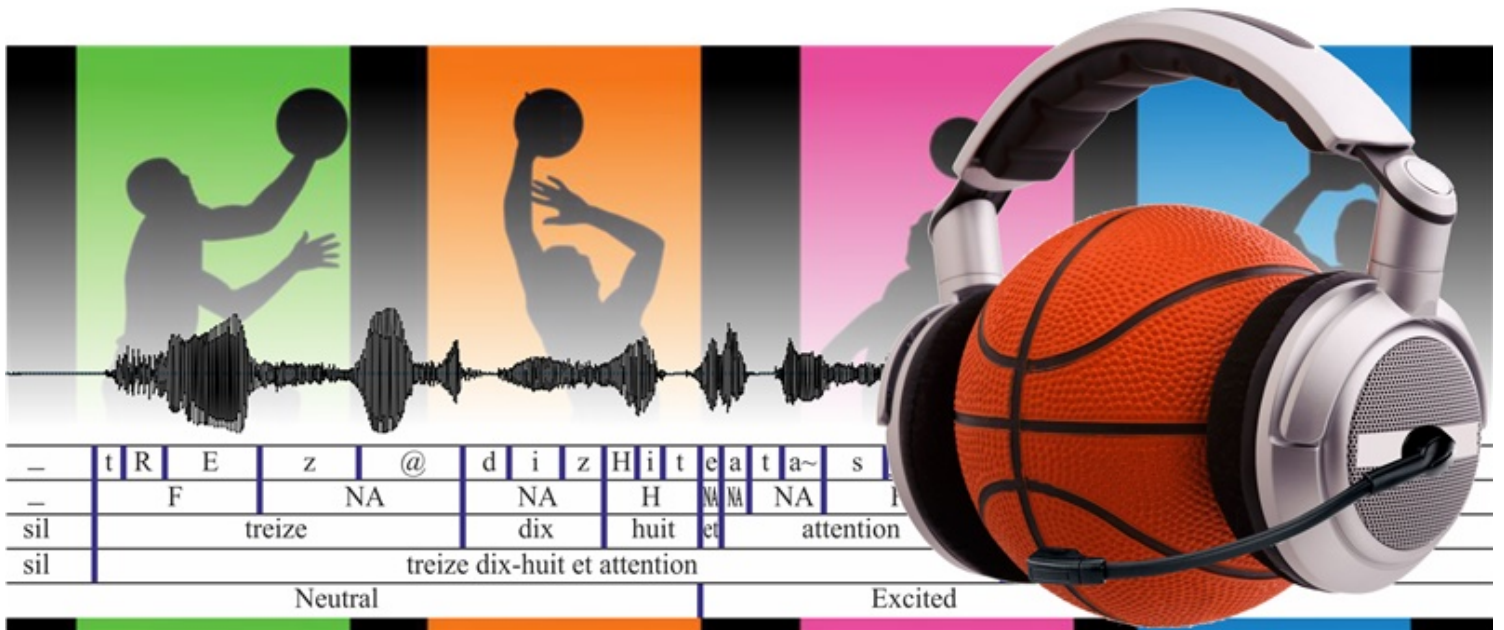


ELLE DONNE UNE VOIX PLUS HUMAINE AUX MACHINES

Publié le 30 janvier 2017



Rendre les voix de synthèse plus naturelles, c'était la mission que s'était donnée Sandrine Brognaux dans sa thèse. Pour son travail, elle a reçu, cet automne, [l'IBM Innovation Award/FRS-FNRS](#) d'un montant de 5.000 euros.

 PROX IBM 21 OCT 2016. Photo Aude Vanlathem

La synthèse vocale a déjà fait beaucoup de progrès. Elle est aujourd'hui utilisée dans de nombreuses applications comme les répondeurs ou les GPS. « Mais on se lasse vite de la monotonie de ces voix », remarque la chercheuse. La demande pour des voix plus expressives est forte notamment dans le secteur des loisirs. « Pour les jeux vidéos par exemple », précise Sandrine Brogneux. Mais pas uniquement. Ainsi, les fabricants d'audiobooks sont également très intéressés.

« Les audiobooks connaissent un gros boom. Or, il est impossible de faire enregistrer tous les ouvrages. Cela coûte cher de passer par des acteurs. Il y a donc une demande très nette de ce secteur pour la synthèse vocale », déclare Sandrine Brognaux. Ces audiobooks sont particulièrement recherchés par les personnes aveugles.

Mais la synthèse vocale est aussi utile pour les personnes qui ne peuvent plus parler suite à une laryngectomie par exemple. « Leur permettre d'utiliser une voix de synthèse plus expressive, plus proche de la voix humaine serait aller un pas

plus loin », note la chercheuse.

Si on pense d'emblée aux répondeurs et aux GPS, on le voit, la synthèse vocale est en réalité bien plus qu'un gadget. D'où l'importance de la rendre moins monotone. Plus naturelle.

Ce qui rend la parole expressive

La chercheuse a d'abord dû analyser ce qui rend la parole expressive. Et sa formation un peu atypique lui a été bien utile. Atypique car Sandrine Brognaux a réalisé un master en germanique suivi par un autre en linguistique informatique. « *Cela m'a permis d'obtenir un doctorat en sciences appliquées* », souligne-t-elle. [Et de se pencher tant sur la langue que sur les aspects plus techniques.](#)

Pour réaliser son analyse, la linguiste a travaillé sur une base de données de commentaires sportifs. « *Parce que ce sont des commentaires très spontanés ; avec beaucoup d'intonations* », confie-t-elle.

Résultat : ce qui rend la parole expressive, ce sont les liaisons réalisées entre certains mots, l'articulation, le placement d'accents... C'est cela qu'elle a tenté de modifier au niveau de la synthèse vocale.

Donner une couleur aux phrases

Elle a réalisé un synthétiseur qui permet de taper n'importe quelle phrase et de lui assigner des « labels ». Elle peut indiquer au programme de mettre l'emphase sur un mot ou un groupe de mots. Ou encore d'adopter un ton neutre, très positif (la joie) ou très négatif (la frustration). « *Cela donne une couleur générale à la phrase* », précise la chercheuse.

Effectuer ce travail sur un livre entier peut s'avérer fastidieux. « *Pour les audiobooks, on pourrait déjà varier les voix en fonction des personnages : un père un peu ronchon, un petit garçon joyeux...* », avance-t-elle. Les logiciels pourraient en effet facilement identifier les personnages grâce aux prénoms. Mais Sandrine Brognaux imagine qu'un jour l'informatique pourra analyser sémantiquement le texte et attribuer des labels appropriés aux différents passages. « *Mais ça c'est une autre recherche* », sourit-elle.

Analyse sémantique

Une recherche en plein développement. « *Il y a déjà des analyses sémantiques des commentaires sur Tripadvisor qui permettent de dire si les remarques sont positives ou négatives. Ces analyses sont aussi utilisées lors des campagnes électorales pour chiffrer le pourcentage de commentaires positifs ou négatifs sur Twitter* », raconte la chercheuse.

Les outils qu'elle a développés durant sa thèse sont utilisés par d'autres laboratoires. En Belgique ou ailleurs. Ainsi, l'UCL a mis sur pied un projet de spin-off pour l'apprentissage du français. Celle-ci recourt au programme de la chercheuse qui aligne un texte et un signal sonore.

« *Le programme repère où se trouvent les a par exemple* », explique Sandrine Brognaux. Et comme il prend en compte la parole expressive, il permet de s'adapter à l'accent de l'apprenant lorsque celui-

ci doit répéter des phrases au logiciel d'apprentissage.

Sa thèse fut réalisée en co-tutelle entre l'UCL et l'UMons. A l'UCL elle a travaillé au sein du [CENTAL](#) (Centre de traitement automatique du langage) et d'[ICTEAM](#) (Institute of information and communication technologies, electronics and applied mathematics). A l'UMons, elle était intégrée au [TCTS Lab](#) (Théorie des circuits et traitement du signal).

Depuis le mois de janvier, elle est conseillère scientifique à l'Université de Mons. Elle y promeut les projets ICT de son université. A l'UCL, elle est impliquée dans la spin-off mentionnée plus haut et donne aussi un cours de logopédie.