STRATEGIES FOR VOICE SEPARATION BASED ON HARMONICITY

Alain de Cheveigné

Laboratoire de Linguistique Formelle, CNRS/Université Paris 7,
case 7003, 2 place Jussieu, 75251, France.

ABSTRACT

A harmonic sound has a spectrum made of discrete components forming a harmonic series, and its wave form is periodic. Voiced speech is approximately harmonic, and this cue appears to be exploited by the auditory system for the perception of speech in a noisy background. This paper addresses the following question: does the auditory system exploit the harmonic structure of a target to segregate it from the background, or else that of a harmonic background to eliminate it.

The issue was considered from two angles. First of all, both strategies were implemented to reduce voiced interference in a speech recognition experiment. The aim was to determine whether one "works better" than the other in this task; the results showed that harmonic cancellation was more effective. In a second experiment human listeners were presented with pairs of vowels, each of which was either harmonic or inharmonic, and requested to identify them both. The responses were scored according to whether the target vowel was harmonic, and whether the other (interfering) vowel was harmonic. It appears that the auditory system exploits harmonicity of the ground (cancellation strategy), a result that is coherent with the outcome of the first experiment.

1. HARMONIC ENHANCEMENT VS CANCELLATION

The regular structure of the spectrum of a harmonic target can be exploited to select its components among those of interfering sounds. We refer to this strategy as *harmonic enhancement*. In a similar fashion the components of a harmonic background may be removed. We refer to this as *harmonic cancellation*. If both target and ground are voiced, both strategies are a priori possible, but their properties and implementation are quite different. It is of theoretical and practical interest to know if one strategy is more effective than the other, and which one (possibly both) is used by the auditory system.

Both strategies have their advantages and drawbacks of principle:

• *Enhancement requires the target to be harmonic but works with any kind of interference.*	• *Cancellation requires the interference to be harmonic but works with any kind of target.*
• *Enhancement requires the fundamental frequency (F0) of the target to be determined. Relatively easy if the Signal-to-Noise Ratio (SNR) is large.*	• *Cancellation requires the F0 of the interference to be estimated. Relatively easy if the SNR is small.*