

Abstract - Résumé

1 Composition du jury

M. Daniel Pressnitzer	Directeur de Recherche, ENS Paris	Rapporteur
M. Steven van de Par	Pr., Université d'Oldenburg	Rapporteur
M. Andrew J. Oxenham	Pr., Université du Minnesota	Examineur
M. Jean-Christophe Valière	Pr., Université de Poitiers	Examineur
Mme. Catherine Marquis-Favre	Chargée de Recherche HdR, ENTPE Lyon	Examinatrice
M. Nicolas Grimault	Chargé de Recherche HdR, CRNL Lyon	Directeur de thèse
M. Mathieu Lavandier	Chargé de Recherche, ENTPE Lyon	Directeur de thèse

2 Abstract

In a context of competing sound sources, the auditory scene analysis aims to draw an accurate and useful representation of the perceived sounds. Solving such a scene consists of grouping sound events which come from the same source and segregating them from the other sounds. This PhD work intended to further develop our understanding of how the human auditory system processes these complex acoustic environments, with a particular emphasis on the potential influence of spatial cues on perceptual stream segregation. All the studies conducted during this PhD endeavoured to rely on realistic configurations.

In a real environment, the diffraction and reflection properties of the room and the head lead to distortions of the sounds depending on the source and receiver positions. This phenomenon is named colouration. Speech-shaped noises, as a first approximation of speech sounds, were used to evaluate the effect of this colouration on stream segregation. The results showed that the slight monaural spectral differences induced by head and room colouration can induce segregation. Moreover, this segregation was enhanced by adding the binaural cues associated with a given position (ITD, ILD). Especially, a second study suggested that the monaural intensity variations across time at each ear were more relevant for stream segregation than the interaural level differences. The results also indicated that the percept of lateralization associated with a given ITD helped the segregation when the lateralization was salient enough. Besides, the ITD *per se* could also favour segregation.

The natural ability to perceptually solve an auditory scene is relevant for speech intelligibility. The main idea was to replicate the first experiments with speech items instead of frozen noises. A characteristic of running speech is a high degree of acoustical variability used to convey information. Thus, as a first step, we investigated the robustness of stream segregation based on a frequency difference to variability on the same acoustical cue (i.e., frequency). The second step was to evaluate the fundamental frequency difference that enables to separate speech items. Indeed, according to the limited effects measured in the two first experiments, it was assumed that spatial cues might be relevant for stream segregation only in interaction with another “stronger” cue such as a F0 difference.

The results of these preliminary experiments showed first that the introduction of a large spectral variability introduced within pure tone streams can lead to a complicated percept,

presumably consisting of multiple streams. Second, the results indicated that a fundamental frequency difference comprised between 3 and 5 semitones enables to separate speech item. These experiments provided results that will be used to design the next experiment investigating how an ambiguous percept could be biased toward segregation by introducing spatial cues.

Keywords: Auditory scene analysis, sequential segregation, spectral differences, spatial cues, speech sounds.

3 Résumé

Dans un contexte sonore constitué de plusieurs sources sonores, l'analyse de scène auditive a pour objectif de dresser une représentation précise et utile des sons perçus. Résoudre ce type de scènes consiste à regrouper les sons provenant d'une même source et à les séparer des autres sons. Ce travail de thèse a eu pour but d'approfondir nos connaissances du traitement de ces scènes auditives complexes par le système auditif. En particulier, il s'agissait d'étudier l'influence potentielle des indices spatiaux sur la ségrégation. Une attention particulière a été portée tout au long de cette thèse pour intégrer des éléments réalistes dans toutes les études menées.

Dans un environnement réel, la salle et la tête entraînent des distorsions des signaux de parole en fonction des positions de la source et du récepteur. Ce phénomène est appelé coloration. Comme première approximation de la parole, des bruits avec un spectre de parole ont été utilisés pour évaluer l'effet de la coloration. Les résultats ont montré que les fines différences spectrales monaurales induites par la coloration due à la tête et à la salle peuvent engendrer de la ségrégation. De plus, cette ségrégation peut être renforcée en ajoutant les indices binauraux associés à une position donnée (ILD, ITD). En particulier, une deuxième étude a suggéré que les variations monaurales d'intensité au cours du temps à chaque oreille étaient plus utiles pour la ségrégation que les différences interaurales de niveau. Les résultats ont également montré que le percept de latéralisation, associé à un ITD donné, favorise la ségrégation lorsque ce percept est suffisamment saillant. Par ailleurs, l'ITD *per se* peut induire de la ségrégation.

La capacité naturelle à résoudre perceptivement une scène auditive est pertinente pour l'intelligibilité de la parole. L'objectif était de répliquer ces premières expériences avec des signaux de parole à la place de bruits gelés. Une caractéristique de la parole est la grande variabilité de ses paramètres acoustiques qui permet de transmettre de l'information. Ainsi, la première étape a été d'étudier la robustesse de la ségrégation basée sur une différence de fréquence à une variation sur la même dimension acoustique (i.e., la fréquence). La seconde étape a été d'évaluer la différence de fréquence fondamentale requise pour séparer des flux de parole. En effet, en accord avec les faibles effets mesurés dans les deux précédentes études, il a été supposé que des indices de position pourraient être utiles pour la ségrégation seulement en interaction avec un indice plus robuste comme une différence de F0.

Les résultats de ces expériences préliminaires ont montré que l'introduction d'une large variabilité spectrale au sein de flux de sons purs pouvait entraîner un percept compliqué, probablement constitué de multiples flux sonores. De plus, les résultats ont indiqué qu'une différence de F0 comprise entre 3 et 5 demi-tons permettait de séparer des signaux de parole. Les résultats de ces expériences pourront être utilisés pour concevoir la prochaine expérience visant à étudier dans quelle mesure un percept ambigu peut évoluer vers de la ségrégation par l'introduction d'indices de position.

Mots-clé: Analyse de scènes auditives, ségrégation séquentielle, différences spectrales, indices de position, signaux de parole.