

SPPAS : segmentation, phonétisation, alignement, syllabation

Brigitte Bigi

Laboratoire Parole et Langage, CNRS & Aix-Marseille Université,
5 avenue Pasteur, BP80975, 13604 Aix-en-Provence France
brigitte.bigi@lpl-aix.fr

RÉSUMÉ

SPPAS est le nouvel outil du LPL pour l'alignement texte/son. La segmentation s'opère en 4 étapes successives dans un processus entièrement automatique ou semi-automatique, à partir d'un fichier audio et d'une transcription. Le résultat comprend la segmentation en unités inter-pausales, en mots, en syllabes et en phonèmes. La version actuelle propose un ensemble de ressources qui permettent le traitement du français, de l'anglais, de l'italien et du chinois. L'ajout de nouvelles langues est facilitée par la simplicité de l'architecture de l'outil et le respect des formats de fichiers les plus usuels. L'outil bénéficie en outre d'une documentation en ligne et d'une interface graphique afin d'en faciliter l'accessibilité aux non-informaticiens. Enfin, SPPAS n'utilise et ne contient que des ressources et programmes sous licence libre GPL.

ABSTRACT

SPPAS : a tool to perform text/speech alignment

SPPAS is a new tool dedicated to phonetic alignments, from the LPL laboratory. SPPAS produces automatically or semi-automatically annotations which include utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription. SPPAS is currently implemented for French, English, Italian and Chinese. There is a very simple procedure to add other languages in SPPAS : it is just needed to add related resources in the appropriate directories. SPPAS can be used by a large community of users : accessibility and portability are important aspects in its development. The tools and resources will all be distributed with a GPL license.

MOTS-CLÉS : segmentation, phonétisation, alignement, syllabation.

KEYWORDS: segmentation, phonetization, alignement, syllabification.

Dans le vaste domaine de la segmentation texte/son, on trouve sur le web de nombreuses « boîtes à outils » pour réaliser différents niveaux de segmentations de la parole et l'apprentissage des modèles sous-jacents. Des ressources (dictionnaires, modèles, etc.) sont également disponibles pour pouvoir les exploiter. Pourtant, lorsqu'il s'agit d'effectuer des alignements (en quantité raisonnable), la plupart des phonéticiens choisissent de le faire manuellement même si plusieurs heures sont souvent nécessaires pour n'aligner qu'une seule minute de signal. La raison principalement évoquée concerne le fait que très peu d'outils sont à la fois disponibles librement, utilisables *de façon simple et ergonomique*, multi-plateforme et, bien sûr, qui traite la langue souhaitée. Ainsi, bien qu'elles soient très utilisées par les informaticiens, des boîtes à outils telles que, par exemple, HTK (Young, 1994), Sphinx (Carnegie Mellon University, 2011)

ou Julius (Lee *et al.*, 2001), ne bénéficient toujours pas d'un développement qui permette une accessibilité à une communauté plus large d'utilisateurs. Développer un outil d'alignement automatique, s'appuyant uniquement sur des ressources libres (outils et données) et regroupant les critères nécessaire à son accessibilité à des non-informaticiens n'est pas uniquement un défi technique. On suppose en effet que si tel était le cas, cet outil existerait déjà.

Le logiciel présenté lors de la session de démonstration s'appelle SPPAS :

<http://www.lpl-aix.fr/~bigi/sppas/>

La segmentation en IPU consiste à aligner les macro-unités d'un texte (segments, phrases, etc) avec le son qui lui correspond. L'algorithme implémenté dans SPPAS s'appuie sur la recherche des pauses dans le signal et leur alignement avec les unités proposées dans la transcription (en supposant qu'une pause sépare chaque unité).

La phonétisation consiste à représenter les unités (mots, syllabes) d'un texte par des symboles phonétiques. SPPAS s'appuie uniquement sur la consultation d'un dictionnaire de prononciations et traite les deux cas suivants : 1/ une entrée peut se prononcer de différentes manières (homographes hétérophones, accents régionaux, phénomènes de réductions propres à l'oral...). Dans ce cas, c'est l'aligneur qui choisit la phonétisation. 2/ une entrée peut être absente du dictionnaire. SPPAS produit une phonétisation automatique dont l'algorithme, indépendant de la langue, cherche les segments les plus longs dans le dictionnaire. Dans un processus semi-automatique, l'utilisateur peut choisir/modifier la phonétisation.

L'alignement en phonème consiste à déterminer la localisation temporelle de chacun des phonèmes d'une unité. SPPAS fait appel à Julius pour réaliser l'alignement. Pour réaliser l'alignement, Julius a besoin d'une grammaire et d'un modèle acoustique. La grammaire contient la (ou les) prononciation(s) de chaque mot et l'indication des transitions entre les mots. L'alignement requiert aussi un modèle acoustique qui doit être au format HTK-ASCII.

Concernant la syllabation, SPPAS encapsule le syllabeur du LPL (Bigi *et al.*, 2010). Il consiste à définir un ensemble de règles de segmentation entre phonèmes. Les phonèmes sont regroupés en classes et des règles de segmentation entre ces classes sont établies.

Références

- BIGI, B., MEUNIER, C., NESTERENKO, I. et BERTRAND, R. (2010). Automatic detection of syllable boundaries in spontaneous speech. *In Language Resource and Evaluation Conference*, pages 3285–3292, La Valetta, Malta.
- CARNEGIE MELLON UNIVERSITY (2011). CMUSphinx : Open Source Toolkit For Speech Recognition. <http://cmusphinx.sourceforge.net>.
- LEE, A., KAWAHARA, T. et SHIKANO, K. (2001). Julius — an open source real-time large vocabulary recognition engine. *In European Conference on Speech Communication and Technology*, pages 1691–1694.
- YOUNG, S. (1994). The HTK Hidden Markov Model Toolkit : Design and Philosophy. *Entropic Cambridge Research Laboratory, Ltd*, 2:2–44.