la notation permet de porter à l'écran les actions scéniques.

Pour les dialogues, nous utilisons le système de transcription phonétique ScriptSync développée par la société américaine Nexidia et intégrée au logiciel de montage Avid Media Composer. Ce système permet d'aligner mot à mot le texte de la pièce avec chacun de nos enregistrements, ce qui nous permet de determiner qui parle à chaque instant de notre enregistrement. Pour l'image, nous développons des solutions spécifiques pour détecter, localiser et reconnaitre les acteurs à chaque instant. En combinant les deux sources d'informations audio et video, nous sommes donc en mesure de determiner à chaque instant où se trouve l'acteur qui parle. Cette information est capitale pour déterminer le choix des cadrages, notamment dans les scenes dialoguées.

Pour le suivi des actions scéniques, nous généralisons l'approche utilisée par le système ScriptSync qui effectue séparément deux transcriptions phonétiques du texte et de l'enregistrement audio, puis calcule une correspondance optimale entre les deux transcriptions. Nous généralisons cette approche au cas des actions scéniques, en développant des algorithmes spécifiques pour retranscrire la partition scénique et la captation vidéo sous forme de "visèmes", primitives visuelles de movement, qui jouent un role similaire aux phonèmes de ScriptSync. Nous appliquons ensuite un algorithme classique d'alignement temporel entre les visèmes de l'image et de la partition, en suivant une méthodologie que nous avons déja proposée pour aligner un film et son scenario[27, 28]. Lors des répétitions, cette approche permet également de comparer les versions successives de la partition, et de déterminer pour chaque version les ajouts et suppressions de visèmes. Cette extension de nos algorithmes fait l'objet d'une autre proposition de projet déposée en mars 2011 auprès de l'ANR, dans le cadre de l'appel à projet "CORPUS pour les sciences humaines" sous le nom de "Spectaclenligne(s)". Dans ce projet, la notation informatique est utilisée comme support pour l'analyse génétique du spectacle[29].

La figure 4 illustre comment nous détectons et reconnaissons les acteurs sur la scène. Nous déterminons pour chaque acteur une composition de « mots visuels » qui les caractérisent. Ces mots visuels sont des régions colorées, caractéristiques du vêtement, de la coiffure, et plus générale de l'apparence visuelle de l'acteur. La combinaison de ces mots visuels forme une signature qui peut être détectée et reconnue de façon efficace dans chaque image de la séquence. Cette information nous permet de proposer automatiquement des solutions de recadrage sur tous les acteurs et groupes d'acteurs présents sur scène. La figure 5 présente quatre exemples de recadrages proposés par notre système.

Figure 4: Indexation visuelle de 6 acteurs pendant les répétitions de "A l'ouest" [30] *de Nathalie Fillion. Mise en scène de Nathalie Fillion. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes.*

Figure 5: Scénoptique : trois recadrages d'une captation HD des répétitions de "A l'ouest" dans la mise en scène de Nathalie Fillion. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes. Coproduction - Théâtre du Rond-Point, Célestins, Théâtre de Lyon, Cie Théâtre du Baldaquin, AskUs , Le Gallia Théâtre-Saintes.

Conclusion

Dans ce chapitre, nous avons fait appel aux théories de Nelson Goodman pour proposer de construire une notation informatique pouvant servir de partition de la mise en scène théâtrale. En choisissant un schéma de notation restreint à un petit nombre d'actions scéniques reconnaissables visuellement, il semble possible de réaliser automatiquement la transcription de cette partition à partir d'enregistrements vidéo. Cette approche ouvre des perspectives nouvelles pour la mise en image et l'analyse du travail théâtral.

Remerciements

Le projet Scénoptique est financé par la Région Rhone-Alpes dans le cadre de son programme CIBLE d'aide à la recherche (2011-2014).

Références

[1] PAVIS, Patrice. Réflexions sur la notation de la mise en scène théâtrale. Revue d'histoire du théâtre, no. 2, 1981.

[2] PINHANEZ et BOBICK, Computer Theater: Stage for Action Understanding. Proceedings of the AAAI '96 Workshop on AI/A-Life Entertainment, AAAI Press Technical Report WS-96-03, Portland, OR, pp. 28-33, August 1996.

[3] RONFARD. Analyse Automatique de Films. Editions Universitaires Européennes, 2010.

[4] PUIG et PRIE, Construire et partager des modes de perception active de films annotés, Figures de l'interactivité, EESI, 2011.

[5] LABAN, Rudolf. La Maîtrise du Mouvement, Actes Sud, coll. "L'art de la danse", Arles, 1994.

[6] BENESH, Rudolf et Joan. An Introduction to Benesh Movement Notation, A. & C. Black, Londres, 1956.

[7] ESHKOL et WACHMAN. Movement notation. London: Weidenfeld and Nicolson; 1958.

[8] SUTTON, Valerie. Movement Shorthand, Book One: The Classical Ballet Key, 1973.

[9] SINGH et al. The Choreographer's Notebook - A Video Annotation System for Dancers and Choreographers. C&C 2011.

[10] GOODMAN, Nelson. Langages de l'art : Une approche de la théorie des symboles. Traduit par Jacques Morizot. Fayard/Pluriel, 2011.

[11] POGGI, Isabella. The Lexicon and the Alphabet of Gesture, Gaze, and Touch. IVA 2001.

[12] ARTAUD, Antonin. Le théâtre et son double, 1938.

[13] DAVIS, Marc. Media Streams, MIT Media Lab, 1996.

[14] WILKE et al. From dance notation to human animation: The LabanDancer project: Motion Capture and Retrieval. Comput. Animat. Virtual Worlds 16, 3-4, July 2005.

[15] CALDARONE, Marina et LLOYD-WILLIAMS, Maggie. Actions: The Actor's Thesaurus. Nick Hern Books; 2004.

[16] PINHANEZ, Claudio. Representation and recognition of action in interactive spaces - MIT Media Laboratory, PhD Thesis, 1999.

[17] SCHANK, Roger et ABELSON, Robert. Scripts, plans, goals and understanding: An inquiry into human knowledge structures, Erlbaum, 1977.

[18] WEINLAND et al. A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding 115(2), 2011.

[19] FEI-FEI, FERGUS, PERONA. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. CVPR 2004, Workshop on Generative-Model Based Vision.

[20] DENG et al. What does classifying more than 10,000 image categories tell us? European Conference of Computer Vision (ECCV). 2010.

[21] WEINLAND et al. Free Viewpoint Action Recognition using Motion History Volumes, Computer Vision and Image Understanding (2006).

[22] MACCOY, Peter. Essentials of Stage Management. Methuen Drama, 2004.

[23] SCHNEIDER, Doris. Arts & Craft of Stage Management. Wadsworth Publishing Co Inc, 1996.

[24] MARCELLO, Linda. The stage management handbook. Department of Drama and Dance, Hofstra University, 1998.

[25] DARMON, Eric et VILPOUX, Catherine. Au soleil même la nuit : scènes d'accouchements. Agat Films, Théâtre du Soleil, La Sept ARTE, 1997.

[26] GODARD, Jean-Luc. Conversation avec R. Allio et A. Bourseiller. Cahiers du cinema n° 177. Avril 1966.

[27] RONFARD et THUONG. A framework for aligning and indexing movies with their script. IEEE International Conference on Multimedia and Expo, 2003.

[28] RONFARD. Reading movies: an integrated DVD player for browsing movies and their scripts. ACM international conference on Multimedia, 2004.

[29] FERAL, Josette. Pour une analyse génétique de la mise en scène. Théâtre/public, Paris, automne 1998, no 144, pp. 54-59.

[30] FILLION, Nathalie. A l'ouest. Actes Sud-Papiers, janvier 2012.