

Miroir de l'âme pour Aristote, la voix est le moyen d'expression le plus singulier de l'être humain. Elle est constitutive de notre identité et exprime, dans ses inflexions et ses intonations, notre personnalité et nos émotions. La parole était propre à l'homme, mais plus maintenant: au cours des deux dernières décennies, les voix de synthèse se sont multipliées et représentent aujourd'hui un enjeu majeur des technologies numériques et de l'économie associée.

La voix de synthèse, l'une des clés de voûte de l'interaction homme-machine, insuffle une « persona », une « âme » à la machine, et définit les contours de sa personnalité. Dans un futur proche, les voix de synthèse seront partout, à commencer dans notre poche, mais nous serons sans doute bien en peine de les différencier de celles de nos semblables.

Les géants du numérique Google, Apple, Microsoft et Amazon sont les fers de lance des grands groupes industriels qui investissent massivement dans la recherche et le développement des technologies numériques liées à la synthèse de la parole. Nos téléphones sont déjà capables d'obéir à nos paroles, mais aussi de nous répondre avec une voix de synthèse. L'avenir est aux assistants personnels, des intelligences artificielles dévolues à notre service et au contrôle des objets connectés de notre domicile, toujours à notre écoute et apprenant à répondre à nos moindres besoins. Tel est le cas des assistants I/O de Google, Siri d'Apple, Cortana de Microsoft et Alexa d'Amazon.

Ces questions mobilisent de nombreux laboratoires de recherche, à l'université de Cambridge en Angleterre, à l'université d'Édimbourg en

L'ESSENTIEL

- Dans la parole, deux caractéristiques essentielles sont le timbre de la voix et la prosodie (accentuation, phrasé, intonation, rythme).
- En créant ou modifiant ces caractéristiques, à l'aide d'outils numériques appropriés, on crée ou transforme à volonté une voix.
- On peut aussi reconstituer une voix perdue ou en inventer d'inédites. Les voix aujourd'hui obtenues ainsi paraissent très naturelles.
- Les applications de ces procédés se multiplient.

© Shutterstock.com/skyboyv

LES TECHNIQUES NUMÉRIQUES, notamment la numérisation des sons illustrée ici, jouent un rôle crucial dans l'analyse et la synthèse de la voix.

Écosse, à l'université de Carnegie Mellon aux États-Unis, à l'Institut des sciences et technologies de Nara au Japon... Parmi ces acteurs mondiaux, l'Ircam (Institut de recherche et coordination acoustique-musique), à Paris, se distingue par le lien qu'il entretient entre l'innovation technologique et le monde de la culture, du divertissement et de la création de contenus (musique, cinéma, jeux vidéo, etc.).

Alors que les géants du numérique et d'autres laboratoires se sont longtemps contentés de voix synthétiques intelligibles et standardisées pour des usages spécialisés, les chercheurs de l'Ircam ont depuis les années 1980 focalisé leurs efforts sur ce qui rend une voix humaine – son grain, son ton, sa personnalité, son expressivité.

Au cours de la dernière décennie, grâce notamment aux travaux de l'équipe Analyse et synthèse des sons, de l'Ircam, on est passé de voix de synthèse intelligibles, mais mécaniques, à des voix de synthèse naturelles et expressives. Cette rupture a été rendue possible par un ensemble d'avancées scientifiques réalisées en traitement du signal, en apprentissage automatique et en linguistique. Aujourd'hui, nous reproduisons plus fidèlement le grain de la voix grâce à

■ LES AUTEURS



Nicolas OBIN est maître de conférences à l'université Pierre-et-Marie-Curie (Sorbonne universités) et chercheur dans l'équipe Analyse et synthèse des sons de l'Ircam, à Paris.



Axel ROEBEL est chercheur et responsable de l'équipe Analyse et synthèse des sons de l'Ircam.

l'amélioration des algorithmes d'analyse et de synthèse de la parole. Par ailleurs, nous maîtrisons mieux la prosodie de la voix et son lien avec la syntaxe d'un texte, ce qui a considérablement amélioré le naturel des voix de synthèse. Enfin, les méthodes d'apprentissage automatique et l'augmentation de la capacité des ordinateurs ont été déterminantes pour concrétiser ces avancées et apprendre à la machine à parler (les bases de données contenant dorénavant des heures d'enregistrements pour une voix, et ce pour des milliers de voix différentes).

Aujourd'hui, des logiciels développés à l'Ircam sont utilisés dans des productions cinématographiques. Ils permettent de sculpter la voix pour en modifier son grain, son ton, son identité et sa personnalité. Nous transformons facilement une voix d'homme en celle d'une femme, une voix jeune en une voix plus âgée, et inversement. Nous sommes également capables de convertir notre voix pour imiter celle d'une autre personne. Nos ordinateurs et nos téléphones sont dotés de voix de synthèse naturelles, expressives et personnalisables, capables de prononcer n'importe quel texte.

Demain, il sera possible de rendre leur voix à des personnes ayant perdu l'usage

COMMENT LA VOIX PREND FORME

Les organes intervenant dans la parole sont nombreux. On les sépare en deux groupes. Les poumons et la trachée excitent, par leurs mouvements, les plis vocaux et les font vibrer. Cela détermine les fréquences sonores émises. Le corps de résonance – du larynx à la cavité nasale, en passant par la bouche – a pour effet de moduler l'amplitude associée à chacune des fréquences du son.

