

# Amazon révèle une synthèse vocale “de pointe” avec BASE TTS

Une équipe de chercheurs affiliée à Amazon revendique la mise au point de BASE TTS, un modèle de synthèse vocale propulsé par l'intelligence artificielle. Décrit comme un outil « *de pointe* », capable de simuler un discours d'une manière naturelle, BASE TTS annonce de riches heures pour les livres audio... Mais un peu moins pour les comédiens et comédiennes ?

Publié le :20/02/2024 à 13:2 [Antoine Oury](#)



L'équipe d'une vingtaine de chercheurs ne dissimule pas sa fierté d'avoir mis au point un modèle de synthèse vocale doté d'un nombre impressionnant [de 980 millions de paramètres](#), et « *entraîné* » à l'aide de 100.000 d'enregistrements vocaux **issus du domaine public**. BASE TTS, ou Big Adaptive Streamable TTS with Emergent abilities (Large modèle adaptatif de synthèse vocale en diffusion continue) est présenté comme une technologie de

synthèse vocale « *de pointe* », capable de prononcer correctement des mots étrangers dans un discours en anglais, de marquer la ponctuation, mais aussi d'exprimer telle ou telle émotion. Le modèle fonctionne à partir d'un texte brut, qu'il analyse et interprète selon ses capacités, avec une relative autonomie. Son « *entraînement* » à partir d'un grand volume de discours et autres extraits de conversation a permis **d'affiner la prosodie** de BASE TTS, autrement dit sa lecture, mais aussi la manière dont l'outil prononce les mots de manière régulière, en respectant les sonorités et les rythmes.

Autant de « *détails* » qui ne passent pas inaperçus, lorsque l'on écoute un texte retranscrit par une synthèse vocale...

## Des livres lus à la pelle ?

Le développement de cet outil par des chercheurs affiliés à Amazon n'est pas anodin, puisque la firme de Seattle mise notamment sur le livre audio pour fidéliser ses clients. Sa filiale Audible a de nombreux concurrents, dont le géant suédois Spotify, et le nerf de la guerre se trouve entre autres du côté **de la taille du catalogue**.

Aussi, l'enjeu de la synthèse vocale apparaît-il comme crucial : en proposant aux auteurs indépendants, mais également aux éditeurs, un outil performant pour transposer le texte à l'oral, Amazon peut s'assurer un flux de récits et de documents audio continu.

Amazon Polly, présenté en 2019, constituait un premier pas en ce sens, avec une solution technique simple et efficace pour faire la lecture avec un certain nombre de voix synthétiques. Fin 2023, la multinationale ouvrait même [une phase de texte auprès des auteurs indépendants](#), pour qu'ils puissent aisément exploiter leurs textes au format audio.

BASE TTS pourrait sans aucun doute **améliorer l'expérience des utilisateurs** de ces livres audio lus par des voix de synthèse. Pour la plus grande inquiétude des comédiens et comédiennes, déjà secoués par [les développements technologiques dans ce domaine](#).

Toutefois, le modèle de synthèse vocale restera pour l'instant réservé à d'autres recherches menées en interne, sans mise à disposition du public. Les chercheurs continueront à étudier les améliorations apportées par la fouille massive des données, pour développer d'autres applications de synthèse vocale.

Photographie : illustration, popkantor\_tv, domaine public

Par [Antoine Oury](#)

Contact : [ao@actualitte.com](mailto:ao@actualitte.com)