

CFA/VISHNO 2016

Le 'voice design' dans le cinéma de science-fiction : le cas du film RoboCop (1987)

C. Vincent

Laboratoire SFL (CNRS/Paris 8), 59/61 rue Pouchet, 75017 Paris, France
coralie.vincent@cnrs.fr



LE MANS

La science-fiction est réputée être un formidable moyen d'explorer notre monde. Le cinéma de science-fiction permet quant à lui de représenter les atmosphères visuelles et sonores de futurs possibles et de délivrer, en retour, un regard critique sur le présent et ses évolutions. Nous nous proposons de présenter ici l'exploration esthétique réalisée sur la voix du personnage principal du film *RoboCop* (Verhoeven, 1987), particulièrement intéressant à notre sens : c'est en effet l'un des premiers du mouvement cyberpunk à mettre en scène un homme augmenté qui oscille entre les statuts d'humain et de robot. C'est principalement la *persona vocale* qui est prise en considération dans cette étude. Notre analyse montre que, du niveau sémantique aux effets audio en passant par la prosodie, tous les moyens à disposition sont mis en œuvre pour instiller chez le spectateur, via la voix, le doute quant à la part d'humanité du personnage.

1 Introduction

La science-fiction est connue pour être un formidable moyen d'explorer notre monde [1]. Le cinéma de science-fiction permet quant à lui de représenter les atmosphères visuelles et sonores de futurs possibles et de délivrer, en retour, un regard critique sur le présent et ses évolutions. Nous nous proposons de présenter ici l'exploration esthétique réalisée sur la voix du personnage principal du film *RoboCop* (Verhoeven, 1987), particulièrement intéressant à notre sens : c'est en effet l'un des premiers du mouvement cyberpunk à mettre en scène un homme augmenté, résultat du "recyclage" d'un corps humain en robot chez qui la condition humaine ressurgit. C'est principalement la *persona vocale* — de *per-sonare*, parler à travers —, au sens défini par Tagg [2] qui est prise en considération pour cette étude : 'any aspect of personality as shown to or perceived by others through the medium of either prosody [or of the singing voice]'.
Le problème considéré ici est donc, en quelque sorte, l'inverse de celui auquel sont confrontés les chercheurs en synthèse vocale : quels sont les moyens mis en œuvre pour qu'un être humain soit identifié, par sa voix, à un robot ? (à l'opposé de : comment faire pour qu'une voix de synthèse soit la plus "humaine" possible ?). Comment les différentes personnes impliquées dans le contenu "sonore" d'un film (scénariste, metteur en scène, acteur, monteur dialogues/son et mixeur) contribuent-elles au rendu final à chaque niveau de l'écriture cinématographique ?

Le problème considéré ici est donc, en quelque sorte, l'inverse de celui auquel sont confrontés les chercheurs en synthèse vocale : quels sont les moyens mis en œuvre pour qu'un être humain soit identifié, par sa voix, à un robot ? (à l'opposé de : comment faire pour qu'une voix de synthèse soit la plus "humaine" possible ?). Comment les différentes personnes impliquées dans le contenu "sonore" d'un film (scénariste, metteur en scène, acteur, monteur dialogues/son et mixeur) contribuent-elles au rendu final à chaque niveau de l'écriture cinématographique ?

2 Méthodes

2.1 Corpus

2.1.1 Le film *RoboCop*

Après avoir été grièvement blessé lors d'une mission, l'officier de police Alex Murphy (*Cop*) est réincarné dans un corps cybernétique (exosquelette de métal), sa mémoire ayant été purgée de tous ses souvenirs. Dans un premier temps, cet hybride homme-machine (*Robo*) se révèle particulièrement efficace pour lutter contre le crime mais, suite à un rêve et après avoir reconnu, lors d'une intervention, l'un des malfrats qui l'avaient abattu, il est assailli de doutes, reconquérant ainsi peu à peu son statut d'être humain (*RoboCop*), et éprouvant également douleur, tristesse et désir de vengeance.

Le corpus est constitué des 110 répliques des trois phases du même personnage joué par l'acteur Peter Weller (voir Figure 1), soit en tout 129 secondes de parole, réparties ainsi :

- *Cop* : 50 secondes (44 répliques)

- *Robo* : 51 secondes (39 répliques)
- *RoboCop* : 28 secondes (27 répliques)



FIGURE 1 – Les trois apparences du personnage de RoboCop : à gauche le *Cop* (policier), au centre le *Robo* et à droite l'hybride homme-machine *RoboCop*.

2.1.2 Le témoignage de l'acteur Peter Weller

En complément des dialogues du film, une courte interview de l'acteur Peter Weller évoquant la manière dont il a travaillé avec le réalisateur Paul Verhoeven sur la voix du personnage a été utilisée [3].

2.2 Pré-traitement

Le DVD du film (publié par MGM) a permis d'obtenir des fichiers informatiques vidéo et audio. Le canal central audio qui contient habituellement les dialogues a été extrait des pistes 5.1 (48 kHz, codec AC-3 à 448 kbps). Puis il a été exporté en PCM wav, 16 kHz, 16 bits afin d'être exploitable dans les logiciels de traitement du signal utilisés par la suite (notamment SPPAS). Du fait de l'encodage d'origine en AC-3 (compression destructive), nous avons été confrontés à une limitation pour les analyses formantiques. De la même manière, la parole interférant à plusieurs reprises avec les bruitages, la musique ou une longue réverbération, nous avons rencontré des difficultés dans le traitement du signal.

Les sous-titres (fichier .srt) ont été importés dans ELAN [4] après une mise en forme adaptée, au format .csv. Cet import a permis d'obtenir une première segmentation relativement précise des dialogues. Les énoncés à analyser ont été déplacés sur une même piste puis exportés au format Praat afin que les bornes soient précisément placées à la main.

Deux types de pauses ont été annotés :

- les pauses inter-lexicales, notées # (pour le bon fonctionnement du logiciel SPPAS), situées entre deux mots d'une même phrase [5]
- les pauses inter-phrastiques, notées [pause], situées entre deux propositions ou phrases d'un monologue.

Un alignement a ensuite été automatiquement réalisé grâce au logiciel SPPAS [6], à l'issue duquel une vérification et une éventuelle correction manuelle ont été effectuées.

Pour la syllabification, nous avons utilisé le dictionnaire gratuit disponible en ligne Wordsmyth – Free On-Line English Dictionary (<http://www.wordsmyth.net/>).

2.3 Analyses

Les mesures ont été réalisées avec le logiciel Praat [7] :

- les durées avec le script “duration logger.praat” écrit par K. Crosswhite et modifié par M. Antoniou
- les fréquences fondamentales d'abord évaluées avec la fonction “Extract visible pitch contour” puis corrigées manuellement et regroupées automatiquement avec le script “pitch logger.praat” développé pour l'occasion
- les valeurs d'intensité mesurées avec la méthode “mean-energy” directement disponible dans Praat.

Les résultats présentés ci-dessous sont organisés par paramètre et par persona : *Cop*, *Robo* et *RoboCop* qui est subdivisé en deux catégories : la parole standard et les cris (*RCop-cri*) au nombre de trois dans l'avant-dernière scène du film — au cours de laquelle les malfrats sont méthodiquement tués.

3 Résultats

3.1 Aspects sémantiques

Même si un mélange des différentes fonctions du langage apparaît dans les répliques de *Cop*, les fonctions référentielle et phatique sont prédominantes. Les interjections sont présentes (et nombreuses), témoignant d'une spécificité humaine : la spontanéité (toute relative, bien évidemment, puisqu'il s'agit de dialogues écrits !).

Néanmoins, à partir du moment où le personnage devient *Robo*, la fonction conative prend le pas : la majorité des paroles de la machine consiste en routines centrées sur l'énonciation de règles (“you are under arrest”) et intègre quelques rares verbes modaux — d'habitude typiquement utilisés quand le point de vue de l'énonciateur est convoqué : “I will notify [...]”, “how can I help you [...]?”. Le *Robo* se définit par sa fonction : “yes, I am a cop”. Son lexique est globalement limité, et seule une petite entorse est faite au statut figé et contrôlé de son vocabulaire : il utilise le mot **creep** (registre argotique) lors d'une intervention.

Enfin, la réplique prononcée par *RoboCop* quand il recherche son identité et se remémore sa famille, a une fonction fortement émotive : “I can feel them... but I can't remember them” (avec donc recours aux verbes modaux). C'est cette phrase qui est au centre du témoignage de Peter Weller dans l'interview.

Les temps des verbes utilisés témoignent aussi d'une évolution entre *Robo* et *RoboCop* : alors que *Robo* ne parle qu'au présent et au futur, la première phrase de *RoboCop* est au passé : “did you bring the gun?”. On remarquera que dans l'une des répliques qui suivent, il mentionne le corps humain et sa fonction digestive (“I'm not hungry”), caractéristique des êtres vivants, annonçant les prémices d'un retour à l'humanité.

3.2 Prosodie

Après avoir considéré quelques aspects sémantiques, nous nous intéressons maintenant à la prosodie des trois personas qui apparaissent dans le film *RoboCop*.

3.2.1 Durée

La longueur moyenne des énoncés est sensiblement la même entre les trois personas, mais l'écart-type est très différent (voir Tableau 1) : la longueur des répliques de *Robo* varie peu, au contraire de celle de *Cop*. Quant à *RoboCop*, il est entre les deux, traduisant un état intermédiaire entre flexibilité de la parole spontanée humaine (*Cop*) et rigidité du carcan d'un programme informatique (*Robo*).

TABLEAU 1 – Longueur des énoncés.

Persona	Longueur	Ecart-type
Cop	5.8 syll.	7.1 syll.
Robo	5.8 syll.	3.3 syll.
RoboCop	5.9 syll.	4.5 syll.

Pour les débits syllabiques, la valeur mesurée pour *RoboCop* est plus élevée que celle de *Cop* et plus faible que celle de *Robo* (voir Tableau 2). Et les cris de *RoboCop*, qui sont des dissyllabes assez longuement tenus, se différencient des autres éléments par leur faible débit.

Les pauses sont, quant à elles, traitées très différemment entre *RoboCop* d'une part et *Cop* et *Robo* d'autre part : les pauses inter-phrases sont près de 3 fois plus longues chez *RoboCop* et il ne produit pas de pause inter-lexicale, comme si ses paroles étaient émises en un souffle ininterrompu, séparées entre elles par d'importantes “respirations”, en particulier dans la séquence à contenu hautement émotif.

Enfin, deux éléments sortant de l'ordinaire peuvent être mentionnés ici :

- une disfluente (typiquement humaine) quand *Cop* mentionne son fils (“my... my son”)
- une pause inter-lexicale dans une phrase prononcée par *Robo* (“or there will be... trouble”), pour laquelle nous n'avons pas d'explication, si ce n'est le fait de surprendre le spectateur en insérant une “hésitation” dans un discours par ailleurs parfaitement régulier.

TABLEAU 2 – Débit syllabique et durée des pauses inter-lexicales et inter-phrases.

Persona	Débit	Durée #	Durée [p]
Cop	5.5 syll./s	0.142 s	0.960 s
Robo	4.4 syll./s	0.254 s	0.951 s
RoboCop	5.2 syll./s		2.917 s
RCop-cri	2.8 syll./s		

3.2.2 Fréquence fondamentale

Les comportements vocaux de *Cop* et de *Robo* sont extrêmement semblables quant à la fréquence fondamentale : leurs moyennes et écarts-types sont identiques, et leurs ambitus très comparables (voir Tableau 3). En revanche, *RoboCop* a lui une voix globalement plus grave (du moins quand il est en registre parlé), ce qui correspondrait a priori à un état moins tonique et à la manifestation d'un sentiment de tristesse [8].

TABLEAU 3 – Fréquence fondamentale F0.

Persona	F0 min	F0 max	Moyenne	Ecart-type
Cop	83 Hz	268 Hz	132 Hz	32 Hz
Robo	76 Hz	274 Hz	132 Hz	33 Hz
RoboCop	79 Hz	241 Hz	117 Hz	30 Hz
RCop-cri	228 Hz	451 Hz	398 Hz	45 Hz

Enfin, on remarquera que les trois cris poussés vers la fin du film se situent dans la gamme de fréquences des cris néonataux [9].

3.2.3 Intensité

Alors que l'intensité moyenne de *Cop* et celle de *RoboCop* sont assez comparables, *Robo*, digne représentant de l'autorité, se distingue, lui, par une intensité de plus de 3 dB plus élevée (voir Tableau 4). Et le cri paroxystique de *RoboCop* dans l'avant-dernière scène du film est en moyenne encore 3 dB plus fort que la voix autoritaire de *Robo*.

TABLEAU 4 – Intensité (mean-energy).

Persona	Int. min	Int. max	Moyenne
Cop	41.7 dB	77.5 dB	68.7 dB
Robo	43.6 dB	79.8 dB	72.0 dB
RoboCop	36.2 dB	77.2 dB	68.2 dB
RCop-cri	51.6 dB	80.0 dB	75.6 dB

Le minimum d'intensité est atteint par *RoboCop*, sur l'une des phrases de la séquence chargée d'émotions : "Murphy had a wife and son".

3.3 Effets audio

Afin de parfaire l'illusion d'hybridation homme-machine déjà constatée pendant nos analyses sémantique et prosodique, deux effets sonores perceptibles sont ajoutés sur la voix de *Robo* quand il est parfaitement fonctionnel (en plus des traditionnels compresseurs de dynamiques, filtres, etc.) :

- un effet fréquentiel *DigiTech Stereo Chorus/Flanger*

(SCF) [10], qui lui confère un caractère synthétique et métallique

- un effet de réverbération, symbole éprouvé de l'autorité [11], dès que le cyborg est en mission pour lutter contre le crime. Ces deux effets disparaissent dès lors que *RoboCop* retire son casque.

Nous avons par ailleurs remarqué qu'à cause de ces effets, la détection automatique de fréquence fondamentale sur Praat était systématiquement prolongée (pendant la réverbération), et que le résultat calculé était quasi-systématiquement trop bas. De ce fait, la fin de chaque courbe détectée a été corrigée manuellement pour *Robo*.

3.4 Récapitulatif (hors cri)

Les différences et similarités entre les voix de *Cop*, *Robo* et *RoboCop* peuvent donc être résumées ainsi :

TABLEAU 5 – Différences et similarités de la voix de *RoboCop* avec les voix de *Cop* et de *Robo*, par paramètre.

Paramètre	Cop	Robo	RoboCop
Ecart-type énoncés	+	-	intermédiaire
Débit syll.	+	-	intermédiaire
Pauses inter-phrases	=	=	+ longues
F0	=	=	+ grave
Intensité	=	+ forte	=
Effets sonores	aucun	SCF	aucun

4 Conclusion

Comme l'indique Jouhaneau dans un chapitre où les voix de robots au cinéma sont abordées [12] : "L'importance de la personnalisation du robot apparaît également dans un détail significatif, sa voix ; jamais totalement synthétique, elle traduit toujours la faiblesse ou la vulnérabilité du personnage". Il est donc primordial que la voix de *RoboCop* soit une hybridation entre les voix de *Cop* (être humain) et de *Robo* (machine) pour certains paramètres. Mais le fait qu'elle outrepassse le caractère humain de la voix de *Cop* est également une nécessité : c'est au moment où son apparence incertaine (corps et crâne robotiques, visage humain) est la plus propice à le faire basculer dans la "vallée de l'étrange" [13] que le subterfuge d'une voix hyper émotionnelle convainc le plus efficacement possible le spectateur de l'humanité en devenir de *RoboCop*. Car une voix émotionnelle, implique forcément un souffle de vie, un corps [14] et donc une part d'humanité.

Et cette humanité retrouvée atteint son paroxysme dans la dernière réplique : à la question de son nom, *RoboCop* répond par "Murphy" (auto-désignation qui signifie une conscience de soi [15], aboutissement d'un processus d'humanisation).

Fondée sur un éventail infini de possibilités, du niveau sémantique aux effets audio en passant par la prosodie, nous avons montré que la voix joue un rôle essentiel dans l'appréciation que nous pouvons avoir de l'humanité d'un personnage. Il pourrait néanmoins être judicieux d'élargir une prochaine analyse à l'étude du *sound design* (bruitages des corps — qu'ils soient de chair ou métalliques), à l'utilisation de la musique ou à l'étude des mimiques, gestes et postures. Il pourrait enfin être intéressant de voir comment tous ces effets ont été traités dans le remake de 2014.

Références

- [1] R. Lehoucq, Un peu de fiction dans votre science. *Les Ernest — 15 minutes pour changer notre vision du monde*. Consulté à l'adresse <http://www.les-ernest.fr/un-peu-de-fiction-dans-votre-science/> (janvier 2014).
- [2] P. Tagg, *Music's Meanings*. Mass Media Music Scholar's Press (MMMSP), Larchmont, NY, USA (2013).
- [3] John at the Movies, Peter Weller Talks RoboCop's Voice. *YouTube*. Consulté à l'adresse https://www.youtube.com/watch?v=71BgjfEsU_E (2013).
- [4] H. Brugman & A. Russel, Annotating multi-media / multimodal resources with ELAN. In *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, pp. 2065-2068 (2004).
- [5] B. Zellner, Pauses and the temporal structure of speech, in E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition*, pp. 41-62. Chichester : John Wiley (1994).
- [6] B. Bigi, SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, International Society of Phonetic Sciences **111-112**, 55-69 (2015).
- [7] P. Boersma & D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.0.14, disponible sur <http://www.praat.org/> (2016).
- [8] K. R. Scherer, Expression of emotion in voice and music. *Journal of Voice* **9(3)**, 235-248 (1995).
- [9] B. Zei, Au commencement était le cri. Note sur la voix humaine, son importance et ses infinies subtilités. *Le temps stratégique* **66**, 96-105 (1995).
- [10] S. Flick, Sound Effects Design for "RoboCop". *Hollywood Lost and Found*. Consulté à l'adresse <http://www.hollywoodlostandfound.net/sound/robocop/> (1988).
- [11] P. Doyle, *Echo and Reverb : Fabricating Space in Popular Music Recording, 1900-1960*. Middletown, CT : Wesleyan (2005).
- [12] J. Jouhaneau, De la réalité scientifique au réalisme cinématographique. In A. Martinet (éd.), *Le cinéma et la science*, pp. 224-247. Paris : CNRS Éditions (1994).
- [13] M. Mori, Bukimi no tani (The uncanny valley, trad. K. F. MacDorman et T. Minato). *Energy* **7(4)**, 33-35 (1970).
- [14] C. Gillie-Guilbert, "Et la voix s'est faite chair...". Naissance, essence, sens du geste vocal. *Cahiers d'ethnomusicologie* **14**, 3-38 (2001).
- [15] A. Morgenstern, *Un JE en construction : genèse de l'auto-désignation chez le jeune enfant*. Paris : Editions OPHRYS (2006).