TALKAPILLAR : L'INFORMATIQUE AU SERVICE DES LINGUISTES

Résumé : Cette présentation vise à faire une démonstration d'un outil élaboré depuis cinq ans dans le laboratoire analyse synthèse de l'IRCAM (Institut de Recherche et de Coordination Acoustique Musique). TALKAPILLAR, offre aujourd'hui de multiples possibilités comme la synthèse de la parole, la transformation dépendante du contexte ou l'analyse statistique de grands corpus oraux, accessibles par simple exploration graphique.

Mots-clés : Base de données, descripteurs acoustiques, expressivités, interface, statistiques.

Disciplines : Phonétique appliquée, linguistique, traitement du signal, informatique statistiques.

1. Introduction

Dans cet article est présenté un outil élaboré depuis cinq ans dans le laboratoire analyse synthèse de l'IRCAM (Institut de Recherche et de Coordination Acoustique Musique). Originellement conçu par Diemo Schwarz lors de sa thèse (Schwarz 2004), le système CATERPILLAR a pour but de synthétiser des phrases musicales de haute qualité par la concaténation de petites unités provenant d'une grande base de données.

Lors de travaux plus récents, ses capacités ont été élargies dans le but de synthétiser de la parole de haute qualité (Beller 2004), pour des applications artistiques telles que le théâtre, la musique contemporaine ou le cinéma (Beller et al. 2005). Renommé TALKAPILLAR, ce système offre aujourd'hui de multiples possibilités comme la synthèse de la parole d'un locuteur disparu, la transformation dépendante du contexte ou l'analyse statistique de grands corpus oraux.

C'est cette dernière exploitation qui est mise en avant dans cette présentation avec en guise d'exemple : L'étude d'un corpus de parole expressive prononcé par un acteur français d'une quarantaine d'année et enregistré dans une chambre anéchoïque (Beller 2005). Plusieurs expressivités ont été simulées par l'acteur: la peur, la joie, la colère, la tristesse, le dégoût, l'indignation, la surprise. Chaque expressivité a été exprimée selon trois niveaux différents d'intensité (faible, moyen et fort). Les résultats présentés sont tous accessibles d'une manière très simple grâce à une interface graphique.

2. Présentation technique

TALKAPILLAR est un ensemble de scripts et fonctions Matlab® qui relient une interface graphique à une base de données relationnelle PostgreSQL. Cette base de données contient de nombreuses « unités » de parole segmentée en semi phones, phones, diphones, syllabes, groupes prosodiques et phrases. Chaque unité est décrite par de nombreux descripteurs catégorisés en trois types :

2.1. Les descripteurs symboliques :

Ils proviennent en majeur partie du projet européen EULER: Phonème, mot, fonction lexicale, contexte (places relatives des unités les unes par rapport aux autres), frontières prosodiques, accentuation des syllabes…

2.2. Les descripteurs dynamiques :

Ils sont les résultats d'analyses acoustiques et évoluent durant les unités : Fréquence fondamentale, énergie, voisement, formants…

2.3. Les descripteurs statiques :

Ils constituent un modèle de l'évolution temporelle des descripteurs dynamiques durant les unités : Moyenne, min/max, variance, coefficients des régressions linéaire et cubique...

Tous ces descripteurs sont accessibles par scripts Matlab® ou via une interface graphique permettant une navigation simplifiée et instantanée dans la base de données. Pour le moment, deux corpus ont été réalisés : L'un contenant approximativement 4 heures de parole neutre lue, et l'autre sus cité, d'une durée d'environ 1 heure de parole expressive.

3. Exemple d'analyse d'un corpus

La reconnaissance des émotions dans la parole signifie la mise en évidence de paramètres acoustiques pertinents, caractéristiques de la façon dont elles s'expriment. La distinction entre la joie et la colère est toujours une difficulté aujourd'hui (Chung 2000) car ces deux expressivités possèdent entre autre, une moyenne de la fréquence fondamentale élevée. Or, cet indice acoustique s'avère primordial d'après de nombreux algorithmes d'apprentissage (Pereira 2000). La figure 1 illustre cette proximité et le fait que de nombreux algorithmes accordant un poids trop grand à la moyenne de la fréquence fondamentale échouent dans la distinction entre la joie et la colère. La figure 1 apporte un autre élément de distinction dans l'observation de la variance du débit syllabique : Elle est beaucoup plus forte dans le cas de la joie ou certains phones sont presque chantés, tandis que la colère s'exprime par une isochronie syllabique forte.

Figure 1 - ***Moyennes et variances de la fréquence fondamentale (abscisse) et de la durée des syllabes (ordonnée) pour chaque expressivité***

4. Conclusion

De nombreuses analyses de ce type (effet de l'expressivité sur les pauses, les respirations, le triangle vocalique, la qualité vocale …) ont été effectuées sur le corpus de l'acteur de manière très simple puisqu'il s'agit de simples requêtes à la base de données. Cet outil se présente donc comme un bon moyen de réaliser des études statistiques sur des corpus oraux et peut amener à des conclusions générales si toutefois le nombre d'individus le permet.

TALKAPILLAR est aussi un bon moyen d'administrer des bases de données puisque plusieurs utilisateurs peuvent s'en servir en même temps (aspect multi client). Enfin, l'interface graphique conçue permet une navigation aisée dans des corpus pouvant dépasser des heures de paroles. Deux modules permettent la synthèse de parole par sélection et concaténation d'unités (analyse par la synthèse) et la transformation dépendante du contexte.

La présentation orale a pour but de montrer les capacités d'un tel système de façon ludique et interactive. Une vue générale du système sera pourvue avant de donner des exemples d'exploitations : Exploration graphique…

Références

(Schwarz 2004) Schwarz, Diemo, 2004. Data-Driven Concatenative Sound Synthesis, *Thèse de doctorat de l'Université Paris 6 - Pierre et Marie Curie*

(Beller et al. 2005) Beller, Grégory, Schwarz, Diemo, Hueber, Thomas & Rodet, Xavier, 2005. Hybrid Concatenative Synthesis In The Intersection of Speech and Music in *JIM 2005*

(Beller 2005) Beller, Grégory, 2005. La musicalité de la voix parlée, *Mémoire de maîtrise Paris 8*

(Beller 2004) Beller, Grégory, 2004. Un synthétiseur vocal par sélection d'unités, *Rapport interne de l'IRCAM*

(Chung 2000) Soo-Jung. Chung, 2000. L'expression et la perception de l'émotion extraite de la parole spontanée: évidences du coréen et de l'anglais, *Thèse de l'Université Paris 3 – ILPGA*

(Pereira 2000) Pereira, Claire, 2000. Perception and expression of emotions in speech, *Thèse de l'Université Macquarie*