Effect of harmonic rank on the streaming of complex tones

SARA M. K. MADSEN[1,*], TORSTEN DAU[1], AND BRIAN C.J. MOORE[2]

[1] *Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Department of Psychology, University of Cambridge, Cambridge, England*

The effect of the rank of the harmonics on sequential stream segregation of complex tones was investigated for normal-hearing participants with no musical training. It was hypothesized that stream segregation would be greater for tones with high pitch salience, as assessed by fundamental frequency (f_0) difference limens. Pitch salience is highest for tones containing some low (resolved) harmonics, but is also fairly high for tones containing harmonics of intermediate rank. The tones were bandpass filtered between 2 and 4 kHz and harmonic rank was varied by changing the f_0. There was a significant trend for less stream segregation with increasing harmonic rank. The amount of stream segregation was inversely correlated with the f_0 difference limens, consistent with the hypothesis.

INTRODUCTION

Fundamental frequency (f_0) discrimination, which provides a measure of pitch salience, is better for complex tones with low harmonics (low harmonic rank) than for tones with only high harmonics. This has often been interpreted in terms of spectral resolvability, i.e., f_0 difference limens (f_0DLs) are smaller for complex tones that contain resolved harmonic components (Bernstein and Oxenham, 2006a;b). However, Bernstein and Oxenham (2003) found that f_0DLs were similar for complex tones presented dichotically, with even harmonics presented to one ear and odd harmonics to the other, and for complex tones with all harmonics presented to both ears, even though dichotic presentation should lead to greater resolvability of the harmonics. They argued that harmonic rank and not resolvability is the key factor governing the magnitude of f_0DLs.

The stream segregation of sequences of sounds is facilitated by perceived differences between the sounds, such as differences in frequency, spectrum and f_0 (Moore and Gockel, 2002). This, combined with the fact that f_0 discrimination is better for tones that contain low harmonics, leads to the hypothesis that stream segregation of complex tones would also be affected by harmonic rank. However, this is not consistent with the results of Vliegen and Oxenham (1999), who found that subjective judgements of stream segregation were similar for pure tones, complex tones with low harmonics, and complex tones with only high unresolved harmonics. The present study investigated subjective stream segregation for pure

*Corresponding author: samkma@elektro.dtu.dk

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

tones and complex tones with variable harmonic rank and compared the results to f_0DLs measured for the same stimuli.

EXPERIMENT 1: SUBJECTIVE STREAM SEGREGATION

Method

The stimuli were sequences of ABA triplets where the f_0s of the A and B tones were varied (see Fig. 1). Each tone had a duration of 90 ms, including 10-ms raised-cosine onset and offset ramps, with gaps of 20 ms within each triplet and 110 ms between triplets. Each sequence lasted approximately 8 s and contained 19 triplets.

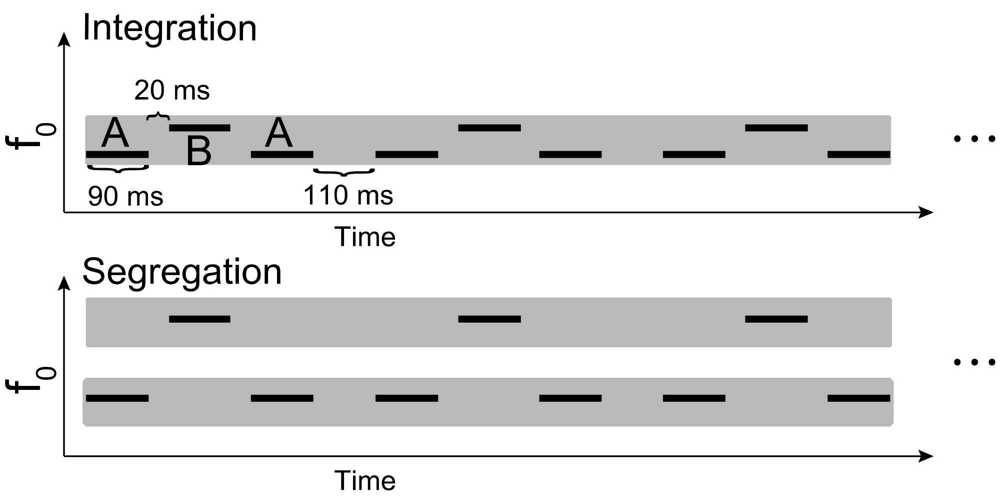

Fig. 1: Schematic illustration of stream segregation and the stimuli used to investigate it.

Both pure tones and complex tones were used. The complex tones were bandpass filtered between 2000 and 4000 Hz. The filter had a spectral slope of 30 dB per octave for frequencies 100 Hz from the edges of the flat bandpass region and 50 dB per octave for frequencies farther away from the passband edges. The harmonic rank of the complex tones was varied by changing f_0. One pure-tone condition with an A-tone frequency of 2000 Hz and five complex tone conditions with A-tone f_0s of 80, 100, 150, 250, and 500 Hz were tested. The B-tone f_0 was always higher than that of the A-tones. Six B-tone f_0s were used with each A-tone f_0, resulting in 36 conditions. The tones had an overall sound pressure level (SPL) of 80 dB and a threshold equalising noise (TEN) with a level of 55 dB SPL/ERN_n was used to mask combination tones and to limit the audibility of stimulus components falling in the filter skirts. An uncorrelated TEN with a level of 25 dB SPL/ERN_n was presented to the other ear.

Nine normal-hearing participants (four female) with audiometric thresholds ≤ 20 dB hearing level (HL) were tested. The participants were between 21 and 28 years of age and had no musical training. The participants were instructed to try to hear out one stream as separate from the other and to indicate via a keyboard key press when their perception changed between one and two streams.

The stimuli were sampled at 44100 Hz and played via a Fireface UXC sound card (RME, Haimhausen, Germany) through Sennheiser HD650 headphones (Wedemark, Germany) in a sound-attenuating booth. The tone stimuli were presented monaurally to the ear that had the lowest mean audiometric threshold across 2, 3, and 4 kHz. Each condition was tested three times in each of 12 blocks. Each participant was trained for at least one two-hour session. The amount of training needed for each participant and the number of sessions was determined based on the mean of the standard deviations calculated for the arcsine-transformed proportions across all trials within a block for each condition. If the value of this measure was greater than 0.2 in one of the three blocks tested within a session, that session was repeated. However, the results were included here if a value greater than 0.2 occurred for only a single block when it was preceded by at least four blocks that had values less than 0.2.

Results

The individual results are shown in Fig. 2; the percentage of time that a given stimulus was perceived as segregated is plotted as a function of the A-tone f_0. Symbols indicate the f_0 difference, Δf_0. The results show large variability across participants. However, there are some general tendencies. The tendency to hear stream segregation usually increased with increasing f_0 difference between the A and B tones. Furthermore, pure tones and complex tones with a high f_0 (more low harmonics) were generally more likely to be perceived as segregated than complex tones with a low f_0.

The effect of f_0 was tested using a one-way within-subjects ANOVA based on a measure of the overall score across all differences in f_0 between the A and B tones after arcsine transformation of the percent scores (in proportions). This measure is called the “normalised segregation score” and its value varies from 0 (no segregation) to 1 (complete segregation). There was a significant effect of f_0 [$F(1,5) = 38.4$, $p < 0.001$]. Bonferoni-corrected pairwise comparisons showed that there were significant differences between the scores for the conditions with f_0s of 500 Hz and 150 Hz [$p = 0.004$], 2000 Hz and 150 Hz [$p = 0.004$], 2000 Hz and 100 Hz [$p < 0.001$], 2000 and 80 Hz [$p < 0.001$], 500 Hz and 100 Hz [$p = 0.002$], 500 Hz and 80 Hz [$p = 0.001$], 250 Hz and 150 Hz [$p = 0.022$], 250 Hz and 100 Hz [$p = 0.0052$], 250 Hz and 80 Hz [$p = 0.0038$], and 150 Hz and 80 Hz [$p = 0.0039$].

EXPERIMENT 2: F_0DLS

Method

Each trial contained three successive tones, two with a base f_0 (or frequency) and one with a higher f_0 or frequency. Each tone had a duration of 500 ms and each was temporally centred in a 700-ms TEN. The tones were separated by a 400-ms gap

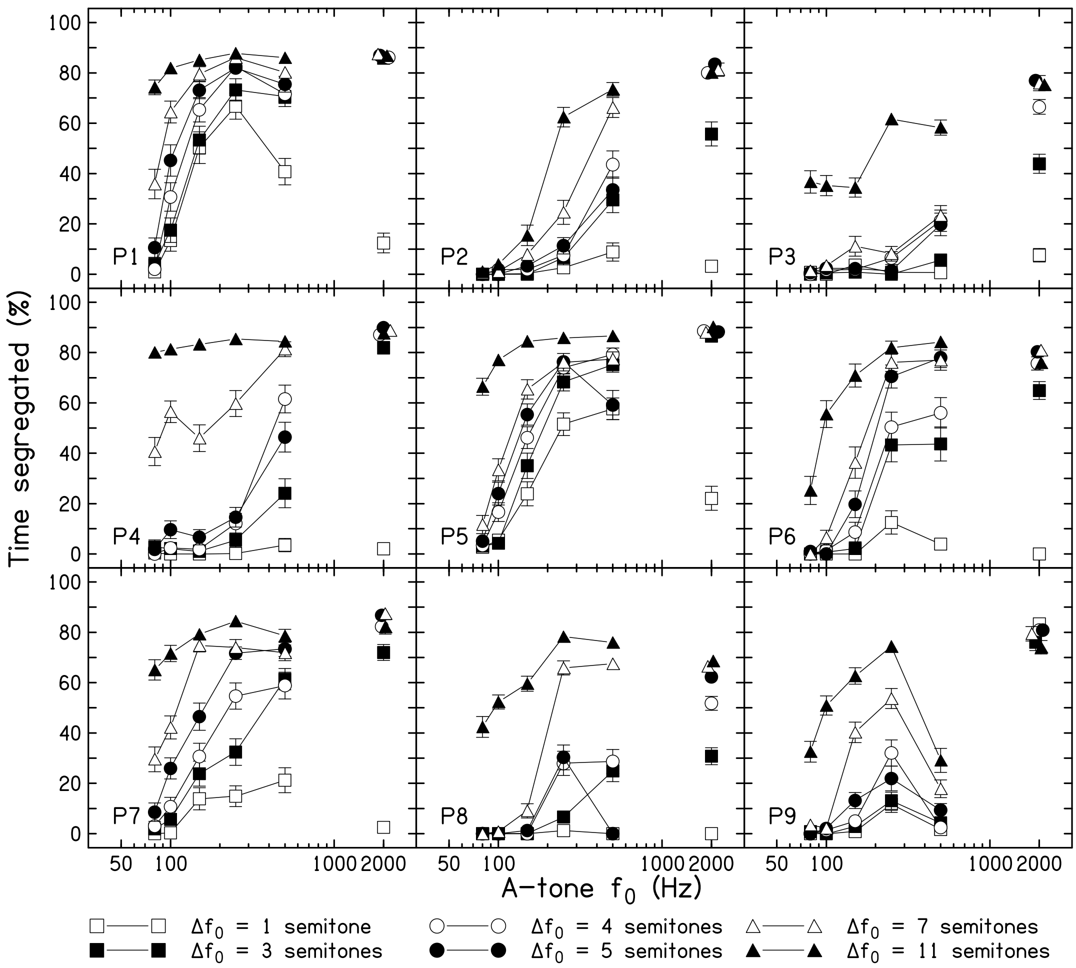

Fig. 2: Percentage of time that participants indicated that they perceived two streams. The mean and standard errors are shown for each participant. Δf_0 is the difference in f_0 between the A and B tones.

The spectrum and level of the tones and the TEN were the same as for experiment 1. The base f_0s of the complex tones were 80, 100, 150, 250, and 500 Hz and the frequency of the pure tone was 2000 Hz, the same as for the A-tones in experiment 1. The participant was instructed to identify the tone with the higher f_0 or higher frequency. A three-interval three-alternative forced-choice weighted up-down adaptive procedure was used to track the 70% correct point on the psychometric function (Kaernbach, 1991). The f_0DL was estimated as the geometric mean of the f_0 differences at the last six reversal points.

As in the study of Bernstein and Oxenham (2006a), the base f_0 was roved over the range ± 5% across trials (uniform distribution) to encourage the participants to listen to the current stimulus instead of comparing the stimulus to the memory of previous stimuli. Also, the levels of the tones were roved within each trial (uniform distribution with a range of ± 2.5 dB) to reduce any loudness cues.

Data are presented for five of the nine participants from experiment 1 since f_0DL measurements have currently not been completed for the remainder. Each condition was tested five times. The participants were trained for one two-hour session and runs were repeated if the standard deviation across the reversal points used to estimate the DL was greater than 0.25.

Results

The f_0DLs are shown in Fig. 3. The f_0DLs varied across participants and were markedly larger for participant nine than for the other participants. Thresholds tended to be lowest (best performance) for the pure tone (2000 Hz) and for the highest f_0 for the complex tones. The f_0DLs were roughly constant for f_0s up to 150 Hz, but decreased somewhat for the f_0 of 250 Hz, even though the lowest harmonic in the passband for the 250-Hz f_0 was the 8th, and this would have been barely, if at all, resolved. This is consistent with the idea that harmonic rank rather than resolvability governs f_0 discrimination (Bernstein and Oxenham, 2003).

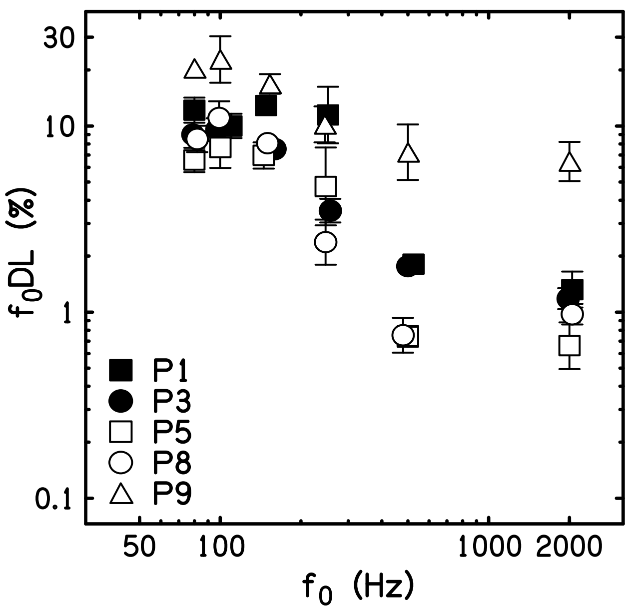

Fig. 3: f_0 discrimination thresholds. Means and standard errors are shown for each participant.

DISCUSSION

The results showed that stream segregation can occur for complex tones with only high harmonics, consistent with results from earlier studies (Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999). However, Vliegen and Oxenham (1999) reported that subjective stream segregation, as measured in our experiment 1, was similar for pure tones, complex tones with low (resolved) harmonics and complex tones with only high (unresolved) harmonics. This contrasts with the results from the present study. Their measurements were made without a noise masker, and so may have been influenced by combination tones, especially for the complex tones with only high harmonics. They also measured stream segregation in the presence and absence of a background noise that would mask combination tones for two conditions, pure tones and complex tones with only high harmonics. Stream segregation seemed to be greater for the pure tones than for the complex tones (as observed in the present study), but this was not discussed in their paper. Furthermore, both the previous and the present study showed large variability across participants, which may also help explain the difference across studies.

Figure 4 shows the normalised segregation score (experiment 1) plotted against the log-transformed f_0DL (experiment 2), for each participant and each f_0 (open symbols). The figure also shows the means across participants (filled squares).

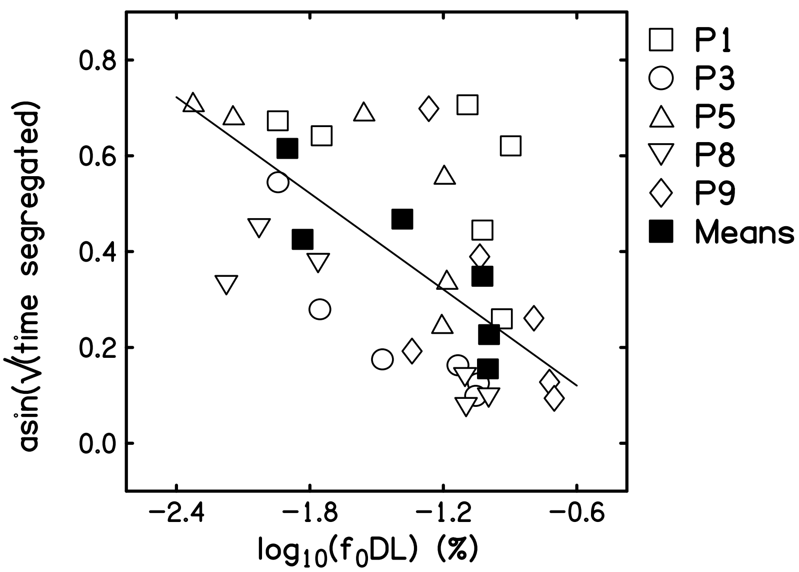

Fig. 4: Arcsine-transformed normalised stream segregation estimate as a function of the log-transformed f_0DL for each participant and the mean across participants (solid squares). The line is a least-squares fit to the mean data.

For the mean across participants, there was a significant Pearson correlation between the normalized segregation score and the f_0DLs [$r = -0.84$, $p = 0.034$], indicating that stream segregation is more likely for tones with small f_0DLs. For the individual participants, the correlations were $r = -0.50$ [$p = 0.31$], $r = -0.89$ [$p = 0.017$], $r = -0.76$ [$p = 0.08$], $r = -0.92$ [$p = 0.01$], and $r = -0.60$ [$p = 0.21$] for P1, P3, P5, P8, and P9, respectively. All correlations were negative, confirming that small f_0DLs, indicating high pitch salience, are associated with greater stream segregation.

CONCLUSIONS

There was a significant effect of harmonic rank on the tendency for a sequence of complex tones to be heard as two streams. More stream segregation occurred for complex tones with resolved harmonics than for complex tones with unresolved harmonics. Also, stream segregation was greater for complex tones that led to low f_0DLs, suggesting that good f_0 discrimination is associated with greater streaming.

ACKNOWLEDGEMENTS

This research was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS). BCJM was supported by the Rosetrees Trust.

REFERENCES

Bernstein, J.G. and Oxenham, A.J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number?," J. Acoust. Soc. Am., **113**, 3323-3334.

Bernstein, J.G. and Oxenham, A.J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: effects of stimulus level," J. Acoust. Soc. Am., **120**, 3916-3928.

Bernstein, J.G. and Oxenham, A.J. (**2006b**). "The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss," J. Acoust. Soc. Am., **120**, 3929-3945.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys., **49**, 227-229.

Moore, B.C.J. and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust., **88**, 320-333.

Vliegen, J., Moore, B.C.J., and Oxenham, A.J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am., **106**, 938-945.

Vliegen, J. and Oxenham, A.J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am., **105**, 339-346.