



HAL
open science

Arborescences et voix de synthèse des serveurs vocaux interactifs

Julien Daillère

► **To cite this version:**

Julien Daillère. Arborescences et voix de synthèse des serveurs vocaux interactifs. JIT 2022 - Journées d'Informatique Théâtrale 2022, Inria Grenoble Rhône Alpes; ENSATT Lyon; Laboratoire Passages XX-XXI de l'Université Lyon 2, Oct 2022, Lyon, France. hal-04153520

HAL Id: hal-04153520

<https://inria.hal.science/hal-04153520>

Submitted on 6 Jul 2023

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

ARBORESCENCES ET VOIX DE SYNTHÈSE DES SERVEURS VOCAUX INTERACTIFS

Julien Daillère

Institut de recherche théâtrale et Multimédia.
Université des Arts de Târgu-Mureș, Roumanie.
j.dailiere@gmail.com

RÉSUMÉ

Dans cet article, j'évoque le cheminement qui m'a amené à m'inspirer des arborescences et voix de synthèse des serveurs vocaux interactifs (SVI) et à les utiliser dans ma pratique artistique, depuis la performance à distance jusqu'à la scène théâtrale. C'est la contrainte du confinement de mars 2020, lors de la crise sanitaire de Covid-19, qui m'a conduit à développer l'usage du téléphone pour partager mes textes avec un public distant. D'abord intéressé par la mise en place d'un véritable SVI, j'ai été confronté aux limites techniques du dispositif. Pour les dépasser, j'ai choisi d'imiter moi-même, en direct, la voix de synthèse d'un SVI selon une arborescence constituée d'hyperliens dans un traitement de texte. Particulièrement stimulé par cette tentative d'imitation de la machine et les stratégies qu'elle m'amenait à mettre en place, j'ai décliné la forme du Serveur Vocal Humain à différentes occasions, jusqu'à en faire une séquence du spectacle *Je t'aime effondrement*, créé en salle à l'automne 2020. Ce fut aussi le point de départ d'une réflexion sur l'usage, pour le spectacle vivant, de l'écriture en arborescence utilisée couramment dans les jeux vidéo.

1. CONTEXTE

En mars 2020, avec la mise en place du premier confinement, j'ai souhaité créer des formes, entre performance et théâtre, destinées à un public distant. Me souvenant avec nostalgie de mes lectures de textes par téléphone, au début des années 2000, j'ai voulu reprendre cette démarche, aussi pour proposer un dispositif non-numérique (du point de vue usager) plus accessible et avec un plus faible impact énergétique que la diffusion vidéo en direct sur Internet ou « livestream ».

J'ai alors lancé le groupe Facebook « Artistes au téléphone » pour rassembler des artistes utilisant ce même médium. En parallèle, j'ai pu observer le foisonnement de propositions artistiques et de médiation, en amateur ou professionnelles, passant par l'audio du téléphone. Rapidement, au-delà des simples lectures de textes, j'ai voulu explorer d'autres formes, notamment pour développer l'interactivité avec le public.

Je me suis intéressé aux spectacles immersifs par téléphone comme les pièces du collectif américain Candle House Collective (*Next Time*) ou encore de la compagnie canadienne Théâtre Déchainés avec par exemple la pièce *Tantôt, demain peut-être* de Marie Ayotte pour lequel une comédienne vous appelle sur WhatsApp en s'adressant à vous comme à un personnage de fiction : « Allô, c'est toi ? Ça me fait quelque chose d'entendre ta voix... ».

En parallèle de mes recherches sur le présentiel par délégation d'action avec la téléperformance (une forme de théâtre d'appartement audioguidé à distance par téléphone, que j'ai pu présenter à différentes reprises en France et, par un audio via Internet, en Roumanie), je me suis intéressé aux serveurs vocaux par téléphone pour leur dimension interactive, à travers une écriture en arborescence dont le déploiement découle des choix des personnes qui appellent.

2. SERVEURS VOCAUX INTERACTIFS

En avril 2020, en partenariat avec la société TLM Com qui conçoit notamment des SVI (serveurs vocaux interactifs) pour des établissements hospitaliers, j'ai commencé à travailler sur un prototype de SVI avec un technicien de leur équipe. Ce SVI était destiné à proposer la diffusion d'enregistrements audio de quelques-uns de mes poèmes et courts textes.

Alors que l'objectif principal d'initiatives similaires - depuis *Dial-A-Poem* de John Giorno en 1968 jusqu'au *Serveur Vocal Poétique* de la compagnie Home Théâtre ou bien encore les textes enregistrés comme messages de répondeur téléphonique par les membres de la compagnie du Théâtre du Rabot, en 2020 - était de faire entendre la lecture d'un texte, j'étais pour ma part surtout intéressé par la construction d'une arborescence longue et complexe, imposant tout un cheminement ludique et poétique en amont de la diffusion d'un texte enregistré, qui tenait presque du prétexte. Cependant, outre les contraintes techniques de programmation informatique qui limitaient mon ambition quant à la complexité de l'arborescence, la forme d'interaction proposée par un véritable SVI est finalement très limitée et, surtout, ne permet pas de retrouver la temporalité partagée propre au spectacle vivant.

C'est à partir de ces réflexions que j'ai conçu le Serveur Vocal Humain. J'ai repris l'arborescence de mon prototype de SVI telle que développée pour TLM Com, transposée dans un simple fichier de traitement de texte avec de multiples hyperliens – afin de pouvoir me déplacer rapidement dans l'arborescence en cliquant dessus, et ainsi *jouer* moi-même le SVI, en direct par téléphone, avec ma voix.

Au moment de proposer ce Serveur Vocal Humain pour la première fois en mai 2020, je l'avais annoncé sur les réseaux sociaux comme étant un véritable SVI composé à partir d'enregistrements retraités de ma propre voix, à la manière des voix de synthèse. Je faisais croire que les appels sur mon numéro de téléphone portable, sur une plage horaire précise, étaient automatiquement redirectionnés vers le SVI. Ce n'était bien sûr pas le cas...

En début d'appel, lorsque je décrochais, j'imitais immédiatement une voix de synthèse pour annoncer : « À des fins de recherche et développement, cet appel est susceptible d'être enregistré après le bip sonore. » Ainsi, je prévenais la personne au téléphone que j'allais possiblement entendre toutes ses réactions en différé – même si j'entendais tout en direct, cela me semblait une manière satisfaisante de sécuriser l'interaction proposée.

Pour jouer ce Serveur Vocal Humain, je me suis inspiré de ce que j'avais constaté lors de mon travail sur un véritable SVI avec la société TLM Com : si le bug technique d'une intelligence artificielle ou d'un robot provoque ce que Salvatore Anzalone, Giulia Filacanapa et Erica Magris appellent une « déconstruction » [1] de « l'illusion de vie » [2], en référence au concept développé par Didier Plassard, à l'inverse, pour mon Serveur Vocal Humain, l'imitation du bug technique pouvait m'aider à construire l'illusion de non-vie, ou de présence machinoïde : cela jouait comme un effet de réel. En conséquence, j'ai cherché à simuler le bug technique de plusieurs manières.

Tout d'abord, j'avais pu observer les limites de l'intelligence artificielle utilisée pour le système de reconnaissance vocale : celle-ci bugue facilement lorsque la diction est imprécise ou lorsque des bruits parasites se mêlent à la parole. Je guettais donc la moindre opportunité de prononcer les messages d'erreur habituels, comme par exemple : « Je n'ai pas compris votre demande. »

Pour simuler des enregistrements de voix maladroitement tronqués, je me coupais parfois la parole, volontairement, avec la touche « silencieux » de mon téléphone portable, en début ou en fin de phrase.

Certaines vocalisations sont impossibles à retranscrire en voyelles et consonnes. Par exemple, dans le menu renvoyant vers différentes thématiques de mes poèmes, l'arborescence comprenait la proposition suivante : « Si vous souhaitez écouter un texte sur le thème des loups, faites le cri du loup. » Mais lors de mes tests avec le véritable SVI que je développais via TLM Com, j'avais pu noter que si une personne au téléphone imitait

vraiment le cri du loup, le système de reconnaissance vocale était incapable de reconnaître ce cri car il ne correspondait à aucune de mes tentatives de retranscription, aussi diverses soient elles : a-ou, aaa-ou, a-ou-ou-ou, etc. Moi seul, avec mon oreille humaine, était capable de le reconnaître. C'est pour cette raison que je m'autoproclamai alors « Meilleur système de reconnaissance vocale au monde », même si ce titre honorifique est bien sûr resté assez confidentiel.

D'autre part, pour me rapprocher des voix de synthèse, je me contraignais à limiter les variations de ma propre voix en terme de hauteur de ton et de vitesse, qui sont les principaux paramètres de programmation disponibles. J'étais parfois une voyelle sur une note continuuuuue, ou j'accentuais l'interprétation caricaturale des signes de ponctuation. Je mettais aussi en exergue la mélodie très marquée d'une phrase destinée à être répétée, pour donner l'illusion de la rediffusion d'une même séquence audio. Par exemple : « Merci de patienter... Merci de patienter... Merci de patienter... » avec par exemple une accentuation très prononcée du « pa- » et un long étirement du « -ter » dans le mot « patienter ».

Au cours des appels, certaines personnes ont rapidement compris que je performais en direct, ont réussi à me faire éclater de rire ou m'ont simplement fait bafouiller d'une manière typiquement humaine (et non buguer). Ce fut le cas lors de l'appel de mon ami Virgile Pons dont j'avais enregistré l'appel.



Figure 1. QR Code pour écouter l'appel¹.

D'autres personnes y ont cru jusqu'au bout. Ainsi, une de mes sœurs a suivi toute l'arborescence une première fois, m'a écouté lire un premier texte en entier, est repartie dans l'arborescence, a commencé de m'écouter lire un deuxième texte puis a subitement raccroché pendant la lecture. Un peu plus tard, alors que je l'appelais au téléphone pour recueillir ses impressions, elle me confirma qu'elle n'avait pas compris que je performais en direct, ce qui fut pour moi un grand encouragement.

La mère d'une amie resta près de vingt minutes à naviguer dans l'arborescence, à écouter des textes. Au bout d'un moment, elle arriva sur un menu qui lui proposait notamment : « Pour plus d'info sur le travail de Julien Daillère, tapez sur étoile. » Et lorsque j'entendis

¹ Alternative au QR Code : https://soundcloud.com/julien_daillere/serveur-vocal-

sonner le « bip » de la touche étoile, je me retrouvai un peu dépourvu car je n'avais pas rédigé de contenu correspondant. Je lui dis simplement, avec ma voix « humaine » : « Eh bien, voilà, c'est moi ! Je suis Julien Daillère ! » Ce à quoi elle répondit : « Oh bah ça alors, je ne m'attendais pas à tomber sur vous ! » Ce qui nous fit bien rire quand je lui expliquai que « c'était moi depuis le début ».

Après ces appels en tête-à-tête, j'ai été accueilli sur les ondes de la webradio Radio Galoche en juillet 2020. Depuis les studios de l'Espace DOC ! à Paris, je proposais aux auditeurs et auditrices de la radio d'appeler le Serveur Vocal Humain par téléphone et notre interaction était donc audible en direct à la radio.



Figure 2. QR Code pour écouter le podcast².

Le plus laborieux, pendant les quarante minutes d'émission, était de devoir répéter une annonce en boucle entre deux appels : « Bonjour, et bienvenue sur Radio Galoche pour le *Serveur Vocal Humain* de Julien Daillère. Appelez dès maintenant le 06 69 18 75 27. Appelez dès maintenant le 06 69 18 75 27 pour explorer les méandres de l'arborescence du *Serveur Vocal Humain* afin de découvrir les poèmes et autres textes de Julien Daillère. »³

Peu de temps après, j'ai également conçu et performé une version spécifique du Serveur Vocal Humain en lien avec une exposition d'art visuel de Théo Poulet à ZAOUM' Espace d'artistes, à Clermont-Ferrand, pour donner à entendre par téléphone des textes écrits par l'artiste en lien avec ses œuvres exposées. La visite de l'espace physique était transposée en un parcours virtuel, à distance, à travers l'arborescence du SVI (que j'interprétais en direct).

J'ai ensuite intégré le serveur vocal humain à une forme théâtrale pour la scène.

Au cours de l'été 2020, alors que je travaillais à la création du spectacle de théâtre *Je t'aime effondrement*⁴, j'imaginai ce qui pourrait être l'objet d'élan nostalgiques en cas d'effondrement de nos infrastructures, moyens de transports et de communication, etc. Si beaucoup de gens se plaignent

aujourd'hui d'être en lien avec des voix de synthèse et des serveurs vocaux interactifs, il se pourrait que certaines personnes les regrettent un jour si l'accès à ces technologies venait à disparaître. Une séquence du spectacle nous propulse donc en 2050, époque pour laquelle j'envisage ouvertement de proposer un simulacre de SVI pour apaiser la souffrance nostalgique de mes contemporaines et de mes contemporains. Pour cette séquence, une personne du public est invitée à composer mon numéro de téléphone (relié à une enceinte afin d'assurer la diffusion sonore de notre conversation dans la salle) pour simuler l'appel de quelqu'un qui, en 2050, souhaiterait se renseigner sur les possibilités de transport collectif entre deux villes. Au cours de l'échange téléphonique, je lui demande notamment, à la manière d'une voix de synthèse de SVI : « De quelle ville souhaitez-vous partir ? » ou « Vers quelle ville souhaitez-vous aller ? » et sur la réponse au nom de ville, afin de mettre en avant la non-empathie de la machine, je réplique aussitôt : « Je suis désolé, cette ville n'existe plus. De quelle ville souhaitez-vous partir ? ». Au fil de l'appel, je propose des horaires de trajets en pédibus ou carriole à cheval, j'invite « les voyageur·euses à se munir de rations d'eau », etc. Et je termine la communication en raccrochant brusquement après un expéditif « Au revoir ! », expliquant que cela est indispensable pour vraiment donner l'illusion de la machine et apaiser l'élan nostalgique des personnes qui feront appel à mes services.

Ainsi, mon expérience de l'imitation d'un SVI pour des interactions à distance m'a permis de créer une séquence entière dans un spectacle destiné à être présenté face à un public en salle.

3. ARBORESCENCES, INTERACTIONS ET HOSPITALITE

Au-delà de la joie que j'ai pu éprouver au cours de ces expériences liées à la conception et à la mise en place de serveurs vocaux, je me suis rendu compte que la construction en arborescence d'un SVI est semblable à celle d'un pas-à-pas, d'un tutoriel permettant de faire comprendre à un public les possibilités d'interaction avec un dispositif au fur et à mesure de son usage. Il y a peut-être là un autre intérêt pour les artistes de théâtre à s'inspirer de l'écriture en arborescence.

Les formes théâtrales actuelles sont d'une grande diversité, et elles impliquent des usages, des codes, des possibilités d'interaction entre les interprètes et le public, qui sont également d'une grande variété, quitte à décontenancer des personnes peu habituées au théâtre. Concernant la question du renouvellement des publics et l'attractivité du spectacle vivant, une piste de recherche pourrait donc consister à ne pas seulement inviter des

² Alternative au QR Code : <http://www.galoche.online/emission/76/2020-07-04-14-30-serveur-vocal-humain>

³ Les mots ici en italique sont, dans mon script, surlignés en bleu. Il s'agit d'un repère visuel me rappelant

d'abaisser le ton de la voix pour créer un effet de décrochage et ainsi donner l'impression qu'il s'agit d'un insert audio issu d'un autre enregistrement.

⁴ Spectacle produit par la compagnie La TraverScène, créé le 11 septembre 2020 à La Goguette, à Clermont-Ferrand.

spectateurs et des spectatrices à rejoindre des artistes en salle ou ailleurs mais de repenser cet accueil comme un processus se déroulant tout au long de la représentation, afin que l'expérience même de ce qui est vécu comme une interaction, soit vécu comme une interaction positive. Il s'agirait alors de travailler sur ce que l'on pourrait appeler l'hospitalité d'une forme théâtrale.

Du 30 septembre au 2 octobre 2022, lors de ma participation à un laboratoire sur le thème de l'attractivité des formes de spectacle vivant, organisé au Théâtre du Minotaure de Béziers par le SNMS et les Chorégraphes associés, la chorégraphe Micheline Lelièvre avait invité le concepteur de jeux vidéo Fabien Delpiano à nous exposer sa manière de travailler. Selon lui, le succès d'un jeu repose essentiellement sur sa capacité à susciter de l'engagement, et donc à rendre palpable la nécessité d'une véritable implication des joueurs et joueuses pour y insuffler de la vie. Pour cela, il lui semble indispensable de prévoir une rétroaction (*feedback*) pour absolument chacune des actions des joueurs et joueuses dans le jeu (qu'il s'agisse de déclencher l'apparition d'un texte, de faire se mouvoir un personnage à l'écran, etc.), quitte à leur signifier que l'action effectuée n'est pas prévue et est inutile (avec par exemple une vibration de la manette de jeu). Ainsi, même en l'absence de tutoriel en amont, le fait même de jouer permet d'explorer les codes de jeu en bénéficiant d'un véritable accompagnement à travers ces différentes rétroactions, d'investir une place depuis laquelle la personne qui joue se sent entendue, prise en compte, accueillie.

L'écriture en arborescence - notamment utilisée pour les SVI - pourrait servir d'outil aux artistes de théâtre qui le souhaitent, à la fois pour la conception en interne de leurs différentes interactions avec le public, mais aussi pour inventer la manière de leur expliciter les codes et usages. L'écriture en arborescence leur permettrait alors de concevoir les différentes rétroactions possibles en réponse aux comportements du public, afin d'explicitement « contractualiser » l'interrelation proposée, en accord avec la dramaturgie de l'œuvre et sa progression tout au long de la représentation.

Au-delà de constituer un sujet de fond ou une couleur supplémentaire dans la palette de jeu théâtral, les arborescences et les voix de synthèse des serveurs vocaux interactifs pourraient aussi enrichir la réflexion et la pratique liées aux modalités mêmes de relation entre des artistes, une forme théâtrale et un public.

4. REFERENCES

[1] Anzalone, Salvatore, Giulia Filacanapa et Erica Magris. "RobotAct - Co-crétions Humain-Robot : une recherche-crétion interdisciplinaire entre théâtre et robotique". In *JIT 2022: 2èmes Journées d'Informatique Théâtrale*, Lyon, France, 10 et 11 octobre 2022.

[2] Plassard, Didier. "Les scènes de l'intranquillité". In : *Puck : la marionnette et les autres arts* 20 (2014), p. 11-16, cité par Anzalone, Salvatore, Giulia Filacanapa et Erica Magris.