

de la parole (*voir l'encadré page 61*), de parler dans une autre langue, mais avec sa propre voix (*voir l'encadré page 60*), de changer de voix comme on change de coiffure, d'interagir au quotidien avec des robots dotés d'une personnalité vocale indiscernable de celle d'un être humain. Ces progrès scientifiques seront sans doute au cœur d'une bataille technologique, commerciale et éthique. Mais pour y parvenir, les chercheurs devront bien comprendre la singularité de la voix humaine et la maîtriser.

Comment la voix est-elle produite ? La source d'émission est constituée de plis vocaux – nommés aussi, de façon trompeuse, cordes vocales (ils ne ressemblent pas à des cordes). Ces plis, situés dans la gorge au niveau de la glotte, sont mis en vibration par l'air que les poumons insufflent à travers la trachée. Ces vibrations entraînent des variations de la pression de l'air, qui se propagent et produisent un son à une certaine fréquence dite fondamentale, la hauteur de la voix, et ses multiples, les harmoniques (*voir l'encadré ci-dessous*).

L'irrégularité des vibrations, la tension des muscles autour des plis vocaux, la quantité d'air qui circule ainsi que les turbulences d'air engendrées modifient la « qualité » du

son émis (voix douce ou tendue, pressée, soufflée, craquée, rauque, cassée, etc.).

Le son émis au niveau de la glotte se propage à travers le conduit vocal (la trachée, puis les cavités buccales et nasales), lequel sert de corps de résonance, amplifie et colore le son. L'effet produit est celui d'un « filtre » qui module l'amplitude des composantes fréquentielles du son. La voix acquiert ainsi une coloration, un « timbre » particulier. L'anatomie propre à chaque personne rend son timbre unique, qui constitue une sorte de signature vocale.

Qui plus est, la coloration du son émis dépend de la configuration des articulateurs de la bouche (les lèvres, la langue et le voile du palais). Le contrôle de ces configurations fait de l'appareil vocal un instrument de communication. Ainsi, chaque langue est pourvue d'un ensemble de sons spécifiques nommés phonèmes, qui en constituent les unités élémentaires. Le français en compte un peu plus d'une trentaine : les voyelles et sons apparentés, comme dans « on » ou « un », les consonnes et sons apparentés (occlusives comme *b*, *d* ou *p*, fricatives comme *f*, *ch* ou *s*, etc.).

Lorsqu'une personne parle, elle est capable de contrôler et de modifier en

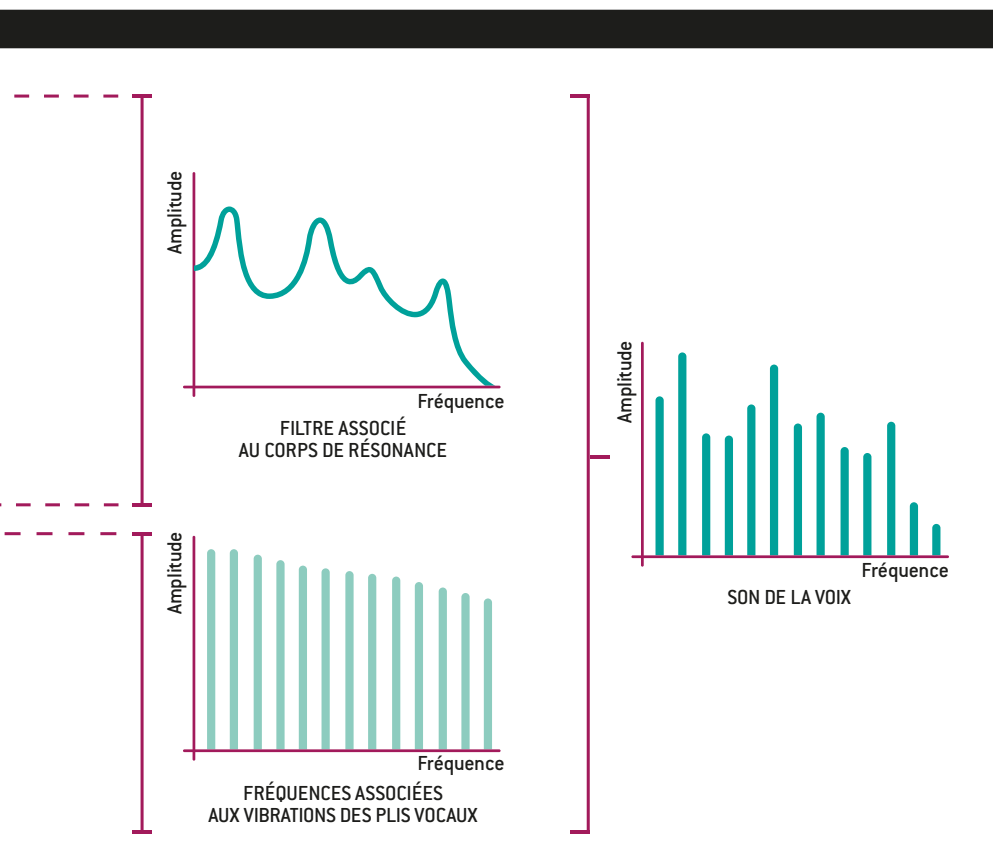
permanence la configuration de son appareil vocal. Ainsi, les articulateurs de la bouche se déplacent en permanence et articulent les phonèmes pour former des syllabes, des mots, des phrases. Également, la dynamique de la glotte et des plis vocaux permet de contrôler la prosodie, ou « musique de la voix ». Elle se traduit par des variations de hauteur, d'intensité, de rythme et de qualité vocale. Son évolution au cours du temps construit, organise et module le sens des mots et des phrases, indique nos origines géographiques et socioprofessionnelles, et traduit nos intentions et nos émotions.

Sculpter la voix

L'identité d'une voix est donc constituée de deux composantes propres à chaque individu : le timbre et la prosodie. Pour changer l'identité ou la personnalité d'une voix, il faut alors être capable de modéliser ces composantes afin de pouvoir les reproduire et les modifier. À cette fin, les chercheurs de l'Ircam ont développé en 2010 le logiciel nommé IrcamTools TRAX. À partir d'un signal sonore issu d'un enregistrement ou d'un microphone, des traitements numériques modifient la hauteur de la voix pour obtenir une voix plus grave ou plus aiguë, ou le filtre de la voix pour en modifier le timbre. Il est alors très facile de transformer une voix d'homme en voix de femme (et inversement), la voix d'un adulte en une voix d'adolescent ou de personne âgée...

En effet, les hommes et les femmes ont des caractéristiques physiologiques (taille, corpulence) qui imposent des différences au niveau de leur appareil phonatoire, et donc de leur voix. Les hommes ont des plis vocaux plus longs (de 17 à 25 millimètres) et un conduit vocal plus grand (17 centimètres en moyenne). La conséquence est que les voix d'homme sont caractérisées par une faible hauteur (de 70 à 160 hertz) et un timbre « grave ». Les femmes ont des plis vocaux plus courts (12,5 à 17,5 millimètres) et un conduit vocal plus court (14 centimètres en moyenne). Une voix de femme sera donc plus haute (de 130 à 300 hertz) et son timbre sera plus « aigu ». En outre, la voix d'une personne évolue au cours de sa vie, à l'instar de son corps. Les adolescents ont une voix aiguë qui devient plus grave à l'âge adulte, puis qui remonte au cours de la vie.

Ainsi, agir au moyen d'un logiciel sur la hauteur et le timbre d'une voix crée l'illusion



d'un changement de sexe ou d'âge. La reproduction des spécificités vocales des personnages âgés, telles que le chevrottement de la voix, nécessite d'affiner les traitements numériques réalisés. L'âge est généralement marqué par des dysfonctionnements dans la mécanique de l'appareil vocal : les plis vocaux se distendent, ce qui a pour effet d'augmenter le souffle et de faire apparaître un déraillement dans la voix, le chevrottement.

Plus surprenant, il est possible de créer des voix non humaines, voire fantastiques, par exemple faire parler une personne avec une voix de lion. Cette transformation est

réalisée à partir d'un enregistrement de voix humaine et d'un rugissement de lion, en substituant l'empreinte sonore des plis vocaux d'un lion à celle des plis vocaux de la voix humaine, tout en conservant le corps de résonance humain. Le résultat est stupéfiant : on entend une voix humaine dont on comprend clairement le sens, mais avec le souffle et les rugissements d'un lion !

Les ingénieurs de production dans le monde du cinéma et des films d'animation utilisent couramment ces traitements numériques pour modifier la voix d'un acteur ou créer des personnages imaginaires.

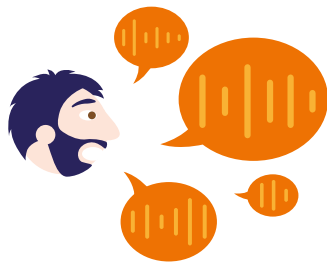
L'étape suivante, aujourd'hui devenue réalité, consiste à reproduire la voix d'acteurs ou de personnalités historiques. Ainsi, à l'Ircam, nous sommes aujourd'hui régulièrement sollicités pour ressusciter des voix à partir d'archives sonores.

Parler avec la voix d'un autre

La prosodie et le timbre font l'identité d'une voix. Mais quelle est la contribution de ces deux caractéristiques sonores lors d'une imitation ? Lorsqu'une personne A (Alice)

TRANSFORMATION DE L'IDENTITÉ DE LA VOIX

Les techniques actuelles permettent de transformer la voix d'une personne en celle d'une personne célèbre, d'une personne plus jeune, plus âgée ou même en celle d'un monstre. Dans l'exemple ci-dessous, la voix d'Alice est transformée en celle de Bernard. Les barres symbolisent la prosodie et le fond coloré des bulles, le timbre.



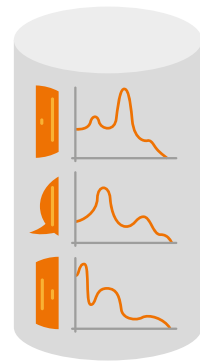
1. ENREGISTREMENT DE BERNARD

Cette étape consiste à récupérer plusieurs heures de paroles prononcées par Bernard.



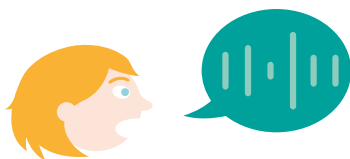
2. DÉCOUPAGE EN PHONÈMES

Un logiciel analyse les paroles de Bernard, isole les phonèmes et détermine le filtre vocal associé au timbre de chaque phonème.



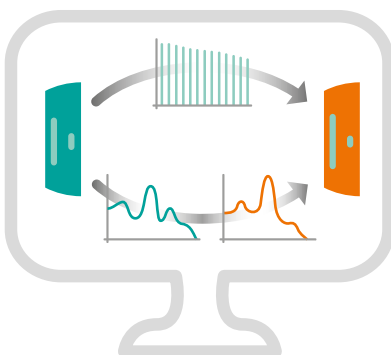
3. CONSTITUTION D'UNE BASE DE DONNÉES

Les phonèmes sont classés avec le filtre vocal associé.



4. ENREGISTREMENT D'ALICE

Alice dit le texte à transformer. Si l'objectif est d'obtenir la voix de Bernard, elle doit imiter sa prosodie. En revanche, elle ne peut imiter parfaitement son timbre.



5. ANALYSE ET TRANSFORMATION

Chaque phonème d'Alice est isolé et analysé. La prosodie est conservée tandis que le filtre vocal associé au timbre d'Alice est remplacé par celui de Bernard.



6. ASSEMBLAGE DES PHONÈMES

La voix obtenue conserve la prosodie d'Alice, mais avec le timbre de Bernard.