

**RENCONTRE** > LES MÉTIERS DE LA MUSIQUE

# Chercheur à l'Ircam

## Sur la bonne voix



Pierre Lanchantin est chercheur à l'Ircam au département Analyse-Synthèse des sons. Nous l'avons questionné plus particulièrement sur la voix, qui est son domaine privilégié d'exploration.

### **KR:** Que recouvre votre fonction à l'Ircam ?

**Pierre Lanchantin :** Je suis chargé de recherche et de développement au sein de l'équipe Analyse-Synthèse des sons. C'est une des huit équipes de recherche de l'Ircam. Son rôle est d'effectuer des recherches et des développements en analyse, transformation et synthèse des signaux sonores. Mes recherches portent plus spécifiquement sur les signaux de parole, et en particulier sur la segmentation en phones, syllabes, etc., transformation de la voix d'un locuteur vers celle d'un autre locuteur (conversion) ainsi que sur la synthèse de parole par modèles statistiques. Les techniques que nous développons le sont tout d'abord pour répondre aux demandes des musiciens pour la création de nouveaux sons et musiques, ce qui contribue au renouvellement de la création contemporaine. Nous participons aussi à des programmes de recherche nationaux et européens et nous menons de nombreuses collaborations de recherche avec l'industrie.

### Quelle a été votre formation ?

Après un bac scientifique, j'ai fait des études de physique à l'université. J'ai suivi en parallèle la formation d'ingénieur

du son de la SAE Paris. Après un master 1 de physique, j'ai intégré sur titre le cursus ingénieur de l'Institut national des télécommunications. Tout en poursuivant ces études, j'ai effectué le master 2 Atiam (Acoustique traitement du signal et informatique appliqués à la musique) de l'université Paris VI en collaboration, notamment, avec l'Ircam. Je me suis ensuite spécialisé en traitement statistique du signal en faisant un autre master 2 en Optimisation et sûreté des systèmes.

J'ai par la suite préparé une thèse en traitement statistique du signal, dont le sujet traitait notamment de modèles statistiques très utilisés dans le domaine du traitement de la parole.

### En quoi consiste(nt) le ou les procédés de synthèse vocale ?

La synthèse vocale est une technique de synthèse sonore qui permet de produire artificiellement de la parole. De manière générale, un système de synthèse de parole prend en entrée un texte et génère en sortie le signal de parole. Pour cela, des méthodes de traitement linguistique sont utilisées pour transformer le texte orthographique en une version phonétique prononçable sans ambiguïté et sur des techniques de traitement du signal pour transformer cette

version phonétique en son numérisé. Différentes méthodes ont été développées parmi lesquelles les méthodes concaténatives qui consistent à assembler des segments de parole pré-enregistrés contenus dans une base de données annotée, et les méthodes fondées sur des modélisations physiques et statistiques du signal de la parole.

### La question de « l'âne » : à quoi ça sert ?

Utilisé conjointement avec un système de reconnaissance vocale, cela permet par exemple de construire des interfaces vocales de manière à établir une communication plus naturelle avec un ordinateur.

### Quelles sont les applications industrielles, domestiques et musicales ?

Dans le domaine industriel, on peut citer les annonces automatiques dans les lieux publics, le doublage dans le domaine du cinéma, les avatars virtuels sur internet. Dans le domaine domestique, la lecture de livres audio, les jeux vidéo. Dans le domaine de la musique, la synthèse de voix chantée, la synthèse de langage imaginaire, la génération de matériaux sonores sous contrainte dans lequel un compositeur peut opérer une sélection.

### Vos recherches peuvent-elles donner naissance à une nouvelle génération d'instruments de musique ?

La voix peut par exemple être utilisée comme un contrôleur. C'est le cas du projet ANR VoxStrument, en collaboration entre l'Ircam et les sociétés Aruria et Voxler, dont l'objectif est de piloter des instruments à partir de l'expressivité naturelle de la voix captée et analysée en temps réel. D'autre part, dans le cas de la synthèse, les caractéristiques de la voix synthétisée peuvent être contrôlées précisément.

### Comment analyse-t-on une voix ?

Analyser une voix consiste à calculer des valeurs décrivant certaines de ses caractéristiques. Ces dernières dépendent de l'application visée. Typiquement, on cherche à décrire la hauteur, l'intensité et le timbre de la voix. La représentation du timbre de la voix est obtenue par un sonogramme qui permet d'afficher dans le temps la décomposition d'un son sur les fréquences qui le composent. Une fois ces caractéristiques extraites, elles peuvent être utilisées pour procéder à des transformations du signal de parole ou pour l'apprentissage de modèles qui seront utilisés pour des tâches de reconnaissance ou de synthèse de parole.

### Peut-on techniquement transformer une voix pour la rendre identique à une autre ? (ça, c'est pour les amateurs de séries policières !)

Oui. Le principe est d'apprendre la fonction de transformation des caractéristiques de la voix source vers celles de la voix dite cible à partir d'un ensemble de phrases spécifiques prononcées par les deux locuteurs. Une fois cette fonction de transformation apprise, on peut transformer les caractéristiques de n'importe quelle autre phrase prononcée par le locuteur source. On peut ainsi parler par exemple avec la voix d'un acteur célèbre. D'autres techniques permettent de modifier en temps réel le type et la nature d'une voix. On peut ainsi changer, entre autres, l'âge ou le genre d'une voix. Ces techniques ont été utilisées dans le film d'Eric Rohmer *Les amours d'Astrée et de Céladon*.

### Est-ce que la machine peut démasquer un bon imitateur ?

En général, un imitateur tend à caricaturer les caractéristiques de la voix à imiter. Il joue principalement sur les caractéristiques prosodiques (hauteurs, durées). Dans ce sens, les techniques dites de reconnaissance du locuteur existantes peuvent aisément démasquer un imitateur. Dans le cas d'un imitateur synthétique (ordinateur), c'est un peu plus difficile, mais c'est peut-être possible en détectant par exemple les artefacts inhérents aux transformations.

### Y a-t-il des applications pédagogiques relatives à ce genre de travaux (nouvelles façons d'enseigner le chant, de placer sa voix parlée, rééducation en phoniatry, assistance aux personnes handicapées, autres) ?

Tout à fait, par exemple pour l'enseignement en ligne, l'apprentissage des langues, l'assistance aux personnes handicapées : lecture pour les aveugles, pour les personnes ayant des difficultés à parler (dyslexiques).

### Vos recherches influencent-elles les modes d'enregistrement de la voix ou de diffusion en situation de concert ?

Dans le cas de traitements fondés sur des modèles statistiques, ces derniers doivent être appris à partir d'enregistrements de la voix à traiter. Il sera donc nécessaire d'enregistrer au préalable un ensemble de phrases dans les mêmes conditions que celles du concert.

Michelle Davène