



HAL
open science

Les apports des Bell Labs à l'ingénierie de la voix (chantée)

Coralie Vincent

► **To cite this version:**

Coralie Vincent. Les apports des Bell Labs à l'ingénierie de la voix (chantée). Revue de Laryngologie Otologie Rhinologie, 2015, 136 (5), pp.193-196. hal-01473594

HAL Id: hal-01473594

<https://hal.science/hal-01473594>

Submitted on 7 Sep 2021

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.

Introduction

Les liens entre sciences et voix sont multimillénaires, comme l'attestent de récentes découvertes relatives au théâtre grec. Que ce soit à travers l'architecture de ces lieux [1], la disposition particulière de vases résonnants en leur sein [2] ou l'utilisation, par les acteurs, de masques aux propriétés acoustiques spécifiques [3], les apports des sciences semblent avoir été présents à de nombreux niveaux dans le but d'optimiser l'intelligibilité de la parole.

Dans cet article, nous nous proposons de montrer dans quelle mesure cette interaction art-sciences est toujours d'actualité et, en particulier, comment les Bell Labs, laboratoires de recherche fondés en 1925 à New York et spécialisés dans les domaines des télécommunications, ont pu révolutionner l'esthétique de la voix (chantée) par le biais de trois avancées technologiques majeures.

- 1) Larynx artificiel, sonovox et talkbox
 - a. Origine et principe

Breveté par les Bell Labs en 1929 [4], le larynx artificiel est un appareil médical conçu pour fournir une prothèse vocale à des patients laryngectomisés : une anche métallique vibre à l'intérieur d'un tube connecté entre bouche et stomie du locuteur (voir Figure 1). La fréquence de vibration de l'anche n'étant pas modifiable, le son est monocorde, mais ce procédé permet néanmoins de compenser le handicap engendré par l'ablation partielle ou totale du conduit vocal.

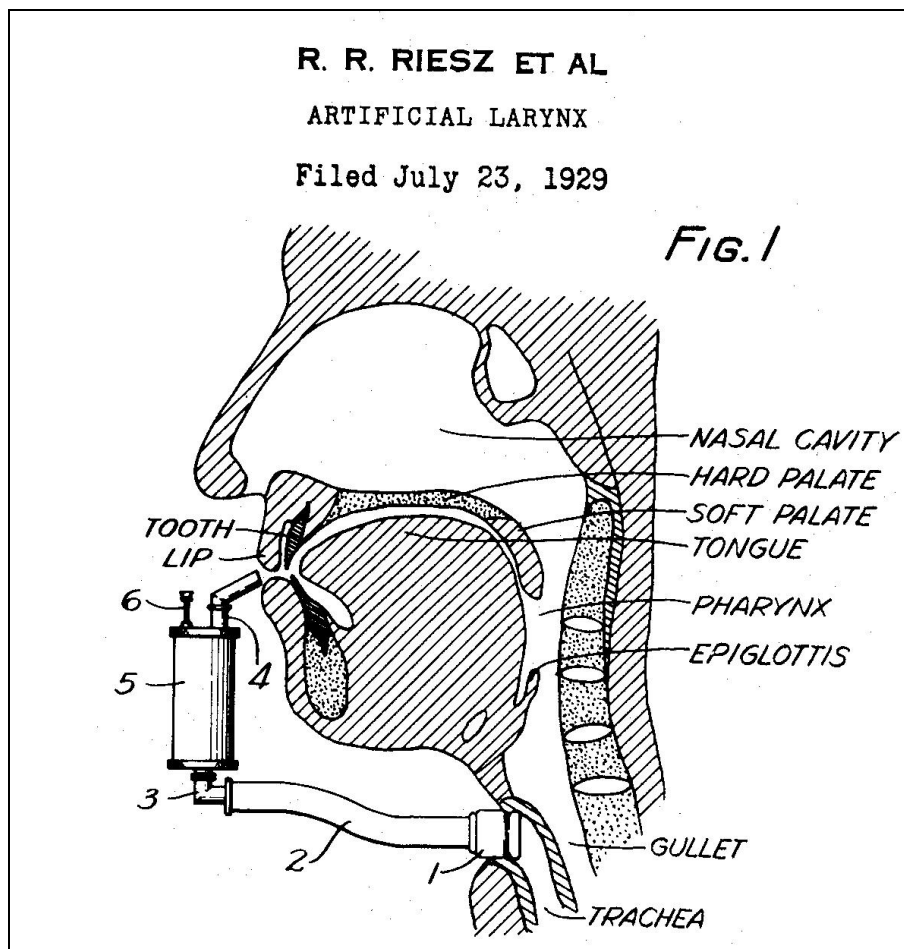


Figure 1 : Schéma d'utilisation d'un larynx artificiel par un patient stomisé, d'après [4]

En 1960, une version électronique de ce dispositif, l'électrolarynx, est inventée : l'anche métallique est remplacée par un générateur électronique placé contre la gorge qui ne nécessite plus de stomie [5]. Cette même année, Fant formalise la théorie sur laquelle reposent ces appareils : la théorie source-filtre [6]. Elle stipule que les signaux vocaux (parole ou chant) résultent de la modulation, par un filtre, d'un flux d'air pulsé. Ainsi, la source créée par les poumons et les plis vocaux est filtrée par le conduit vocal : en fonction de la configuration de ce dernier, certaines fréquences sont amplifiées alors que d'autres sont atténuées, modifiant ainsi de manière perceptible le signal produit par la source glottique (voir figure 2).

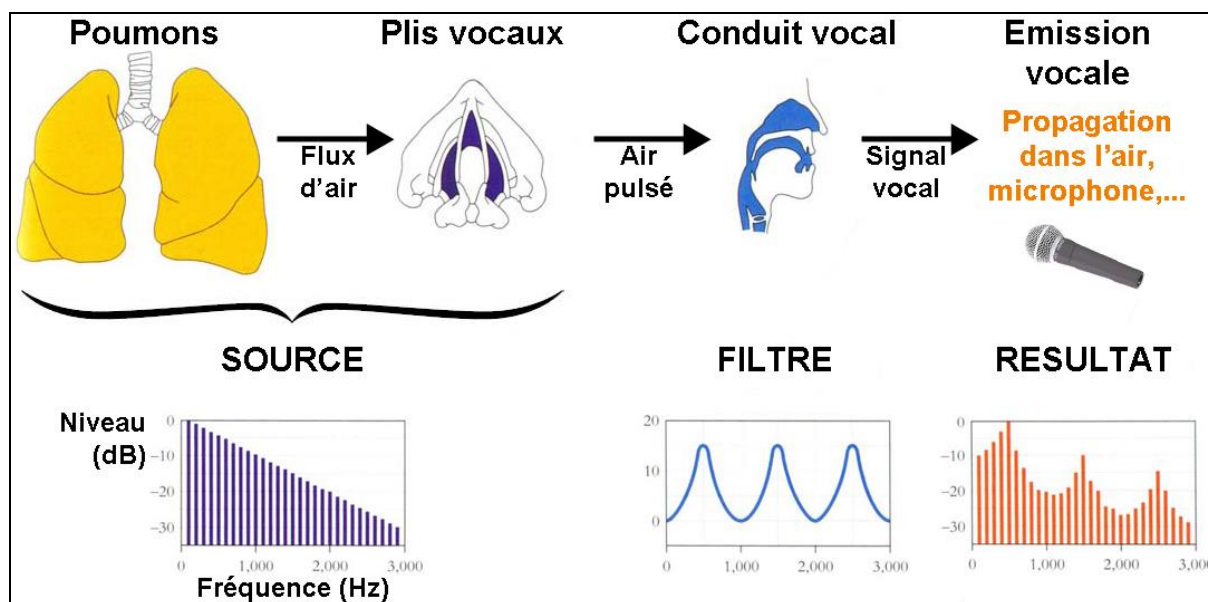


Figure 2 : Théorie source-filtre, d'après [7]

Pour résumer, ce qui est mis en œuvre dans le larynx artificiel et les procédés décrits ci-dessous (sonovox et talkbox), c'est le remplacement de la source glottique (poumons et plis vocaux) : dans le cas du larynx artificiel, une anche (respectivement un générateur électronique) se substitue à cette source glottique dans la version mécanique (respectivement électronique) de cet instrument.

Le principe de cette innovation médicale est, quelques années après l'invention du larynx artificiel, exploité par l'industrie musicale avec le sonovox puis la talkbox. Avec ces outils, la source glottique est remplacée par de petits hauts-parleurs de contact – dans le sonovox – et par un tube branché à un circuit d'attaque de haut-parleur – dans la talkbox (voir Figure 3). Les sonorités produites sont proches de celles de la parole d'origine, le conduit vocal modulant le contenu spectral de la source de substitution, mais elles possèdent un contenu spectral différent. Dans le cas du sonovox comme de la talkbox, la gorge ne produit pas de son. L'utilisateur mime silencieusement l'articulation des paroles avec son conduit vocal, qui est donc utilisé pour moduler la source glottique de substitution.

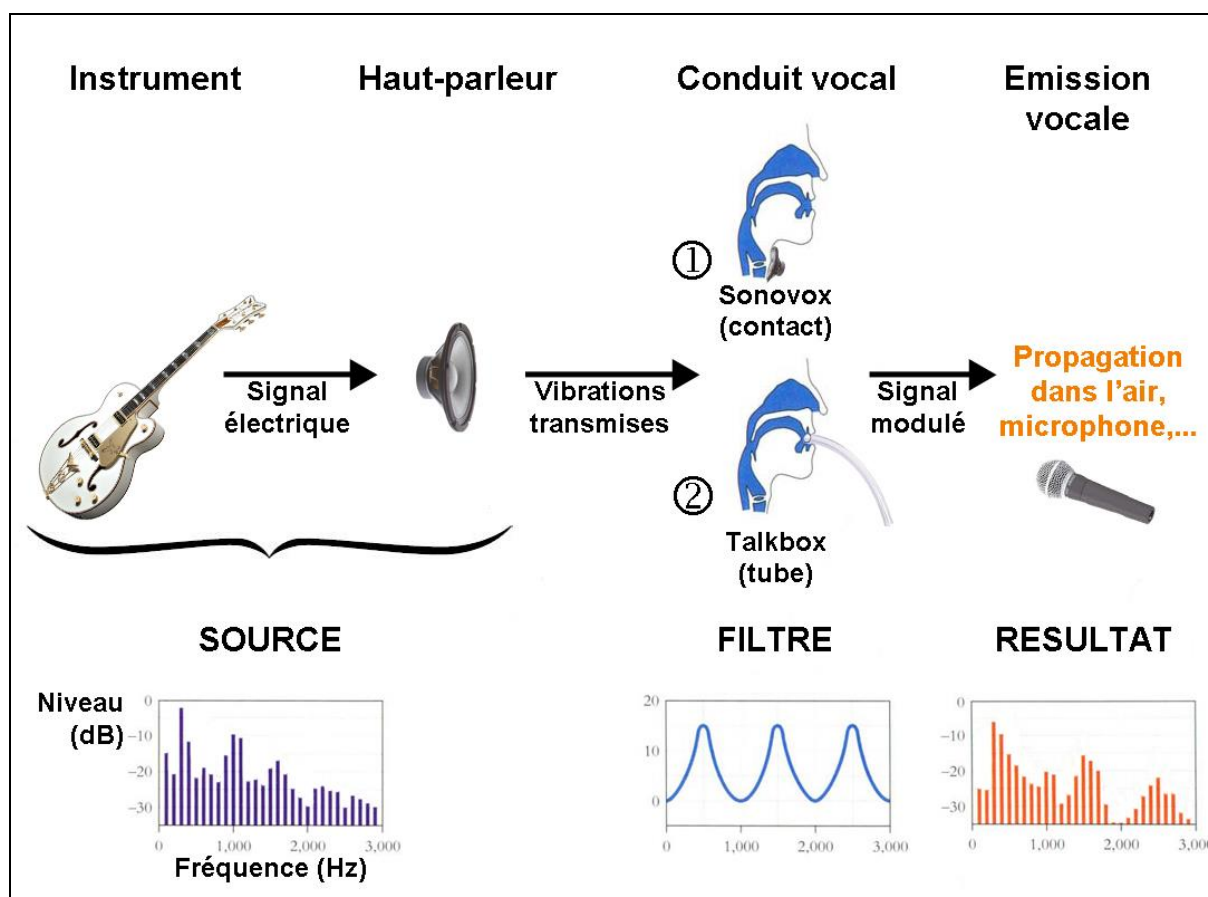


Figure 3 : Principes de fonctionnement de la sonovox ① et de la talkbox ②, d'après [7]

b. Historique et utilisations

Le sonovox est inventé par Gilbert Wright, ingénieur et opérateur radio, en 1939. Cette idée lui serait venue en se rasant [8] :

Gilbert Wright was meditatively scratching the bristly whiskers on his Adam's apple & he noticed that queer sounds came out of his mouth. When he silently mouthed words, the sounds caused by scratching his whiskers were formed into words.

Ce procédé est très utilisé dans les années 1940, dans des publicités et des dessins animés (pour la voix du train dans le film « Dumbo » [9], par exemple). Le résultat, considéré comme susceptible d'impressionner les jeunes auditeurs, est parfois précédé d'un message de mise en garde. Le sonovox peut être montré à l'écran, dans une perspective didactique [10] ou de pur divertissement [11].

La « Talking Steel Guitar », précurseur de la talkbox, est, quant à elle, conçue par Bill West en 1963 : diffusant le son d'une guitare, un circuit d'attaque de haut-parleur est fixé à un entonnoir lié à la bouche du performeur par un tube plastique. Peu puissant en terme d'intensité sonore, ce procédé peut uniquement être utilisé en studio (et non en *live*). Il est pour la première fois entendu sur le titre « Lock, Stock and Teardrops » de Roger Miller & Pete Drake (1963) [12]. Dès 1969, il est vendu sous l'appellation de The Bag, et est utilisé, entre autres, par les musiciens Mike Pinera puis Jeff Beck.

C'est encore Bill West qui, en 1971, aurait créé la talkbox utilisée sur le titre « Rocky Mountain Way » de Joe Walsh [13]. Le procédé aurait, par la suite, été amélioré et breveté par Bob Heil, fondateur de

la société Heil Sound qui commercialise à grande échelle la Heil Talk Box. Le chanteur Peter Frampton, virtuose de la talkbox, a contribué à la rendre populaire dans les années 1970. Plus récemment, on a pu par exemple entendre cet effet sur les titres « California Love » de 2Pac & Dr. Dre (voix de Roger Troutman) [14] et sur « Digital Love » de Daft Punk [15].

2) Vocodeur

a. Principe

Le vocodeur est la deuxième innovation technologique créée par les Bell Labs toujours largement utilisée en production musicale.

L'idée de base de cette invention est d'analyser, en temps réel, les caractéristiques spectrales d'un signal filtré par bandes de fréquences afin de séparer, à peu près, l'excitation (la vibration des plis vocaux) et la réponse (les résonances) du conduit vocal (voir Figure 4). Le son d'un instrument de remplacement est ensuite envoyé à une banque de filtres passe-bande de synthèse qui accentuent les mêmes gammes de fréquence que les filtres d'analyse.

En musique, l'enveloppe mesurée sur le modulateur – typiquement, une voix – est appliquée au porteur – typiquement, un synthétiseur :

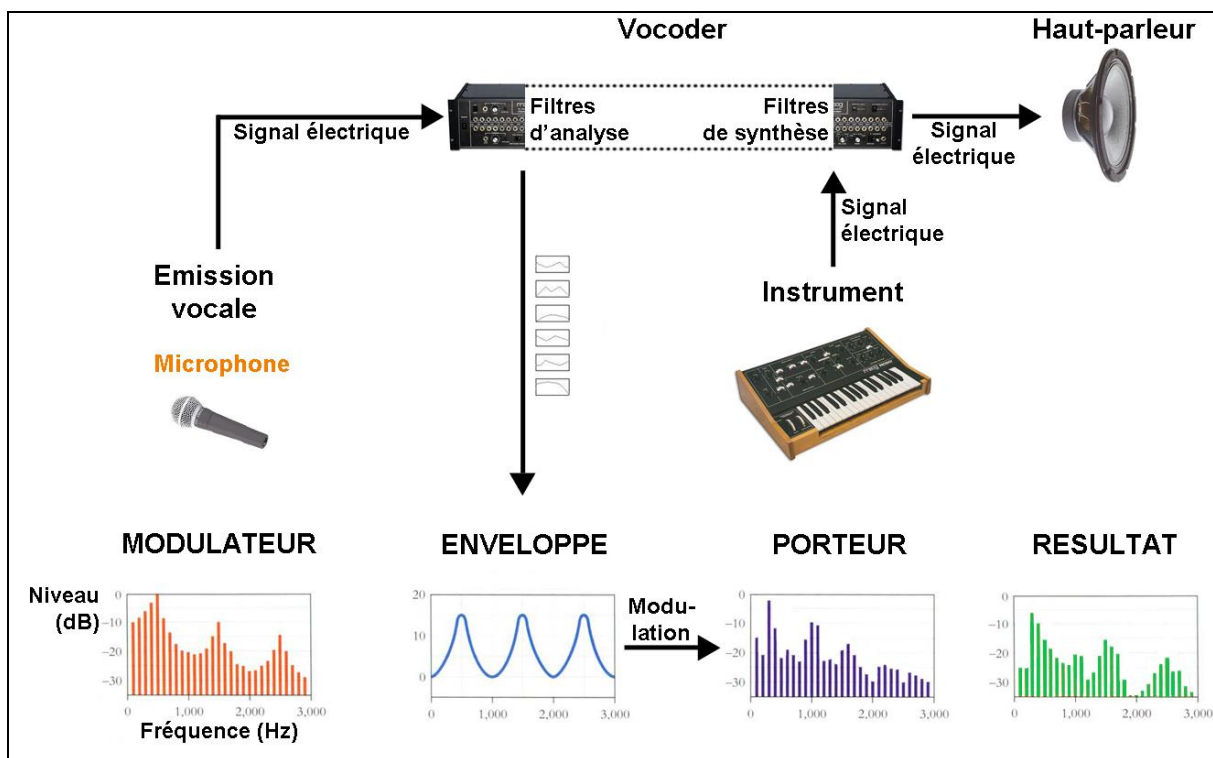


Figure 4 : Principe de fonctionnement du vocodeur, d'après [7]

b. Historique et utilisations

Le vocodeur (*vocoder*) est breveté par Homer Dudley (Bell Labs) sous l'intitulé *signal transmission* en 1935 [16]. Une implémentation sur ordinateur est effectuée en 1966 par Flanagan & Golden [17]. Mais ce n'est qu'en 1970 que la première utilisation musicale est réalisée, par Bruce Haack [18].

Depuis cette date, il a été adopté à de nombreuses occasions pour créer des voix fictives, de robot par exemple chez Kraftwerk [19] et Daft Punk [20].

3) Synthèse vocale

Dans les deux parties précédentes, nous avons vu que :

- Le conduit vocal peut être utilisé pour moduler une source glottique de substitution (larynx artificiel, sonovox et talkbox).
- Le signal acoustique produit par une voix peut être analysé de manière à appliquer sur une source sonore de synthèse un filtrage similaire à celui appliqué par le conduit vocal (vocodeur).

Nous nous intéressons maintenant à la modélisation du signal produit par une voix chantée afin de la synthétiser.

a. Principe

Les techniques de synthèse de la voix chantée sont nombreuses et variées. Le but de cet article n'étant pas d'en faire un inventaire détaillé, on se référera aux articles de Cook [21] et Rodet [22] pour plus d'informations à ce sujet.

Nous retiendrons simplement ici que l'un des principes les plus utilisés actuellement est la synthèse par concaténation, technique qui assemble des échantillons de durées variables (mais courts, inférieurs à la seconde) de sons précédemment enregistrés.

b. Historique et utilisations

L'année 1957 voit la naissance de deux innovations majeures dans l'histoire des techniques : l'enregistrement numérique et la synthèse de sons musicaux. La première voix chantée de synthèse est émise par un ordinateur quatre ans plus tard, en 1961, dans les Bell Labs. La pièce choisie pour cette première est le morceau « Daisy Bell », programmé par John Kelly et Max Mathews [23]. Pour l'anecdote, Arthur C. Clarke, qui visite le laboratoire à ce moment-là, fasciné par les sons qu'il entend, décide de rendre un vibrant hommage à cette invention dans le film « 2001, L'odyssée de l'espace » [24] : c'est en effet cette chanson qu'entonne, dans la version originale, l'ordinateur HAL alors qu'il est progressivement débranché.

Depuis son invention, la voix de synthèse est notamment utilisée pour créer des « chimères » vocales dans la musique savante, mais aussi pour répondre à des demandes de l'industrie cinématographique : pour donner par exemple corps à des voix disparues ou irréelles (dans les films « Farinelli » en 1994 [25] ou « Les amours d'Astrée et de Céladon » en 2006 [26], auxquels a contribué l'IRCAM). Hormis dans l'évocation nostalgique d'un monde informatisé et déshumanisé (« Fitter Happier » de Radiohead [27]), ce type de voix était peu exploité dans la musique populaire. Et pour cause, l'un des points centraux du *star system* était jusqu'à récemment d'idolâtrer des êtres humains réels. Mais cette tendance est sérieusement remise en question depuis l'apparition du logiciel VOCALOID, avec notamment les débuts en 2007 de la chanteuse virtuelle Hatsune Miku [28] (dont la page Facebook dépasse à l'heure actuelle les 2,5 millions de fans). Le succès de ce personnage réside dans cette idée géniale de Yamaha : n'importe qui peut créer une chanson avec la

voix de Hatsune Miku, tant qu'il a les logiciels appropriés [29]. En quelques années, Hatsune Miku est devenue un vrai phénomène de société. Le personnage et la communauté de ses fans sont au cœur de travaux universitaires [30].

Conclusion

Il apparaît que les interactions entre science et musique sont omniprésentes dans les traitements de la voix chantée. Les laboratoires de recherche, comme les Bell Labs, y ont largement contribué. Les inventions et principes mis au jour dans cet établissement (larynx artificiel, vocodeur ou synthèse vocale) ont ainsi pu être ré-exploités dans de nombreux cadres, notamment celui de la musique populaire, comme en témoignent les centaines de titres où peuvent être entendus sonovox, talkbox et vocodeur. Il est également intéressant de noter que le principe de la théorie source-filtre a été utilisé de manière empirique pendant plusieurs dizaines d'années avant d'être formalisé.

Un autre point important est la disparition progressive du corps humain du processus de création vocale, au fil des ans : indispensable pour le larynx artificiel, il n'est plus utilisé directement par le vocodeur (le son produit est simplement analysé par le vocodeur) et enfin, dans la synthèse de voix chantée, c'est une modélisation antérieure qui sert à générer les sons perçus.

En conclusion, l'ingénierie de la voix (chantée) semble bien être devenue une composante incontournable du processus de création musicale. Et avec l'apparition du logiciel VOCALOID, la problématique du rapport au corps est devenue centrale. Il est donc légitime de se demander ce que sera la prochaine évolution. La réponse pourrait bien être un logiciel mettant en scène un avatar vocal : l'utilisateur pourrait ainsi « greffer » à sa voix la voix de chanteurs célèbres en temps réel !

Références bibliographiques, cinématographiques et discographiques

- [1] N. F. Declercq et C. S. A. Dekeyser, « Acoustic diffraction effects at the Hellenistic amphitheater of Epidaurus: Seat rows responsible for the marvelous acoustics », *J. Acoust. Soc. Am.*, vol. 121, n° 4, p. 2011-2022, avr. 2007.
- [2] R. Godman, « The enigma of Vitruvian resonating vases and the relevance of the concept for today », *J. Acoust. Soc. Am.*, vol. 122, n° 5, p. 3054-3054, nov. 2007.
- [3] A. Tsilfidis, T. Vovolis, E. Georganti, et J. Mourjopoulos, « Function and Acoustic Properties of Ancient Greek Theatre Masks », *Acta Acust. United Acust.*, vol. 99, n° 1, p. 82-90, janv. 2013.
- [4] G. W. Burchett et R. R. Riesz, « Artificial larynx », US1910966 A, 23-mai-1933.
- [5] R. E. Polk, « Artificial larynx », USD190812 S, 04-juill-1961.
- [6] G. Fant, *Acoustic Theory of Speech Production*. The Hague: Mouton, 1960.
- [7] C. Vincent, « De l'anti-pop à l'Auto-Tune™ », in *La voix chantée, entre sciences et pratiques*, N. Henrich Bernardoni, Éd. Paris: De Boeck-Solal, 2014, p. 123-142.
- [8] « Science: Sonovox », *Time Magazine*, vol. 34, n° 4, 24-juill-1939.
- [9] S. Armstrong et N. Ferguson, *Dumbo*. 1941.
- [10] A. L. Werker, H. Luske, J. Cutting, U. Iwerks, et J. Kinney, *The Reluctant Dragon*. 1941.
- [11] D. Butler, *You'll Find Out*. 1940.
- [12] R. Miller et P. Drake, *Lock, Stock and Teardrops*. RCA Camden, 1963.
- [13] J. Walsh, *Rocky Mountain Way*. ABC-Dunhill, 1973.
- [14] 2Pac, Dr. Dre, et R. Troutman, *California Love*. Death Row, Interscope, 1995.
- [15] Daft Punk, *Digital Love*. Virgin Records, 2001.
- [16] H. W. Dudley, « Signal transmission », US2151091 A, 21-mars-1939.
- [17] J. L. Flanagan et R. M. Golden, « Phase Vocoder », *Bell Syst. Tech. J.*, p. 1493-1509, nov. 1966.
- [18] B. Haack, *Electric to Me Turn*. Columbia, 1970.
- [19] Kraftwerk, *The Robots*. Kling Klang (EMI), 1978.
- [20] Daft Punk, *Around the World*. Virgin Records, 1997.
- [21] P. R. Cook, « Singing Voice Synthesis: History, Current Work, and Future Directions », *Comput. Music J.*, vol. 20, n° 3, p. 38-46, 1996.
- [22] X. Rodet, « Synthesis and Processing of the Singing Voice », in *Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgique, 2002, p. 99-108.
- [23] J. Kelly et M. Mathews, *Daisy Bell*. Bell Laboratories, US, 1961.
- [24] S. Kubrick, *2001: A Space Odyssey*. 1968.
- [25] G. Corbiau, *Farinelli*. 1995.
- [26] E. Rohmer, *Les amours d'Astrée et de Céladon*. 2008.
- [27] Radiohead, *Fitter Happier*. Parlophone, 1997.
- [28] H. Miku, *World is Mine*. Sony Entertainment, 2009.
- [29] H. Kenmochi, « Singing synthesis as a new musical instrument », in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, p. 5385-5388.
- [30] K. F. (傅嘉雯) Fu, « Playing with the voice : Hatsune Miku and vocaloid culture in contemporary Japan », Master's Thesis, University of Hong Kong, Hong Kong, 2014.