

Emotions, speech and the ASR framework

Louis ten Bosch, Nijmegen

SpeCom 03

Présentation : Clément de Groc, Orsay

Outline

- Contexte
- Système
- Émotions
- Conclusion

Domaine

Traitement automatique du langage (TAL)

- Linguistique + Informatique + Intelligence Artificielle
- Techniques et programmes informatiques appliqués au langage humain

Traitement du langage parlé (TLP)

- Traitement du signal + TAL
- Reconnaissance vocale, Synthèse vocale, Reconnaissance du locuteur...

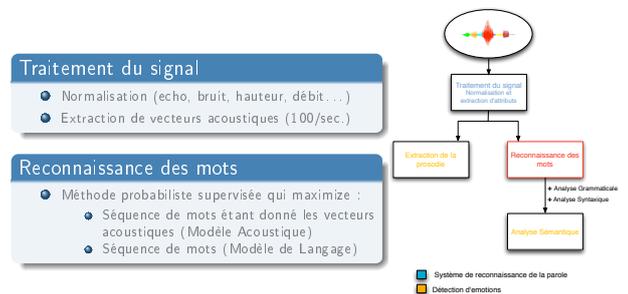
Les émotions dans le langage

- Domaine de recherche assez récent
- Choix du nombre d'émotions à détecter (absence/présence, 3, 6...)
- Compréhension, réalisme, agents réactifs à l'humeur / aux émotions...

Étude de la parole ⇒ Mot - Sémantique - Prosodie - Émotion - Fonction [Manning00]

Système de reconnaissance de la parole (ASR)

Mais qu'est-ce qu'y dit?



Désambiguïsation Sémantique

De quoi ky parle?

Exemple

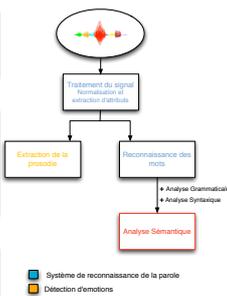
- "La pièce de cet auteur est très mauvaise"
- "La pièce de 10 euros est très petite"

Méthode

- Analyse grammaticale (Nom, Verbe, Adjectif, ...)
- Analyse syntaxique (Groupe Nominal, ...)
- Méthode à base de règles ou statistique (WordNet)

Intérêt

- "Compréhension" de ce qui est dit
- Utilisation pour la reconnaissance d'émotions



Prosodie

Définition

- Phénomènes dits "supra-segmentaux"
- Hauteur, accent, durée, débit, intonation.

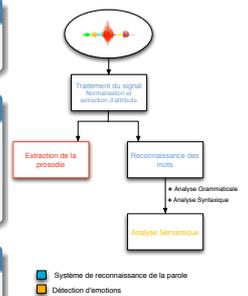
Intégration au système

Deux possibilités :

- Comme vecteurs acoustiques séparés puis utilisés lors de la reconnaissance des mots
- Intégrés dans les vecteurs acoustiques comme tous les autres attributs

Intérêt

- Améliore la reconnaissance des mots
- Utilisation pour la reconnaissance d'émotions



Émotions

Méthode d'analyse standard

- Constitution d'une base de données de phrases lues avec différentes émotions
- Étude des corrélations entre les attributs acoustiques et les émotions
- Vérification et ajustement en utilisant un système de TTS et un interlocuteur humain (subjectif)

Corrélations Prosodie-Émotions

- Attributs principaux : Hauteur, Énergie, Durée syllabique, Débit
- Exemples [Montero98] :
 - Tristesse ⇒ Débit faible
 - Peur ⇒ Débit rapide

Sémantique et émotions

- Actuellement : Input ⇒ Action
- La sémantique contient des informations sur les émotions (surprise, colère...)

Travaux

1/3

Slaney et Mc Roberts. 1998

- Paramètres : hauteur uniquement
- Cas des relations enfants/parents
- 3 états : Approbation, attention et prohibition
- 55% de réussite indépendamment de l'orateur et entre 60 et 92% sinon.
- ⇒ Très dépendant du locuteur

Whiteside 1998 / Nogueiras et al. 2001

- Paramètres : hauteur, énergie, écart-type
- Art1 : Colère, peur, surprise, dégoût, joie, tristesse, neutre
- Art2 : Colère, intérêt, joie, exaltation, tristesse, neutre
- ⇒ Tristesse et colère sont facilement reconnaissables
- ⇒ Surprise et joie sont très souvent confondus

Li et Zhao. 98

- Paramètres : Utilisation d'attributs court et long terme. Analyse en Composantes Principales.
- 6 émotions : Colère, peur, surprise, joie, tristesse, neutre
- Trois groupes d'émotions difficilement séparables :
 - Neutre-Triste
 - Colère-Peur
 - Joie-Surprise
- 62% de réussite (5 locuteurs, 15 phrases d'entraînement, 5 phrases de test)
- ⇒ **Prise en charge du contenu et de la sémantique**

Koike et al. 1998

- 5 émotions : colère, surprise, tristesse, haine et joie
- Paramètres prosodiques pour le Français et le Japonais
- ⇒ **Colère, tristesse et haine peuvent être correctement identifiés par ces paramètres à plus de 85%. Les autres nécessitent un apport sémantique.**

Shigeno. 1998

- 7 Japonais, 5 Nord-Américains.
- 6 émotions : Colère, peur, surprise, dégoût, joie, tristesse
- Enregistrement vocal + expression faciale
- ⇒ **En audio : reconnaissance d'émotions inter-culturelles plus difficile**
- ⇒ **Combiné : reconnaissance d'émotions repose plus sur le visuel que sur l'audio**

Klein, 1999

- Affective computing = Agents réactifs aux émotions
- ⇒ **Agent anti-frustration plus efficace qu'un agent neutre pour aider les personnes à récupérer d'une situation de frustration.**

Résultats

- Reconnaissance des émotions est difficile
- Dépendante du locuteur et de la langue/culture
- Prise en compte du contenu (sémantique)
- Prosodie est porteuse de beaucoup d'informations (Hauteur + Énergie)

Données et validation

- Jeux de données souvent très petits
- Pas de données en situation réelle
- Validation souvent subjective