Are fine structure cues an important feature for temporal streaming ?

Nicolas Grimault [1]

[1] *UMR CNRS 5020, Université Claude Bernard Lyon 1, IFR 19, Institut fédératif des neurosciences de Lyon, Lyon, France, Email: nicolas.grimault@olfac.univ-lyon1.fr*

Introduction

Alternating sequence of two ***A*** and ***B*** bursts of sound tend to be perceptually organized in different auditory streams when introducing either a spectral [i.e. 1] or a temporal difference between ***A*** and ***B*** [i.e. 2]. The spectral difference between ***A*** and ***B*** sounds can be simply a frequency shift for pure tones or a spectral-pitch shift for complex sounds. In general, the channeling theory of streaming predicts that any salient difference between the excitation patterns evoked by ***A*** and ***B*** sounds would lead to a segregated percept [3]. Some previous studies have however evidenced that a sequence of sounds with similar spectral properties but with different temporal properties can be heard as segregated [i.e. 4]. In particular, temporal cues are undoublty responsible for the segregated percept when hearing a sequence of bursts of white noises that are amplitude-modulated at widely different rates [2]. All previous streaming experiments involved stimuli that had frequency components bellow 5000 Hz. As a consequence, the individual contribution of temporal fine structure cues and envelope cues could not be dissociated and remains largely undetermined. The current experiment is dedicated to test further the relative importance of both envelope and fine structure cues to segregate sequences of unresolved harmonic complex tones with different fundamental frequencies (F_0). Two high-frequency regions above and bellow 5000 Hz and several phase relationships leading to several temporal peak factors have been used in a subjective streaming task.

Experiment

Methods

Stimuli: Temporal sequences with a global duration of 9.6 s have been digitally generated and individually saved on a CD-R. Each sequence was build as the repetition of two alternating complex tones ***A*** and ***B*** using a ***ABA-***... temporal pattern (Figure 1). All ***A*** and ***B*** were 0.1 s long including a 0.01 s rise/fall time. Each triplet ***ABA-*** was separated from the following triplet by a 0.02 s silent gap as illustrated on figure 1.

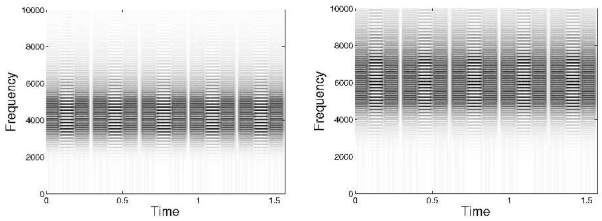

Figure 1: Extracts from two ***ABA-***... sequences filtered in the lower frequency region (condition 2) on the left and in the upper frequency region (condition 5) on the right.

The F_0 of ***A*** was always equal to 100 Hz. The F_0 of ***B*** was fixed within a sequence but varied across sequences from 100 Hz to 400 Hz in 8 ¼-octave steps. Six experimental conditions summarised in table 1 have been tested.

Region	H			H+		
Phase	SCH+	SINE	SCH-	SCH+	SINE	SCH-
Condition	1	2	3	4	5	6

Table 1: Experimental conditions

A and ***B*** were either band-pass filtered from 3535 Hz to 5000 Hz (lower region H, conditions 1, 2 and 3) or from 5000 Hz to 7071 Hz (upper region H+, conditions 4, 5 and 6). These two regions were high-pitched enough to assure all harmonics to be unresolved and to prevent then the perception of any spectral cues [5]. Moreover, the limitation of phase locking in the peripheral auditory system prevent the perception of fine structure cues in H+ but not in H [6]. Moreover, ***A*** and ***B*** harmonic components were generated either in sine phase (SINE, conditions 2 and 5), in positive Schroeder phase (SCH+, conditions 1 and 4) or in negative Schroeder phase (SCH-, conditions 3 and 6). These phase relationships presumably led to distinct physiological temporal envelope peak factors from the highest for SCH+ to the lowest for SCH- [7]. All complex tones were delivered at the SPL level leading to the same loudness than a band-pass noise filtered in the same frequency region and played at 30 dB/Hz. These levels were experimentally determined for each individual and each condition using an adjustment method. All sequences were presented in the right ear along with a continuous low-pass noise (cut-off frequency 1000 Hz) played at 30 dB SPL/Hz to prevent the perception of combination tones.

Procedure: A typical subjective streaming task has been used. The subjects had to report on a response box (RBOX2, PI2) the number of streams they had heard at the end of each sequence presentation. Three blocks of measurements each including the 6 experimental conditions (along with 6 additional conditions that are not reported here) in a random order have been recorded in ten 2-hours sessions. One block consisted of the presentation of 90 sequences in a random order (10 repetitions, F_0 for ***B*** from 100 Hz to 400 Hz).

Material: All digital stimuli were delivered to the subjects through a Tucker Davis Technology system including an analogue converter (AD1), an anti-aliazing filter (FT6-2), two programmable attenuators (PA4), a mixer (SM3), a headphone buffer (HB6) and a Sennheiser HD 250 headphone. The low-pass noise was independently generated by leading the output of a BK 1405 noise generator to an analogue filter (PF1), an attenuator (PA4) and ultimately to the same mixer (SM3). All subjects were comfortably seated in a sound-treated booth that provided an acoustic attenuation better than 70 dB in the frequency range from 125 to 8000 Hz.

Subjects

Four young normal hearing subjects participated. All had pure tones detection thresholds at or better than 20 dB HL in the right ear. All completed the experiment and were paid for their participation. None had previously participated to any psychoacoustical experiment. This experiment have been approved by a local ethical committee (CCPPRB) and all subjects filled and signed an inform consent.

Results and Discussion

The averaged results across subjects for all six conditions are plotted in figure 2.

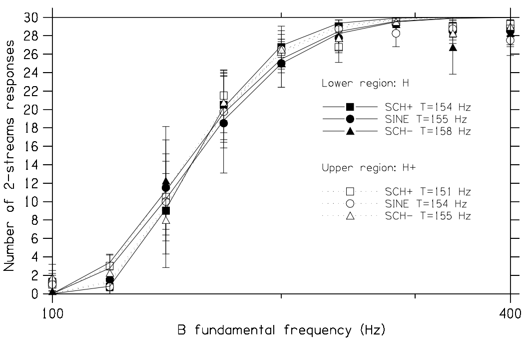

Figure 2: Number of 2-streams responses averaged across subjects for all six conditions. The dotted and continuous lines are the results from the fitting (see text for details).

Both individual and averaged results have been fitted using a two parameter cumulative gaussian curve in order to extrapolate a streaming threshold value defined as the 50% point of 2-streams responses. The streaming thresholds extrapolated from the results plotted in figure 2 are plotted in figure 3.

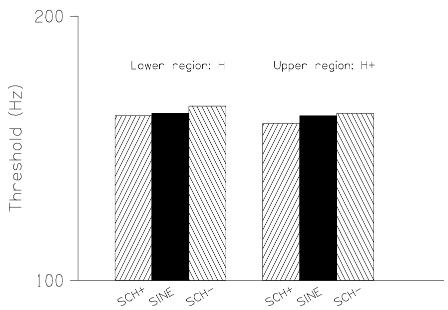

Figure 3: Streaming thresholds extrapolated from the results plotted in figure 2 (see text for details).

A two-way repeated measure ANOVA with factor frequency region and factor phase relationship has been performed on the individual streaming thresholds extrapolated from the individual results. Neither the frequency region [$F(1,3)=0.39$, $p=0.58$] nor the phase relationship [$F(2,6)=0.23$, $p=0.80$] have been showed to have any significant effect in spite of the slight decrease in streaming performances when reducing the envelope peak factor from SCH+ to SCH- (figure 3). Temporal fine structure cues are assumed to be absent in the upper frequency region. The similar streaming performances in this region would then probably indicate a negligible effect of fine structure cues for temporal streaming with unresolved stimuli. This result must however be tempered by the fact that the presence of fine structure cues is strongly correlated with the frequency region. The depreciative effect of the absence of fine structure cues might then be counterbalanced by the broadening of the auditory filters in the upper frequency region leading to additional envelope cues in the upper frequency region.

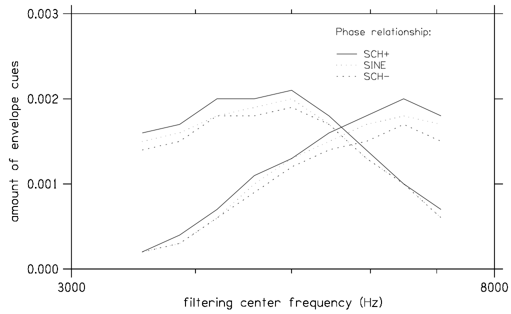

Figure 4: Average magnitude of the slope of the envelope at the output of a gammachirp auditory filterbank.

The amount of envelope cues for each condition has been estimated by computing the average magnitude of the slope of the envelope [8] at the output of a gammachirp auditory filterbank including a phase curvature module to mimic the phase curvature of the cochlea [9]. According to this estimation plotted in figure 5, the amount of envelope cues is in average lower in the upper frequency region than in the lower frequency region. This would presumably indicate that temporal streaming with unresolved complex tones is better based upon temporal envelope cues than upon temporal fine structure cues.

References

[1] van Noorden L.P.A.S. *Temporal coherence in the perception of tone sequences*, Doctoral Dissertation, Technische Hogeschool Eindhovern, The Netherlands (1975).

[2] Grimault N. Bacon S.P. and Micheyl C. *Auditory stream segregation on the basis of amplitude-modulation rate*, J. Acoust. Soc. Am. 111, 1340 (2002).

[3] Hartmann W.M. and Johnson D. *Stream segregation and peripheral channeling*, Mus. Perc. 9, 155 (1991).

[4] Vliegen J. and Oxenham A.J. *Sequential stream segregation in the absence of spectral cues*, J. Acoust. Soc. Am. 105, 339 (1999).

[5] Shackleton T.M. Carlyon R.P. *The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination*, J. Acoust. Soc. Am. 95, 3529 (1994).

[6] Kiang N.Y.S. *Discharge patterns of single fibers in the cat's auditory nerve*, Cambridge Mass., MIT press (1965).

[7] Recio A. and Rhode W.S. *Basilar membrane responses to broadband stimuli*, J. Acoust. Soc. Am. 108, 2281 (2000).

[8] Strickeland E.A. and Viemeister N.F. *Cues for discrimination of envelopes*, J. Acoust. Soc. Am. 99, 3638 (1996).

[9] Oxenham A.J. and Dau T. *Toward a measure of auditory-filter phase response*, J. Acoust. Soc. Am. 110, 3169 (2001).