

Le jazz et l'intelligence artificielle De la présence aux traces

Marc Chemillier

Février 2020

Les logiciels d'improvisation utilisés dans ce disque ont tous en commun de se servir de traces laissées par un musicien (en l'occurrence Bernard Lubat). Ces traces peuvent provenir d'un microphone qui capte un signal audio dans le piano. Elles peuvent aussi venir d'un dispositif optique de captation des notes comme le système Piano Bar de Moog que l'on place au fond du clavier du piano pour détecter le mouvement des touches (ou les systèmes MIDI du même type intégrés dans les pianos Disklavier de la firme Yamaha). Enfin ces traces peuvent provenir d'un détecteur de hauteurs inclus dans les logiciels qui extrait des motifs mélodiques à partir du signal audio, c'est le système qu'on utilise pour détecter des mélodies dans la voix de Bernard. Les traces obtenues sont ensuite traitées par trois logiciels différents. Le logiciel OMax découpe les phrases jouées par le musicien en fragments qui sont ensuite recombinaés selon des propriétés statistiques. Le logiciel Djazz effectue la même opération mais en se calant sur un repérage temporel en unités régulières qui donne un sentiment de pulsation. Enfin le logiciel SoMax crée instantanément une co-improvisation polyphonique qui suit le musicien live en puisant dans un corpus de modèles appris au préalable. Lorsque ces logiciels manipulent non pas le signal audio, mais les notes extraites de ce signal, ils peuvent réorchestrer les phrases avec toute une palette d'instruments stockés dans un échantillonneur. Aucun son n'est produit par les ordinateurs qui ne soit dérivé, par l'un des processus décrits ci-dessus, de ce que Bernard a joué. C'est un choix que nous avons fait.

Ainsi les échanges présentés dans ce disque entre Bernard Lubat et les logiciels d'improvisation peuvent-ils se résumer à la formule suivante : « du Lubat, rien que du Lubat ». Mais ces traces manipulées par les logiciels sont-elles vraiment « du Lubat » comme on dit de certains yaourts ou jus de fruit qu'ils ont « de vrais morceaux de fruit dedans » ? Lorsqu'il a neigé pendant la nuit et qu'on découvre autour de la maison des traces de pas, on imagine qu'un animal est passé à proximité. Il n'est plus là car on ne le voit pas, mais il est peut-être encore caché dans le voisinage à l'abri dans un buisson. Ces traces laissées dans la neige indiquent la présence de l'animal bien qu'il ne soit plus présent à proprement parler.

Il y a quelque chose de paradoxal à vouloir improviser du jazz avec les traces laissées par un musicien qui sont, en quelque sorte, détachées de son corps. Car le jazz est d'abord une affaire de corps et de présence corporelle. L'engagement physique des musiciens dans le jazz est considérable et se traduit de multiples façons qui appartiennent au folklore de cette musique : grognements chez Elvin Jones, Erroll Garner ou Keith Jarrett (on entend Bernard grogner dans l'enregistrement de Philadelphie), gouttes de sueur chez Cecil Taylor, langue pendante chez Art Blakey, etc. De plus la relation entre les musiciens et le public se traduit, elle aussi, par différentes manifestations corporelles (frappements de mains, interjections « Come on », « One more time », danse). Cette relation participative avec le public doit être rattachée au contexte rituel dans lequel la musique était pratiquée au sein des sociétés africaines d'où provenaient la plupart des ancêtres des musiciens afro-américains.

Dans la tradition du jazz, la musique se jouait à un instant donné, dans un lieu donné et avec des personnes présentes en chair et en os, musiciens et spectateurs. Lors du Workshop ImproTech que nous avons organisé à Philadelphie en décembre 2017, le théoricien des médias et grand spécialiste de l'improvisation avec machines Bob Ostertag avait donné une brillante conférence sur le thème de l'improvisation. Il soulignait en particulier ce caractère essentiel de la présence pour définir ce qu'est l'improvisation : « Ce que l'improvisation signifie, c'est qu'un humain doit encore être là. Le sens de l'improvisation place celle-ci comme le dernier rempart pour affirmer que des humains doivent être

présents dans la production de musique ». C'est pour cette raison que les collectionneurs de disques de jazz s'attachaient à relever avec soin les informations discographiques mentionnées sur les pochettes (qui joue de quel instrument ? à quelle date et où?). De grands chroniqueurs de disques comme Hugues Panassié, concernant la période du jazz dit « classique », savaient reconnaître à l'oreille les musiciens d'une séance d'enregistrement en analysant les caractéristiques de leur jeu, de leur toucher, de leur phrasé et de leur sonorité, et pouvaient le cas échéant rectifier les erreurs de la pochette. On notera que cette approche tend à devenir anachronique car la plupart des plateformes de streaming aujourd'hui ne donnent presque plus d'informations discographiques (Qobuz en donne un peu plus que Spotify ou Deezer) de sorte que pour un amateur de jazz cherchant un enregistrement particulier dans la grande tradition de cette musique, il est difficile de savoir s'il s'agit bien de la version enregistrée en telle ou telle année.

Une trace enregistrée est toujours une réduction. C'est un débat sans fin pour les jazzophiles qui se heurtent au paradoxe selon lequel l'enregistrement de studio est jugé le plus souvent comme une trace inférieure à la performance réelle en concert alors que ce sont pourtant les traces en studio qui ont fait connaître le jazz à la plupart de ces amateurs. Mais le caractère réducteur de la trace produite par les technologies du son va bien au-delà de la distinction entre enregistrements de studio et concerts. La technologie transforme en profondeur et de façon radicale la pratique du concert lui-même. Bernard Lubat nous a souvent parlé de ce qu'il appelle les « technologies de la représentation » (dispositifs technologiques qui s'interposent entre l'artiste sur scène et le public) pour nous dire son attachement à la « dialectique de vivant à vivant » qui doit s'instaurer, selon lui, pendant le concert entre celui qui joue la musique et celui qui l'écoute. Sur ce point il reprend à son compte les réflexions sur la notion de « relation » développées par le poète antillais Édouard Glissant (*Poétique de la relation*, Gallimard, 1990). Or force est de constater que l'accroissement des dispositifs technologiques dans le spectacle vivant tend à réduire la relation entre musiciens et spectateurs. Le gigantisme des concerts (stades, arènes, etc.) rendu possible par la technologie du son en est l'un des symptômes les plus évidents et parfois, les musiciens ne sont visibles pour le public que grâce au déploiement d'écrans géants sur lesquels leur image est projetée.

Au-delà des technologies d'amplification permettant de rassembler des publics toujours plus nombreux, d'autres technologies contribuent à réduire les effectifs et, de ce fait, à appauvrir les relations. Ainsi le synthétiseur doté à l'origine d'une expressivité propre a-t-il finalement remplacé l'orchestre en permettant une baisse des coûts pour les organisateurs de spectacle. Puis le DJ a remplacé le groupe lui-même et on observe qu'aujourd'hui, dans de nombreux bars à musique, les patrons préfèrent payer un DJ plutôt qu'un groupe de plusieurs musiciens. Dans d'autres musiques que le jazz, où la relation entre l'artiste et son public prend des formes plus marquées par les phénomènes extra-musicaux de starification (variété, pop, rock), on assiste au développement de pratiques comme le playback qui éliminent toute forme de relation d'ordre strictement musical. Lors de l'avènement du disco à la fin des années 1970, le groupe Boney M a construit sa carrière entièrement sur l'utilisation du playback. Parmi les quatre personnes qui étaient présentes sur scène (un homme et trois femmes), la seule qui chantait vraiment était Liz Mitchell et les autres se contentaient de faire semblant. Aujourd'hui, Justin Bieber chante en playback devant ses fans ce qui permet, on l'imagine, de limiter les risques du direct devant des publics qui peuvent comporter jusqu'à 20000 personnes comme à Paris-Bercy.

Philippe Le Guern, professeur en sciences de la communication à l'Université de Nantes, s'intéresse à l'utilisation des robots en art. Il nous avait invités, Bernard Lubat, Gérard Assayag et moi-même, à une rencontre intitulée « Artefacts sonores » qu'il organisait en juin 2018 à la BPI de Beaubourg à Paris. Il a publié avec Maël Guesdon une étude passionnante dans laquelle il analyse le cas emblématique des hologrammes (*Où va la musique ?*, Presses des Mines, 2016). Cette technologie se répand aujourd'hui pour faire jouer sur scène à titre posthume des célébrités disparues comme Whitney Houston programmée Salle Pleyel en mars 2020 alors qu'elle est décédée en 2012. Mais le

cas le plus troublant est celui de la star virtuelle Hatsune Miku qui ne correspond à aucune identité réelle. Apparue il y a un peu plus de dix ans comme une silhouette de manga dotée d'une voix générée par le logiciel de synthèse vocale Vocaloid, elle totalise aujourd'hui environ 2,5 millions de fans sur sa page Facebook et elle était en 2013 la vedette d'un opéra *The End* donné au théâtre du Châtelet. Les internautes peuvent faire chanter Hatsune Miku en utilisant le logiciel et réaliser des clips vidéo à l'aide d'un modèle en 3D du corps de la star. Les clips postés sur le web sont discutés et évalués par la communauté des fans. C'est cette dimension participative qui est revendiquée dans les forums comme l'un des aspects les plus nouveaux et les plus fascinants de l'expérience Hatsune Miku. Mais la question de la présence est fortement perturbée dans une telle situation. Le lien classique observé dans les phénomènes de starification entre les images omniprésentes de la star partagées par des millions de fans, et l'inaccessibilité de la star elle-même dont l'intimité est protégée par des dispositifs de sécurité, est rendu complètement obsolète dans ce cas puisqu'il n'y a pas d'intimité.

La société Crypton Future Media qui gère l'hologramme vend les images et la voix modélisés, mais elle le fait dans une conception pyramidale de l'organisation de cette communauté. Les concerts doivent rester le modèle de toutes les autres images produites par les fans. Ceux-ci se retrouvent lors des événements organisés par la firme qui simule un vécu émotionnel partagé entre la star et son public, par exemple en la faisant pleurer sur scène à l'occasion d'une chanson. Les vidéos des concerts filmés sont diffusés sur Internet et permettent aux participants de garder une trace du regroupement physique de la communauté. Trois chercheuses en études culturelles et artistiques de l'Université de Copenhague ont analysé le modèle économique de Crypton Future Media dans la revue *Digital Creativity* (Volume 28, Issue 4, 2017). L'un des aspects originaux de ce modèle est d'encourager les fans à créer de la musique ou de la vidéo avec la voix et les images de la star, ce qui est très différent de l'approche observée avec d'autres personnages vedettes comme Batman dont les produits dérivés sont contrôlés par la société DC Comics qui en détient les droits et qui ignore les productions des fans. Les auteurs replacent cette approche originale de Crypton Future Media dans la perspective de l'économie de l'attention, terme qui désigne le fonctionnement de marchés dans lesquels l'offre est abondante et donc économiquement dévalorisée (par exemple la musique sur Internet), et la ressource qui est rare et recherchée devient le temps et l'attention des consommateurs.

L'apparition de fausses identités ou « fausses présences » sous forme d'avatars numériques, n'est pas un phénomène isolé, il est au contraire utilisé massivement sur le web dans de nombreux domaines. Il existe des « fermes » de faux sites liés entre eux qui pointent vers un site cible et permettent de renforcer son poids dans les moteurs de recherche. On voit également se développer les faux avis dans l'hôtellerie ou le commerce émis par des petites mains en Inde ou à Madagascar qui sont payées pour mettre des étoiles sur les produits proposés en faisant croire que des vrais gens ont exprimé des vraies opinions. David Chavalarias, chercheur au Centre d'analyse et de mathématique sociales de l'EHESS (CAMS) où est développé le logiciel d'improvisation Djazz, est un spécialiste de la manipulation d'information sur le web dans le champ politique et l'un des concepteurs d'un outil d'analyse appelé Politoscope. Son équipe a ainsi obtenu des résultats spectaculaires sur les phénomènes de désinformation populaire planifiée qu'on appelle l'astroturfing (voir son article sur le site web AOC « Analyse Opinion Critique », 7 novembre 2018). Il s'agit de techniques destinées à créer l'illusion d'un consensus ou d'un soutien spontané dans le champ politique ou dans celui des industries culturelles. Ces techniques peuvent être manuelles lorsqu'un groupe se coordonne sur Internet pour démultiplier un message de façon artificielle. C'est ce que font de nombreux groupes militants. Elles peuvent devenir algorithmiques lorsqu'elles utilisent des robots, ou à l'échelle d'acteurs plus puissants comme les États, certains logiciels spécialisés pour gérer des centaines d'identités virtuelles, chacune ayant un compte Facebook, Instagram etc. avec des photos et des publications pour faire croire à de vrais comptes (logiciels dits sock puppets). Le Politoscope a montré grâce à une cartographie de la circulation des messages sur Twitter que des groupes

étrangers d'extrême-droite, notamment américains, avaient utilisé certaines de ces techniques pour s'immiscer dans la campagne présidentielle française de 2017 pour favoriser la candidature de Marine Le Pen.

Le problème de la manipulation d'information est jugé suffisamment préoccupant pour qu'un rapport sur le sujet ait été commandé au Ministère de l'Europe et des Affaires étrangères et au Ministère des Armées en août 2018. D'après ce rapport, le risque n'est pas que les manipulations changent l'opinion des électeurs, mais plutôt qu'elles sèment chez eux le doute et la confusion pour encourager le passage à l'acte, c'est-à-dire le fait d'aller voter (*Les Manipulations de l'information, un défi pour nos démocraties*, p. 25). On sait que dans les scrutins, c'est à l'intérieur de la masse des indécis que se joue le résultat, et qu'il suffit de convaincre un faible pourcentage de votants avec ces messages pour obtenir la victoire. Il ne faut certes pas exagérer le potentiel nuisible des technologies numériques dans ce contexte. La propagande existe depuis toujours et l'on se souvient que l'émir Ben Kalish Ezab envoyait par avion des tracts sur le campement de Bab El Ehr dans *Tintin au pays de l'or noir*. Mais l'une des forces du numérique est de permettre le « microciblage » dans des campagnes de ce type. C'est ce que pratiquait la société californienne Cambridge Analytica lorsqu'elle s'est retrouvée au centre d'un scandale qui a éclaté en 2018. On s'est aperçu qu'elle exploitait les données personnelles de 87 millions d'internautes pour établir des profils psychologiques individuels extrêmement précis afin de pouvoir cibler les électeurs. Face à ces dérives dans l'utilisation du numérique, il convient cependant de ne pas jeter le bébé avec l'eau du bain. Avant de donner naissance à ces pratiques qui s'industrialisent et représentent une menace, le web a d'abord été investi par de simples militants. Fabien Granjon, sociologue qui mène depuis plusieurs années un travail de grande ampleur sur ce qu'il appelle le « front culturel de résistance populaire » déployé par Bernard Lubat à Uzeste, avait été l'un des premiers à analyser l'appropriation du web au sein des mouvements sociaux dans son livre *L'Internet militant* paru en 2001 (Éditions Apogée). D'un côté, la circulation des idées sur le web est stimulée par le cybermilitantisme étudié par Fabien Granjon et d'un autre côté, elle est brouillée par les opérations de manipulation de l'information décrites par David Chavalarias. Il faut prendre garde qu'à trop vouloir le nettoyer, on risquerait de refermer cet espace de libertés qui a été ouvert par le monde numérique.

L'une des caractéristiques des traces laissées sur le web est de pouvoir se prêter à des mesures : par exemple on peut compter les clics des internautes. La notion de mesure a connu un essor considérable depuis les politiques néolibérales des années 1980 avec sa systématisation dans le champ de la décision publique et la généralisation de la politique des indicateurs. Pour des chercheurs comme Gérard Assayag et moi-même, l'usage de mesures quantifiées pour évaluer notre travail prend la forme de ce qu'on appelle la « bibliométrie ». On désigne par ce terme le fait de compter le nombre de fois où un article scientifique a été cité par d'autres. La plupart des publications étant aujourd'hui disponibles en ligne, il devient possible d'utiliser des algorithmes pour compter toutes les traces sur le web laissées par les chercheurs quand ils citent des articles de leurs collègues. L'idée est que plus un chercheur est cité par d'autres, plus son travail peut être considéré comme digne d'intérêt. Des agences de notation ont été développées pour effectuer ces calculs et leurs résultats tendent de plus en plus à s'immiscer dans la gestion de la carrière des chercheurs (recrutement, avancement, etc.). Dans l'ouvrage *Derrière les grilles* paru en 2014 aux éditions Mille et une nuits, Barbara Cassin et Roland Gori s'insurgent avec véhémence contre les dérives absurdes de ce système. L'une de ses aberrations les plus visibles est le fait qu'il n'accorde aucune attention au contenu des articles, c'est-à-dire à ce qu'écrivent effectivement les chercheurs. Les auteurs dénoncent également la tendance perverse du système à fonctionner comme une « prophétie auto-réalisatrice » : par essence la nouveauté n'est pas déjà connue à l'avance et n'est donc pas citée de sorte que les mécanismes mis en cause ne peuvent mesurer que ce qui est connu, et les décideurs qui s'en servent ne peuvent encourager et soutenir financièrement que ce qui est déjà découpé et répertorié. Comme le soulignent les auteurs, l'invention est la grande absente de tels dispositifs.

Le calcul sur les traces des internautes a pris une forme qui connaît aujourd'hui un succès retentissant avec le développement des algorithmes de recommandation. Les traces laissées par les utilisateurs sur le web (il s'agit ici de vraies traces attestant d'une réelle présence qui est traquée en quelque sorte par les algorithmes, et non des fausses traces dont on a parlé plus haut à propos de l'astroturfing) sont utilisées pour essayer de prédire le futur à partir du passé. Ce principe est appliqué dans des situations très variées, par exemple dans la suggestion de clavier de nos téléphones portables qui fonctionne en calculant la distance du mot saisi par rapport à des mots conservés en mémoire, notamment des mots qu'on a saisis antérieurement. Dominique Cardon, sociologue spécialiste de l'Internet, a étudié les mécanismes de recommandation mis en œuvre dans ce qu'on appelle le « filtrage collaboratif » : on propose à l'utilisateur des choses qui ont été achetées par les personnes ayant acheté auparavant la même chose que lui. Les plateformes de distribution de musique en ligne comme Deezer ou Spotify ont fait des outils de recommandation leur pierre angulaire et ils permettent parfois au mélomane de faire de réelles découvertes. D'une certaine manière, le rayonnage dans les bibliothèques publiques provoque des effets comparables quand on découvre un livre parce qu'il se trouvait rangé dans le même rayon qu'un autre livre qu'on avait emprunté. Les techniques de recommandation comportent toutefois le risque d'enfermer l'utilisateur dans les limites de ses propres routines puisqu'elles le conduisent à « ajuster ses désirs sur la régularité de ses pratiques » comme le souligne Dominique Cardon dans son livre *À quoi rêvent les algorithmes ?* (Le Seuil, 2015, p. 88). Plus généralement elles entérinent un certain ordre dont on voudrait bien être assuré qu'il ne deviendra pas la norme imposée sans recours possible. C'est un problème inhérent à tous les automatismes: les usagers doivent pouvoir les débrayer afin de « passer en manuel ». Tous les utilisateurs de téléphones portables ont fait l'expérience des mécanismes de suggestion de clavier récalcitrants qui refusent d'écrire les mots qu'on essaie désespérément de taper. Une autre dimension inquiétante de ces mécanismes vient de ce qu'ils reposent sur les traces que nous laissons au fil de notre navigation sur des interfaces numériques. Les sociétés qui travaillent dans le domaine du marché de la donnée publicitaire sont ainsi amenées à disséminer des cookies dans nos ordinateurs et nos appareils afin de collecter ces traces de navigation. Le recours à de telles pratiques se situe à la limite de la légalité et pose des problèmes aigus de respect des données personnelles.

Les logiciels d'improvisation que nous avons développés sont basés sur un principe comparable à celui que l'on vient de décrire consistant à « prédire » le futur à partir du passé. Les phrases musicales calculées par ordinateur prolongent en quelque sorte celles de Bernard Lubat et le jeu de Bernard est traité comme un ensemble de traces collectées par les algorithmes. Ce processus relève de ce qu'on appelle l'apprentissage automatique qui est un champ d'études de l'intelligence artificielle. Apparue au milieu des années 1950, cette discipline a évolué au cours du temps en fonction des problèmes posés et des techniques disponibles pour les résoudre mais son ambition depuis le début est d'accomplir des tâches normalement réservées aux humains. La création musicale en fait partie et à ce titre, on peut parler de créativité artificielle. Depuis les années 1980, l'évolution dans ce domaine s'est traduite par la montée en puissance des systèmes basés sur la collecte de traces et l'apprentissage au détriment des systèmes basés sur des règles explicites. George Lewis, musicien de l'AACM (Association for the Advancement of Creative Musicians), fut l'un des pionniers de l'improvisation par ordinateur en faisant interagir des programmes informatiques avec lui-même jouant du trombone ou avec d'autres grandes figures du jazz comme le saxophoniste Roscoe Mitchell, lui aussi membre de l'AACM. Il a conçu ses premiers systèmes dès la fin des années 1970 en implémentant des ensembles de règles qui formalisaient le comportement d'un musicien quand il improvise.

Dans le cas des logiciels de créativité artificielle sur lesquels nous travaillons, il ne s'agit pas véritablement de prédire le futur à partir des traces du passé (cela éliminerait l'idée de créativité). Il s'agit plutôt de produire des choses qui ressemblent à celles qui ont été enregistrées, mais en invitant

également le système à produire de la surprise et de l'invention. Cela est rendu possible par l'existence de choix aléatoires dans le processus de recombinaison. En définitive ce qui nous intéresse, ce n'est pas ce que produit la machine, c'est plutôt sa confrontation avec le musicien lui-même. Et pour être plus précis, il faudrait ajouter qu'il s'agit de confronter la machine, le musicien et les deux opérateurs informaticiens qui paramètrent les calculs. Car les résultats engendrés par les logiciels d'improvisation dépendent aussi d'une intervention humaine, et donc de la présence de personnes en chair et en os pour les piloter. En fin de compte, les logiciels d'improvisation sont surtout l'occasion d'une rencontre à trois, Bernard Lubat, Gérard Assayag et moi-même, certes médiatisée par des dispositifs informatiques, mais dans laquelle l'idée de présence est aussi importante que si l'on pratiquait du jazz « à l'ancienne » c'est-à-dire sans les machines. De ce point de vue, le travail que nous menons depuis près de vingt ans avec Bernard Lubat a été une source constante d'émerveillement par la richesse des réactions que celui-ci a exprimées au sujet de cette confrontation. On pourra en juger par ses commentaires à l'issue de la séance d'écoute du 24 février 2020 au cours de laquelle nous avons sélectionné les enregistrements du présent disque : « Ça m'impulse, ça change ma façon de jouer du piano. Je prends conscience d'aller vers le pire. Ça soigne ma peur, on a toujours peur de déplaire, on a toujours peur de soi ».

Addendum — Nous sommes en avril 2020. Près de la moitié de l'humanité se trouve actuellement confinée pour tenter d'enrayer l'épidémie de coronavirus. Le monde du spectacle vivant est durement touché par cette crise sanitaire avec des annulations en cascade et des salles fermées. La notion de présence prend un éclairage singulier dans de telles circonstances. Les musiciens explorent différentes voies utilisant les nouvelles technologies pour s'efforcer de recréer des liens. Le pianiste Chick Corea se filme tous les jours sur sa page Facebook au piano chez lui en train d'improviser ou de jouer des partitions (Chick's Workshop). Par moment il se lève et vient derrière la caméra pour voir si des personnes sont connectées et le cas échéant, leur adresse un message amical. On assiste également à la multiplication des vidéos en forme de mosaïques d'écrans montrant des performances orchestrales enregistrés non pas en direct, mais séparément et à distance pour être ensuite mixées ensemble (le *Boléro* de Ravel interprété de cette façon par l'Orchestre national de France a été vu 2,4 millions de fois sur YouTube). Du côté des auditeurs, les plateformes d'écoute musicale partagée (*social listening*) où des internautes se connectent simultanément voient leur audience s'envoler. La technologie permet ces relations virtuelles médiatisées par Internet grâce à la vidéo et aux réseaux sociaux. Il est difficile de dire ce qu'il en sera après la crise du coronavirus. On reviendra sans doute à un mode de relation musicale traditionnel, mais beaucoup de scènes musicales ne rouvriront pas et le spectacle vivant subira de lourdes pertes dans cette crise. Verra-t-on également se pérenniser des pratiques musicales basées sur un mode de présence « virtuelle » entièrement dépendante de la technologie et dans ce cas, quelles seront les conséquences à terme sur le plan esthétique, social, psychologique et environnemental ?