

Cover story (French)

Cover story for the paper Nadia Guerouaou, Guillaume Vaiva & Jean-Julien Aucouturier (2021) "The shallow of your smile: The ethics of expressive vocal deep-fakes", Royal Society Philosophical Transactions B

—

La société Voxpressive est une startup française issue des travaux de recherche d'un laboratoire d'informatique de Sorbonne Université, et qui utilise les technologies de l'intelligence artificielle pour modifier les émotions d'une voix en temps-réel, par exemple pour la rendre plus souriante ou moins colérique.

La société Voxpressive est en train de tester la mise sur le marché d'un produit, appelé MyVoicePlus™, constitué d'un ensemble de deux prothèses connectées permettant d'utiliser cette technologie de transformation émotionnelle de la voix pour modifier soit la voix de l'utilisateur, soit la façon dont l'utilisateur perçoit la voix de ses interlocuteurs.

Le produit MyVoicePlus™ est constitué, d'une part, d'une prothèse vocale se plaçant sur le cou de l'utilisateur, et ressemblant à un bijou. Cette prothèse, conçu en collaboration avec des médecins ORL, utilise une technique brevetée de stimulation vibrotactile (SVT; Khosravani et al. 2019) pour modifier la vibration des cordes vocales au moment où la personne parle. Lors de tests en laboratoire, l'utilisation de MyVoicePlus™ a notamment permis de rendre la voix de l'utilisateur plus joyeuse et souriante (augmentation de 70% du score de sourire perçu par l'interlocuteur), ou de la rendre moins nerveuse et tremblante (diminution de 68% du score d'anxiété perçu par l'interlocuteur).



Gauche: Tests en laboratoire du dispositif de stimulation vibrotactile de MyVoicePlus™. Droite: premier prototype du dispositif.

MyVoicePlus™ est constitué, d'autre part, d'une prothèse auditive se portant dans l'oreille, et ressemblant à des écouteurs de musique, *qui utilise une technologie de traitement du signal sonore (le vocodeur de phase wavenet; Wu et al. 2019) capable de modifier le ton émotionnel des voix qui sont entendues par l'utilisateur. Lors de tests en laboratoire, l'utilisation de MyVoicePlus™ a permis à ses utilisateurs d'entendre les voix de leurs interlocuteurs comme étant plus joyeuses ou souriantes (augmentation de 69% du score de sourire perçu par l'utilisateur de la prothèse), ou moins agressives et colériques (diminution de 73% du score d'agressivité ressenti par l'utilisateur de la prothèse).*



Prototype du dispositif auditif connecté MyVoicePlus™, embarquant une technologie brevetée de transformation de voix émotionnelle.

Utilisées ensemble, les deux prothèses de MyVoicePlus™ permettent également de créer une situation de “feedback”, où l'utilisateur entend sa propre voix avec un ton émotionnel modifié, par exemple plus joyeuse ou moins anxieuse. Dans ce cas, la modification de la voix n'est perceptible que par l'utilisateur via la prothèse auditive: l'entourage, quant à lui, ne perçoit pas la transformation et continue d'entendre la voix originale de l'utilisateur. Lors de tests en laboratoire, l'utilisation de MyVoicePlus™ en situation de feedback a permis d'avoir plusieurs effets positifs sur l'utilisateur, par exemple de devenir plus joyeux en entendant sa voix plus souriante (augmentation de 67% des émotions positives ressenties par l'utilisateur), ou de devenir plus calme en entendant sa voix moins anxieuse (diminution des symptômes anxieux de 71%).

Le développement de technologies d'intelligence artificielle pour transformer les émotions d'une voix pose un certain nombre de questions éthiques (Westerlund, 2019). La société Voxpressive est particulièrement attentive au fait de développer des bonnes pratiques pour encadrer et de n'encourager que des utilisations socialement acceptables de ses technologies. Dans le cadre de la mise sur le marché de son produit MyVoicePlus™, la société VoxPressive souhaite réaliser une enquête d'opinion sur différentes possibilités d'application de sa technologie.

Dans la suite de cette enquête, nous allons vous présenter plusieurs situations possibles dans lesquelles l'utilisation du produit MyVoicePlus™ est envisagée ou en cours de test. Pour chacune, nous vous demanderons d'évaluer à quel point cette utilisation vous semble acceptable. Vos réponses, analysées par un laboratoire de recherche indépendant, serviront à définir une stratégie pour le développement commercial de la société, par exemple en excluant certaines applications que, en tant que participant à cette expérience, vous aurez jugées moralement inacceptables.

Références:

Khosravani, S., Mahnan, A., Yeh, I. L., Aman, J. E., Watson, P. J., Zhang, Y., ... & Konczak, J. (2019). Laryngeal vibration as a non-invasive neuromodulation therapy for spasmodic dysphonia. *Scientific reports*, 9(1), 1-11.

Wu YC, Hayashi T, Tobing PL, Kobayashi K and Toda T (2019) Quasi-periodic wavenet vocoder: a pitch dependent dilated convolution model for parametric speech generation. *arXivpreprint arXiv:1907.00797*

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11)