

mini-chapitre pour le "Livre blanc de l'Acoustique", Juin 2008, Alain de Cheveigné

Perception de la parole.

La parole est une activité proprement humaine, sans doute la plus importante. Sa perception fait intervenir, outre les mécanismes auditifs génériques semblables à ceux qu'on trouve chez d'autres espèces, des mécanismes spécifiques qui nous sont propres. Comprendre ces mécanismes est important sur le plan scientifique, et aussi pour les applications médicales (prothèses, implants, etc.) et industrielles (téléphonie, interfaces homme-machine, etc.). La perception de la parole est aussi essentielle pour calibrer les modèles de *production*, en particulier pour déterminer quelles sont les dimensions pertinentes pour l'évaluation de la qualité d'un modèle.

La perception de la parole est abordée par plusieurs disciplines. La *psychoacoustique* caractérise les facultés auditives de base, en relation avec la *physiologie* (chez les modèles animaux), *l'imagerie cérébrale* (chez l'homme) et la *modélisation* de l'audition. La *phonologie* et la *phonétique* étudient respectivement la structure abstraite et la forme acoustique détaillée des sons du langage. La *psycholinguistique* s'intéresse aux mécanismes perceptifs et cognitifs de haut niveau, et à leur développement. La *reconnaissance de la parole*, et plus généralement les techniques d'*apprentissage automatique* (machine learning) fournissent des modèles utiles dont on peut s'inspirer pour mieux comprendre les mécanismes de perception de la parole. En retour les mécanismes naturels peuvent inspirer de nouveaux algorithmes applicables dans les systèmes artificiels.

Les implants cochléaires et appareillages pour malentendants visent en premier à rétablir les facultés de perception et communication verbales. Il est essentiel de préserver la part de l'information acoustique nécessaire à la perception dans la conception des systèmes et algorithmes de traitement, sachant que les contraintes techniques (miniaturisation, nombre de canaux réduit, etc.) sont sévères. Un problème important est l'intelligibilité dans le bruit, en particulier la présence de voix concurrentes (effet cocktail ou "cocktail party effect"). Les porteurs d'implants, malentendants, et même les sujets normaux âgés, souffrent particulièrement dans ces situations.

L'acoustique des salles de conférence et de cours (écoles) vise à concevoir des espaces aptes à préserver l'intelligibilité. Il est essentiel de connaître les traits acoustiques nécessaires à la perception de la parole afin de les préserver. Les métriques globales (par exemple basées sur la puissance) risquent d'attribuer trop de poids à des dimensions moins utiles, et ainsi perdre de leur pouvoir prédictif. Il en est de même des métriques qui mesurent la qualité de modèles de production, ou de systèmes de synthèse de la parole: l'erreur RMS n'est pas le meilleur prédicteur de la qualité. Les tests perceptifs, et modèles de perception, font partie de la boucle de synthèse.

La téléphonie est la principale application qui s'intéresse à la perception de la parole, et les premiers progrès dans ce domaine datent des travaux dans les laboratoires Bell dans les années 20 à 50 du siècle dernier. Les progrès techniques ont amené à s'intéresser à la mesure de la "qualité" (définie comme ce qui affecte la satisfaction de l'utilisateur à intelligibilité égale). Avec le développement des réseaux cellulaires l'intelligibilité redevient à l'ordre du jour. La quantification des effets de variations de taux de compression ou pertes de paquets est complexe.

La reconnaissance de la parole substitue à l'oreille et au cerveau de l'auditeur un microphone et un ordinateur. Le développement de ces techniques artificielles s'est inspiré de la perception et la cognition de la parole chez l'homme, même si la trace de cet apport est difficile à reconnaître dans les techniques actuelles (réseaux de markov, modèles de langage, etc.). Les capacités humaines d'écoute dans le bruit (parole concurrente notamment) continuent inspirer le développement de techniques de séparation de source et reconnaissance robuste (par exemple les modèles CASA, Computational Auditory Scene Analysis).

La parole est extrêmement robuste face aux dégradations du signal, telles que filtrage, distortion non-linéaire, interruption, ce qui complique la détermination des traits acoustiques importants pour la perception. Face à un locuteur, l'auditeur utilise aussi des indices visuels (mouvements des lèvres et du visage) et non seulement acoustiques. La compréhension du message acoustique prend appui sur le contexte et les connaissances partagées, qui elles-mêmes progressent au cours du dialogue. Ce sont des phénomènes difficiles à caractériser et quantifier, et reproduire dans des systèmes artificiels.

- Psychoacoustique, Psycholinguistique, Physiologie, Imagerie Cérébrale.
- Importance de la perception de la parole pour les telecomms
- Importance pour la synthèse, et l'acoustique de la production.