

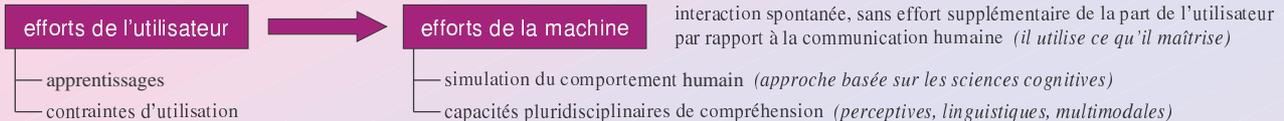
Compréhension automatique du geste et de la parole spontanés : apport de la Théorie de la Pertinence

LORIA – UMR 7503, campus scientifique BP 239, 54506 Vandœuvre-lès-Nancy



Abstract: With an approach based on the spontaneous character of man-machine communication, we need to provide multimodal systems with intelligent comprehensive abilities. The design of such systems must rely on theories from social sciences. But these theories have to handle concepts which can be formalized in a computing point of view. The exemple we present with Relevance Theory appears to fit well this concern.

Communication homme-machine spontanée : pour que les systèmes de dialogue ne soient pas uniquement utilisables par des spécialistes



CAPACITÉS PERCEPTIVES

Capture et exploitation du contexte visuel

- connaissance ou reconnaissance des objets constituant la scène visuelle
- structuration de ces objets en groupements perceptifs selon les lois de la gestalt
- identification de la dynamique de la scène : lignes de force, saillance
- éventuellement capture de la direction du regard

Capture et traitement de la parole

- un ou plusieurs micros, éventuellement couplage avec une lecture sur les lèvres
- modélisation du vocabulaire, performances maximisées sur les mots attendus
- modélisation de la grammaire, acceptation des constructions du langage parlé

Capture et traitement du geste

- dispositifs non intrusifs
- acceptation de tous les types de gestes

CAPACITÉS LINGUISTIQUES ET MULTIMODALES

Connaissance des possibilités d'appariements des modalités

- indices temporels (*synchronisation*)
- indices lexicaux (*détermination, quantification*)
- indices syntaxiques (*coordination, genre et nombre*)
- indices sémantiques (*appel à des connaissances encyclopédiques*)
- indices pragmatiques (*plausibilité par rapport au contexte*)

Interprétation en contexte visuel

- capacités d'interprétation du geste : compréhension d'un geste imprécis, extension d'un geste à un groupe perceptif
- capacités d'interprétation de la parole : exploitation de la saillance visuelle, confrontation avec l'historique du dialogue
- capacités d'interprétation multimodale : couplage des filtres spatial et catégoriel

Interaction multimodale sur écran tactile : quand le dispositif limite les efforts de la machine tout en permettant une interaction spontanée

exploitation de tous les dispositifs

faisabilité : choix du micro et de l'écran tactile

- conséquences :
- la parole, support de la communication, reste spontanée
 - l'utilisation des gestes, guidée par le dispositif, est restreinte seule la partie significative est captée, pas les phases d'approche et de retrait

exploitation de toutes les expressions multimodales

faisabilité : performances maximisées pour les plus pertinentes

- conséquences :
- si un même geste peut avoir plusieurs interprétations (synchronisateur, expressif, référentiel), la plus pertinente est retenue en priorité
 - la compréhension dépend du modèle de pertinence : apport d'information utile à l'interlocuteur ; optimisation du ratio : effets contextuels / effort de traitement

GESTES COMMUNICATIFS EXPLOITÉS

| | | |
|---------------------|-------------------------------------|----------------------------------------------------------------------|
| Quasi-linguistiques | <input type="checkbox"/> | (<i>car écran tactile</i>) |
| Co-verbaux | | |
| référentiels | <input checked="" type="checkbox"/> | |
| expressifs | <input type="checkbox"/> | (<i>car communication homme-machine</i>) |
| paraverbaux | <input type="checkbox"/> | (<i>pertinence : apport d'information utile à l'interlocuteur</i>) |
| Synchronisateurs | <input type="checkbox"/> | (<i>car communication homme-machine</i>) |

Met l'accent sur le principal problème en communication homme-machine avec un support visuel : celui de la **référence** (comment faire le lien entre les objets de la scène et les expressions langagières ou multimodales utilisées)

FORMALISATION DE LA PERTINENCE

effets contextuels = conséquences de l'interprétation d'un énoncé sur le contexte
effort de traitement = fonction de la complexité de l'énoncé, de la complexité du contexte et de l'accessibilité des informations dans celui-ci

Ce qui donne, au niveau d'une référence :

effets contextuels = fait nouveau qu'il y a eu une référence, et résultat de celle-ci
effort de traitement = complexité de l'expression verbale (*nb de traits sémantiques*)
complexité de la trajectoire gestuelle (*nb de singularités*)
effort nécessaire pour associer geste et parole
taux de présence des informations requises dans la scène visuelle et dans l'historique du dialogue