# Effect of harmonic rank on sequential sound segregation

**Madsen, Sara Miay Kim; Dau, Torsten; Moore, Brian C.J.**

*Published in:*
Hearing Research

*Link to article, DOI:*
10.1016/j.heares.2018.06.002

*Publication date:*
2018

*Document Version*
Peer reviewed version

*Citation (APA):*
Madsen, S. M. K., Dau, T., & Moore, B. C. J. (2018). Effect of harmonic rank on sequential sound segregation. *Hearing Research*, *367*, 161-168. https://doi.org/10.1016/j.heares.2018.06.002

---

# Effect of harmonic rank on sequential sound segregation

Sara M.K. Madsen, Torsten Dau, Brian C.J. Moore

PII: S0378-5955(18)30115-1

DOI: 10.1016/j.heares.2018.06.002

Reference: HEARES 7570

To appear in: *Hearing Research*

Received Date: 24 March 2018

Revised Date: 1 June 2018

Accepted Date: 8 June 2018

Please cite this article as: Madsen, S.M.K., Dau, T., Moore, B.C.J., Effect of harmonic rank on sequential sound segregation, *Hearing Research* (2018), doi: 10.1016/j.heares.2018.06.002.

# Effect of harmonic rank on sequential sound segregation

Sara M.K. Madsen[1,*], Torsten Dau[1], Brian C.J. Moore[2]

[1] *Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark*

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

* Corresponding author. Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark.

*E-mail address:* samkma@elektro.dtu.dk (S.M.K. Madsen)

Revised version submitted June 2018

*Keywords:* Stream segregation, Fundamental frequency, Fundamental frequency discrimination

## ABSTRACT

The ability to segregate sounds from different sound sources is thought to depend on the perceptual salience of differences between the sounds, such as differences in frequency or fundamental frequency (F0). F0 discrimination of complex tones is better for tones with low harmonics than for tones that only contain high harmonics, suggesting greater pitch salience for the former. This leads to the expectation that the sequential stream segregation (streaming) of complex tones should be better for tones with low harmonics than for tones with only high harmonics. However, the results of previous studies are conflicting about whether this is the case. The goals of this study were to determine the effect of harmonic rank on streaming and to establish whether streaming is related to F0 discrimination. Thirteen young normal-hearing participants were tested. Streaming was assessed for pure tones and complex tones containing harmonics with various ranks using sequences of ABA triplets, where A and B differed in frequency or in F0. The participants were asked to try to hear two streams and to indicate when they heard one and when they heard two streams. F0 discrimination was measured for the same tones that were used as A tones in the streaming experiment. Both streaming and F0 discrimination worsened significantly with increasing harmonic rank. There was a significant relationship between streaming and F0 discrimination, indicating that good F0 discrimination is associated with good streaming. This supports the idea that the extent of stream segregation depends on the salience of the perceptual difference between successive sounds.

Keywords: stream segregation, pitch, fundamental frequency discrimination, perceptual differences

## 1 Introduction

The ability to segregate sounds from different sound sources is thought to depend on the perceptual salience of differences between the sounds, such as differences in frequency or fundamental frequency (F0) (Moore and Gockel, 2002; Paredes-Gallardo et al., 2018). It is therefore easier to understand speech produced by a female speaker in the presence of one or more male speakers than when in the presence of other female speakers (Brungart et al., 2001). The ability to segregate sounds into different auditory objects is used constantly in daily life and is essential for understanding speech in the presence of background sounds. The ability also makes it possible to hear out individual instruments or voices in music.

Speech and music are complex signals and sound segregation is often investigated using simpler, more controlled, stimuli such as sequences of interleaved A and B sounds where A and B differ in some way (e.g., Bregman, 1990; van Noorden, 1975). These sequences can be heard either as one stream (integrated) or as two streams (segregated). The perceptual construction of two streams is called sequential stream segregation or streaming. Several studies have used such sequences to explore the effect of differences between the A and B sounds in frequency (pure tones) or in F0 (complex tones) and have shown that the ability to segregate increases with increasing frequency or F0 difference (e.g., Grimault et al., 2000; Grimault et al., 2001; Rose and Moore, 1997; van Noorden, 1975; Vliegen and Oxenham, 1999; Vliegen et al., 1999).

Studies of F0 discrimination have shown that F0 difference limens (F0DLs) are relatively small when the tones contain low harmonics (with harmonic numbers, also called ranks, up to about 8), but increase when the rank of the lowest harmonic increases above about 8, indicating that pitch salience decreases when only high-rank harmonics are present (e.g., Bernstein and Oxenham, 2006a; Hoekstra and Ritsma, 1977; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). The increase in F0DLs with increasing harmonic rank might be explained by better resolution of lower than of higher harmonics (Bernstein and Oxenham, 2006b; Shackleton and Carlyon, 1994). However, some lines of evidence suggest that resolution of harmonics is not the key factor. Firstly, for very low F0s, the harmonics that dominate the pitch percept are not the lowest resolved harmonics (Jackson and Moore, 2013). Secondly, Bernstein and Oxenham (2003) compared F0DLs for tones with all harmonics presented to both ears (diotic) and tones with odd harmonics presented to one ear and even harmonics to the opposite ear (dichotic). If F0 discrimination were governed by the degree of resolution of the harmonics, performance should have been better for the dichotic condition, since the frequency separation of harmonics within each ear was twice as large as for the diotic condition. In fact, F0DLs were similar for the diotic and dichotic conditions. The results suggest that harmonic rank *per se* is important. The effect of harmonic rank has been explained by ‘place dependence’, i.e. for each place in the cochlea (corresponding to a specific auditory filter with a certain center frequency) there is a limited range of periodicities that can be analyzed, and this range is closely tied to the center frequency of that filter (Bernstein and Oxenham, 2005; Moore, 2003).

If stream segregation depends on the salience of the perceptual differences between successive sounds (Hartmann and Johnson, 1991; Moore and Gockel, 2002; Paredes-Gallardo et al., 2018), one might expect that the ease with which a sequence of complex tones (tones A and B, differing in F0) can be segregated into streams would be affected by pitch salience (strength). If so, then for a fixed difference in F0 between successive tones, stream segregation should be more likely to occur for tones containing low harmonics than for tones containing only high harmonics. A few studies have investigated the effect of harmonic rank on streaming, but with differing results. Vliegen and Oxenham (1999) measured sequential stream segregation for pure tones, complex tones with low harmonics, and complex tones with only high harmonics. For each of these, the F0 of the B tone was between one and 11 semitones higher than the F0 of the fixed A tone. The listeners were instructed to try to hear the sequence as segregated and to indicate whether they heard each sequence as one or two streams. The proportion of trials that were perceived as segregated was similar for all conditions, indicating no effect of harmonic rank. Grimault et al. (2000) measured streaming for complex tones with fixed F0s for the A and B tones. The tones were filtered into three regions (low, mid, and high) to vary the ranks of the harmonics in the tones. They found that the percentage of segregation decreased with increasing harmonic rank and argued that this was an effect of the resolvability of the harmonics in the tones. They did not instruct the listeners to try to hear the streams as segregated or integrated, as in the study of Vliegen and Oxenham (1999). The instruction to try to segregate used by Vliegen and Oxenham might have increased the proportion of segregation, especially for the difficult conditions with only high harmonics, and Grimault et al. (2000) suggested that the difference in instructions might explain the difference between studies. Also, they proposed that the difference across studies might be explained by their conditions being more extreme in terms of resolvability than the ones used by Vliegen and Oxenham (1999). If so, this would indicate that large differences in harmonic rank are required to reveal differences in stream segregation.

The aims of the present study were: (1) to determine the effect of harmonic rank when the listeners were instructed to try to hear the sequence as segregated; 2) to establish whether there is a relationship between F0DLs and streaming. Sequential stream segregation was investigated for pure tones and complex tones with harmonic ranks ranging from low (with well resolved harmonics) to high (with all harmonics clearly unresolved), i.e. representing conditions with harmonic rank less than 8 or larger than 14, respectively (Moore and Gockel, 2011). Preliminary data from this study was previously presented in a conference paper (Madsen et al., 2015).

# 2 General method

## 2.1 Listeners

Thirteen normal-hearing listeners (audiometric thresholds ≤20 dB HL at octave frequencies between 250 and 8000 Hz; five females, eight males) between 21 and 27 years of age (mean = 23.6 years, SD = 1.6 years) were tested. The listeners had no musical training. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark.

## 2.2 Stimulus generation and presentation

The stimuli were generated in MATLAB at a sampling rate of 44100 and presented via a Fireface UCX sound card (RME, Haimhausen Germany) and Sennheiser HD 650 headphones (Sennheiser, Wedemark, Germany). All stimuli were presented monaurally at a sound pressure level (SPL) of 80 dB to the ear with the lowest audiometric threshold averaged across the frequencies 2, 3, and 4 kHz. This level was chosen since this study was meant to be the first in a series of experiments in which hearing-impaired listeners would also be tested. This allows the comparison of results for normal-hearing and hearing-impaired listeners at the same sound pressure level. All measurements were made in an acoustically shielded booth.

# 3 Experiment 1: Sequential stream segregation

## 3.1 Rationale

The goal of this experiment was to determine whether subjective sequential stream segregation is affected by harmonic rank. F0 discrimination is better for tones with low harmonic rank and it was therefore hypothesized that the presence of low harmonics would facilitate the segregation of sequences of complex tones.

## 3.2 Method

[Insert Fig. 1 approximately here]

### *3.2.1 Stimuli*

The stimuli consisted of sequences of ABA-ABA tones where A and B are different tones and “-” represents a brief pause. This type of stimulus has been used in many experiments on stream segregation (e.g., Bregman, 1990; van Noorden, 1975). As illustrated in Fig. 1A, such a sequence can be perceived as one stream (upper panel; integration) that is heard as having a galloping rhythm or as two separate streams, one twice as fast as the other (lower panel; segregation). As in the study of Vliegen and Oxenham (1999), each tone had a duration of 90 ms including 20-ms raised-cosine ramps. The time interval between tones within each triplet was 10 ms and consecutive triplets were separated by 110 ms. Each tone sequence consisted of 19 triplets and had a duration of approximately 8 s.

Both the A and B tones were either complex tones or pure tones. As illustrated in Fig. 1B, the complex tones were initially generated to contain all harmonics with equal amplitude, added in sine phase. The tones were then bandpass filtered between 2 and 4 kHz (3-dB down points), using a filter slope of 30 dB/octave for the first 100 Hz on each side of the flat passband and 50 dB/octave beyond that range. The edge frequencies of the passband were 2125 and 3798 Hz. The filter slope was chosen to avoid abrupt changes in level of individual harmonics as they passed into and out of the passband when the F0 was changed. The harmonic rank was varied by varying the F0; the higher the F0 the lower was the harmonic rank. For the pure-tone stimuli, the frequency of the A tone was 2000 Hz. For the complex tones, the A-tone F0 was 80, 100, 150, 250, or 500 Hz. Hence, the rank of the lowest harmonic in the passband varied from 27 (F0 = 80 Hz) to 5 (F0 = 500 Hz). The B-tone frequency or F0 was always higher than that of the A tone. The frequency or F0 difference between the A and B tones ($\Delta$F0) was 1, 3, 4, 5, 7, or 11 semitones (ST), resulting in 36 conditions. The frequencies or the F0s of the A and B tones were fixed within each trial.

A threshold-equalizing noise (TEN) (Moore et al., 2000) was used to mask combination tones and to limit the audibility of stimulus components falling on the filter skirts. According to Oxenham et al. (2009) the $2f_1$-$f_2$ combination tone produced by interaction of the two lowest components in the passband may just be audible when the component level is 15 dB higher than the TEN level, expressed as dB SPL/ERB$_N$, where ERB$_N$ is the average value of the equivalent rectangular bandwidth of the auditory filter for listeners with normal hearing (Glasberg and Moore, 1990). The component level needs to be about 30 dB higher than the TEN level for the next lower combination tone to be audible. The present study used a TEN level of 55 dB SPL/ERB$_N$, which meant that the level of each component in the complex tones was 20-24 dB higher than the level/ERB$_N$ of the TEN. Hence, the $2f_1$-$f_2$ combination tone may have been just audible, but no lower combination tones were audible. This does not create a problem in the interpretation of the results presented here, since the only consequence of the $2f_1$-$f_2$ combination tone being audible would be to lower the harmonic rank by one. This would not affect whether the tones in the different conditions were resolved or unresolved.

### *3.2.2 Procedure*

The aim was to assess the proportion of time that two streams were perceived when listeners were actively trying to segregate the sequence. The listeners were therefore asked to try to hear the sequence as segregated and to press one key when they heard one stream and a different key when they heard two streams. They could switch between the two keys during presentation of a sequence if the percept appeared to change. The listeners were trained for at least two hours and tested in four 2-hour sessions. Each condition was tested 36 times for each listener in blocks that each contained one presentation of each condition. The conditions were randomized such that the order of conditions within a block was always different across blocks for each listener. The order was different for each listener. Nine blocks were tested in each session.

To ensure that the listeners had been sufficiently trained, the standard deviation of the streaming scores (percentage of time that two streams were reported) for each condition was calculated across each set of three successive blocks and then averaged across conditions. If the mean standard deviation was larger than or equal to 20% for at least one of the three sets of three blocks tested in the first test session, these blocks were considered as training and they were repeated in the following session.

### *3.2.3 Statistical analysis*

Due to large deviations from normality, the data were transformed using the aligned rank transform (Wobbrock et al., 2011) and then analyzed with a linear mixed-effects model with harmonic rank and ΔF0 as fixed factors and listener as a random factor, using the ARTool library (Wobbrock et al., 2011) in R. Post-hoc analysis was performed using the lsmeans library (Lenth, 2016) and Tukey corrections were used to correct for multiple comparisons.

## *3.3 Results*

Subjective sequential stream segregation was assessed as the proportion of time that the listeners indicated hearing two streams (no galloping rhythm), assessed over the whole duration of the sequence. Figure 2 shows the individual data and Fig. 3 shows the mean data. Results for the complex tones are plotted on the left as a function of F0 and results for the pure tones are plotted on the right. While there were large individual differences, the streaming scores generally increased with increasing ΔF0 or ΔF (pure tones) and with increasing F0, i.e. decreasing harmonic rank. All conditions, including the pure tone conditions, were included in the analysis. Both main effects were significant (ΔF0: $F(5, 420) = 142.77$, $p < 0.001$; harmonic rank: $F(5, 420) = 205.34$, $p < 0.001$) and the interaction was also significant ($F(25, 420) = 9.96$, $p < 0.001$). Pairwise comparison of conditions with different F0 (harmonic rank) showed that the differences between all pairs were significant ($p < 0.01$) except between F0 = 500 and 250 Hz. Similarly, pairwise comparison of conditions with different ΔF0 showed that all differences were significant ($p < 0.01$) except between ΔF0 = 4 and 5 ST.

## *3.4 Discussion*

It is possible that the listeners judged the perceptual difference between the A and B tones rather than judging stream segregation per se. To assess this possibility, it was determined whether a build up of “two-stream” responses occurred over time, since build up is generally regarded as a key characteristic of stream segregation (e.g., Anstis and Saida, 1985; Bregman, 1978b). This was done for a condition leading to an intermediate percentage of two-stream responses (ΔF0 = 7 and F0 = 250 Hz) to avoid floor and ceiling effects. Figure 4 shows the percentage of time where the two-stream key was pressed after every half second for each of the listeners. As expected, the proportion of two-stream responses increased with time for most listeners, confirming that judgements were based on stream segregation rather than on the perceptual difference between the A and B tones.

[Please insert Fig. 4 approximately here]

The significant increase in segregation with increasing ΔF0 is consistent with the results of many other studies (e.g., Grimault et al., 2000; Grimault et al., 2001; Rose and Moore, 1997; van Noorden, 1975; Vliegen and Oxenham, 1999; Vliegen et al., 1999) and with the idea that the extent of stream segregation increases with increasing perceptual difference between successive sounds (Moore and Gockel, 2002). This idea is also supported by the decrease in stream segregation with increasing harmonic rank. The harmonic rank was varied by varying the F0. In theory, therefore, the observed effects could be a result of variations in F0 rather than variations in harmonic rank. However, this seems unlikely, since F0DLs for sounds with fixed harmonic content (e.g. harmonics 1-5 or 6-12) are similar (when expressed as Weber fractions) for F0s within the range tested in this study (e.g., Moore and Moore, 2003).
The effect of harmonic rank found here differs from that reported by Vliegen and Oxenham (1999) but is consistent with the findings of Grimault et al. (2000). However, the results from the present study are not consistent with the suggestions made by Grimault et al. (2000) to explain the difference between their results and those of Vliegen and Oxenham (1999). Firstly, the present results showed an effect of harmonic rank when the listeners were instructed to try to segregate, as was done by Vliegen and Oxenham (1999) but not by Grimault et al. (2000). Secondly, streaming differed between conditions that did not differ greatly in terms of the resolvability of the harmonics in the complex tones. For example, the harmonics can be assumed to be mostly completely unresolved for the F0s of 80 and 100 Hz

(lowest harmonics in the passband were 27 and 22, respectively, for the A tones and 15 and 12, respectively, for the B tones for ΔF0 = 11 ST), but streaming differed significantly for these two conditions.

One difference between the present study and that of Vliegen and Oxenham (1999) is that segregation here was quantified as the percentage of time that the listeners indicated that they heard two streams, measured over the whole duration of the sequence, while Vliegen and Oxenham (1999) obtained a single response for each sequence, presumably made towards the end of the sequence or after the sequence was finished. Stream segregation tends to build up over time for stimuli with small perceptual differences between successive sounds (Anstis and Saida, 1985; Bregman, 1978a) but can occur very rapidly when there are large perceptual differences. Using the measure of segregation of the present study, this build-up effect might have had a greater influence for stimuli where the build up was slow (small perceptual differences) than for stimuli where the build up was rapid (large perceptual differences), thus increasing differences across conditions. In the study of Vliegen and Oxenham (1999), the build up was probably near-complete for all stimuli. This might have contributed to the difference across studies.

To assess this possibility, the percentage of trials for which the two-streams key was the last key pressed was determined, giving a measure similar to that of Vliegen and Oxenham (1999). Analysis with a logistic generalized mixed-effects model for binary data using the lme4 library in R (Bates et al., 2015) showed significant effects of harmonic rank ($\chi^2(5) = 100.04$, $p < 0.001$) and of ΔF0 ($\chi^2(5) = 822.96$, $p < 0.001$) and a significant interaction ($\chi^2(25) = 402.52$, $p < 0.001$). Thus, it does not seem that the measure used here to assess the amount of segregation can explain the difference in results across studies.

Another difference between the two studies is the sequence length. In this study, each sequence contained 19 triplets whereas Vliegen and Oxenham (1999) used 12 triplets per sequence. However, since segregation builds up slowly over time when perceptual differences are small (Anstis and Saida, 1985; Bregman, 1978a), it would have been less likely to occur for conditions with high harmonic rank in the study of Vliegen and Oxenham (1999) than in the present study, so this factor also cannot explain the difference between the results of the two studies.

Another factor that might have influenced the results is combination tones. The present study used a TEN to mask combination tones while Vliegen and Oxenham (1999) did not use any noise in their main experiment. Vliegen and Oxenham (1999) presented a preliminary experiment showing a small deleterious effect of masking noise on the stream segregation of complex tones containing only high harmonics, but a similar effect occurred for pure tones. They argued that combination tones were unlikely to explain their results. The results of the preliminary experiment, did, however, generally show more segregation for pure tones than for the complex tones, which is similar to the findings of this study but different from the results of their main study. Vliegen and Oxenham (1999) argued that this difference “may simply illustrate the large inter-subject variability”. In the present study, the results also varied markedly across listeners, so it is possible that inter-listener variability can explain the difference between studies. However, the fact that all listeners in the present study showed some effect of harmonic rank and the fact that Grimault et al. (2000) also found a significant effect of harmonic rank indicate that stream segregation does worsen with increasing harmonic rank, at least for most listeners.

The results of the present study are also consistent with studies that investigated F0 discrimination for pairs of tones preceded and followed by complex tones with fixed F0 (fringes) (Gockel et al., 1999; Micheyl and Carlyon, 1998). In these studies, it was proposed that the fringes interfere with F0 discrimination when the fringes and target tones are perceived as a single stream, but that interference is small when the fringes are perceived as being in a separate stream from the target tones. The results showed that when the mean F0 of the fringes differed from that of the target, there was more interference when both fringes and target tones contained unresolved harmonics than when they both contained resolved harmonics. This suggests that stream segregation of the fringes and target was more likely to occur when they both contained resolved harmonics, which is consistent with the results presented here.

The results from the present study confirm that stream segregation is possible for complex tones without resolved components. This is consistent with results from several studies showing that stream segregation can be induced using temporal cues alone, without any excitation-pattern cues (e.g., Dannenbring and Bregman, 1976; Grimault et al., 2002; Hong and Turner, 2009; Paredes-Gallardo et al., 2018; Roberts et al., 2002; Stainsby et al., 2004; Vliegen et al., 1999).

The present study found that the percentage of segregation increased with decreasing harmonic rank (increasing F0; Fig. 2) except that there was no difference between F0s of 250 and 500 Hz. The mean streaming scores were very similar for those two conditions. Most individual scores were also similar for these conditions, but a few listeners (L2, L6 and L11) showed consistent decreases in streaming when the F0 was increased from 250 to 500 Hz. These decreases may be explained by the relatively small number of harmonics in the conditions with the A-tone F0 = 500 Hz. Assuming that all harmonics with a level of 55 dB SPL or above (which was the level/ERB$_N$ of the TEN) were audible, the A tone had six audible harmonic components and the number of audible harmonics in the B tone decreased with increasing ΔF0. For the conditions with ΔF0 = 7 and 11 ST, the B tones had only four audible harmonics. Due to the limited number of well-resolved harmonics, a few listeners may have heard individual harmonics (spectral pitch) instead of the fundamental pitch of the tone complex (Schneider et al., 2005). They may have focused their attention on non-corresponding harmonics in the A and B tones. For example, when ΔF0 = 11 ST they may have attended to the 4th harmonic of the A tone (2000 Hz) and the second harmonic of the B tone (1879 Hz), which might have led to reduced segregation, since these harmonics differ in frequency by only slightly more than 1 ST.

## 4 Experiment 2: Relation between stream segregation and discrimination of pure tones and complex tones

### *4.1 Rationale*

F0DLs were measured to determine the relationship between streaming and the salience of the F0 differences between the A and B tones. Furthermore, F0DLs were measured for tones whose harmonics were added in sine phase or in random phase, to provide an indirect measure of the resolvability of the harmonics. It is generally assumed that harmonic phase has an influence on F0DLs only when the harmonics interfere, and therefore are at least partly unresolved (Houtsma and Smurzynski, 1990; Moore, 1977; Wang et al., 2012). The outputs of auditory filters in response to tones with unresolved harmonics have a higher peak factor for sine-phase tones than for random-phase tones. This is expected to affect F0DLs based on the use of envelope cues. Therefore, F0DLs are expected to be smaller for sine-phase than for the random-phase tones when all harmonics are unresolved.

### *4.2 Method*

#### *4.2.1 Stimuli*

F0DLs were measured for pure tones and complex tones similar to the ones used in experiment 1. The tones were bandpass filtered between 2 and 4 kHz and the nominal F0s of the reference tones were the same as the F0s of the A tones in experiment 1. Each tone had a duration of 500 ms including 10-ms raised-cosine onset and offset ramps. The interval between the three tones in each trial was 250 ms. The stimuli were presented at the same level and in the same TEN as for experiment 1. The TEN had the same purposes as for experiment 1. In addition, it was intended to promote synthetic rather than analytic listening (listening to the pitch corresponding to the missing F0 rather than to individual harmonics), since background noise promotes synthetic listening (Hall and Peters, 1981; Houtgast, 1976).

#### *4.2.2 Procedure*

F0DLs were measured using a 3-alternative-forced-choice (AFC) weighted up-down paradigm (Kaernbach, 1991) to estimate the 75% point on the psychometric function. The listeners were asked to indicate which of the intervals contained the tone with the different pitch (the deviant). The F0 of this tone was always higher than the reference F0. The reference F0 was roved by ±5 % between trials using a uniform distribution around the nominal value. For each run, the initial F0 difference between the reference and the deviant $(F0_{\mathrm{deviant}} - F0_{\mathrm{reference}})/F0_{\mathrm{reference}}$ was 20%. In the following trials, the F0 difference was decreased logarithmically by a step size that decreased after every second reversal. The F0DL was calculated as the geometric mean of the F0 difference at the last six out of 10 reversals. Each condition was tested twice during training and the final F0DLs were calculated from values obtained over five runs. To check whether the F0DLs had stabilized after training, a straight line was fitted to the five F0DLs for each condition and one more block for each condition was added if the slope of the line was significantly lower than 0 for more than two conditions.

### *4.3 Results*

As shown in Fig. 5, the F0DLs decreased (improved) with increasing F0 (decreasing harmonic rank) for conditions with both sine and with random phase and were larger for random than for sine phase for the conditions with high harmonic rank but similar across phases for the conditions with lower harmonic rank. Analysis using a mixed-effects model with F0 and phase as fixed factors and listener as a random factor confirmed that both main effects (F0: $F(4, 628) = 223.9$, $p < 0.001$, phase: $F(1, 628) = 32.77$, $p < 0.001$) and the interaction ($F(4, 628) = 10.61$, $p < 0.001$) were significant. Comparisons of pairs of conditions with different phase but the same F0 (adjusted for multiple comparisons controlling the false discovery rate (Benjamini and Hochberg, 1995)) showed significant effects of phase for F0s of 80 Hz ($t(628) = -6.25$, $p < 0.001$), 100 Hz ($t(628) = -5.46$, $p < 0.001$) and 150 Hz ($t(628) = -2.23$, $p = 0.031$) but not for the higher F0s, consistent with the idea that an effect of component phase occurs when the harmonics are unresolved.

Both stream segregation and F0 discrimination improved with increasing F0 (decreasing harmonic rank). The left panel of Fig. 6 illustrates this relationship. In this scatter plot, the mean percentage segregation for each A-tone F0 (averaged geometrically across ΔF0s and across listeners) is plotted against the mean F0DL (across listeners) obtained for the same F0. There was a strong negative Pearson correlation between the two measures ($r = -0.95$, $p = 0.002$, one tailed, since a negative correlation was hypothesized), indicating that small F0DLs are associated with greater streaming. To investigate the relationship between stream segregation and F0DLs for the individual listeners, for each listener the mean segregation score was plotted against the mean F0DL (right panel of Fig. 6). There was a general tendency for stream segregation to decrease with increasing F0DL indicating that good F0 discrimination for an individual is associated with greater segregation for that individual. The Pearson correlation was moderate but significant ($r = -0.54$, $p = 0.03$, one tailed).

### *4.4 Discussion*

The increase in F0DLs with increasing harmonic rank and the effect of phase seen in Fig. 5 are consistent with what has been found in earlier studies (e.g., Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b; Houtsma and Smurzynski, 1990; Wang et al., 2012). The better performance for sine than for random phase for tones containing only high harmonics is thought to reflect the use of envelope cues resulting from the interference of harmonics in the cochlea. No phase effects are expected when one or more resolved harmonics are present, since performance is then dominated by the resolved harmonic(s). The results therefore suggest that the complex tones with F0s of 80, 100 and 150 Hz did not contain any resolved harmonics whereas the tones with higher F0s did. However, the F0DLs for the random-phase tones did increase significantly as the F0 decreased from 150 to 80 Hz ($t(628) = 5.25$, $p<0.001$), suggesting that F0 discrimination worsens with increasing harmonic rank even when all harmonics are unresolved. This is consistent with the idea that the worsening of F0DLs with increasing harmonic rank reflects an effect of harmonic rank *per se*, rather than an effect of resolvability.

Figure 6 shows a clear relationship between stream segregation and F0 discrimination, supporting the idea that the extent of stream segregation depends on the salience of the perceptual difference between successive sounds. This is consistent with result from two recent studies that both showed a relationship between pitch salience and sequential stream segregation performance (Paredes-Gallardo et al., 2018; Shearer et al., 2018).

Some studies have shown a significant relationship between speech-in-speech perception and performance in a stream segregation task (Gaudrain et al., 2012; Hong and Turner, 2006; Mackersie et al., 2001). This raises the possibility that F0 discrimination might be related to speech-in-speech perception. Furthermore, musical training is associated with enhanced frequency discrimination and F0 discrimination (e.g., Bianchi et al., 2016; Brown et al., 2017; Madsen et al., 2017; Micheyl et al., 2006; Ruggles et al., 2014), so it is possible that musical training would be associated with better stream segregation and better speech in speech perception. However, two recent studies have shown that musicians are not better than non-musicians at using F0 differences between competing voices to understand speech (Deroche et al., 2017; Madsen et al., 2017) and the latter specifically found no relationship between F0DLs and speech-in-speech perception.

## 5 Overall summary and conclusions

Experiment 1 investigated the effect of harmonic rank on the subjective sequential stream segregation of complex tones in a task where the listeners were instructed to try to segregate. Stream segregation scores were compared to F0DLs measured using similar stimuli in experiment 2.

The results of experiment 1 showed that: (1) segregation increased with decreasing harmonic rank; (2) the effect of harmonic rank was continuous and progressive and even small differences in harmonic rank led to differences in segregation.

In experiment 2, F0DLs were measured for pure tones and complex tones similar to the A-tones used in experiment 1. F0DLs increased with increasing harmonic rank. Significant correlations were found between the mean percentage segregation for each A-tone F0 (averaged across ΔF0s and across listeners) and the mean F0DL (across listeners) obtained for the same F0 and between the mean percentage segregation for each listener (averaged across all conditions) and the mean F0DL for that listener (averaged across all F0s). This supports the idea that the extent of stream segregation of successive sounds depends on the salience of the perceptual difference between those sounds.

## Acknowledgements

This study was supported by the Carlsberg Foundation. We thank Andrew Oxenham, Bob Carlyon, Federica Bianchi, Sebastién Santurette and two reviewers for useful discussions and comments.

## References

Anstis, S., Saida, S. 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. Hum. Percept. Perform. 11, 257-271.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1-48.

Benjamini, Y., Hochberg, Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing J. R. Stat. Soc. Ser. B-Methodol. 57, 289-300.

Bernstein, J.G., 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., Oxenham, A.J. 2005. An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. J. Acoust. Soc. Am. 117, 3816-3831.

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

Bianchi, F., Santurette, S., Wendt, D., Dau, T. 2016. Pitch discrimination in musicians and non-musicians: effects of harmonic resolvability and processing effort. Assoc. Res. Otolaryngol. 17, 69-79.

Bregman, A.S. 1978a. Auditory streaming is cumulative. J. Exp. Psychol. Hum. Percept. Perform. 4, 380-387.

Bregman, A.S. 1978b. The formation of auditory streams. In: Requin, J., (Ed.), Attention and Performance VII. Erlbaum, Hillsdale, NJ.

Bregman, A.S. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. Bradford Books, MIT Press, Cambridge, Mass. pp. 455-528.

Brown, C.J., Jeon, E.K., Driscoll, V., Mussoi, B., Deshpande, S.B., Gfeller, K., Abbas, P.J. 2017. Effects of long-term musical training on cortical auditory evoked potentials. Ear Hear. 38, E74-E84.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R. 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527-2538.

Dannenbring, G.L., Bregman, A.S. 1976. Stream segregation and the illusion of overlap. J. Exp. Psychol. Hum. Percept. Perform. 2, 544-555.

Deroche, M.L.D., Limb, C.J., Chatterjee, M., Gracco, V.L. 2017. Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. J. Acoust. Soc. Am. 142, 1739-1755.

Gaudrain, E., Grimault, N., Healy, E.V., Bera, J.C. 2012. The Relationship Between Concurrent Speech Segregation, Pitch-Based Streaming of Vowel Sequences, and Frequency Selectivity. Acta Acust. United Acust. 98, 317-327.

Glasberg, B.R., Moore, B.C.J. 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103-138.

Gockel, H., Carlyon, R.P., Micheyl, C. 1999. Context dependence of fundamental-frequency discrimination: Lateralized temporal fringes. J. Acoust. Soc. Am. 106, 3553-3563.

Grimault, N., Bacon, S.P., Micheyl, C. 2002. Auditory stream segregation on the basis of amplitude-modulation rate. J. Acoust. Soc. Am. 111, 1340-1348.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., Collet, L. 2000. Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. J. Acoust. Soc. Am. 108, 263-271.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., Collet, L. 2001. Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. Br. J. Audiol. 35, 173-182.

Hall, J.W., Peters, R.W. 1981. Pitch for nonsimultaneous successive harmonics in quiet and noise. J. Acoust. Soc. Am. 69, 509-513.

Hartmann, W.M., Johnson, D. 1991. Stream segregation and peripheral channeling. Music Perception 9, 155-184.

Hoekstra, A., Ritsma, R.J. 1977. Perceptive hearing loss and frequency selectivity. In: Evans, E.F., Wilson, J.P., (Eds.), Psychophysics and Physiology of Hearing. Academic, London, England. pp. 263-271.

Hong, R.S., Turner, C.W. 2006. Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. J. Acoust. Soc. Am. 120, 360-374.

Hong, R.S., Turner, C.W. 2009. Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. J. Acoust. Soc. Am. 126, 291-299.

Houtgast, T. 1976. Subharmonic pitches of a pure tone at low S/N ratio. J. Acoust. Soc. Am. 60, 405-409.

Houtsma, A.J.M., Smurzynski, J. 1990. Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304-310.

Jackson, H.M., Moore, B.C.J. 2013. The dominant region for the pitch of complex tones with low fundamental frequencies. J. Acoust. Soc. Am. 134, 1193-1204.

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

Lenth, R.V. 2016. Least-Squares Means: The R Package lsmeans. J. Stat. Softw., 1-33.

Mackersie, C.L., Prida, T.L., Stiles, D. 2001. The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. J. Speech. Lang. Hear. Res. 44, 19-28.

Madsen, S.M.K., Dau, T., Moore, B.C.J. 2015. Effect of harmonic rank on the streaming of complex tones. In: Santurette S., D.T., Christensen-Dalsgaard J., Tranebjærg L., Andersen T, (Ed.), Proceedings of the International Symposium on Auditory and Audiological Research: Individual Hearing Loss - Characterization, modelling, compensation strategies, Vol. 5.

Madsen, S.M.K., Whiteford, K.L., Oxenham, A.J. 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. Sci. Rep. 7.

Micheyl, C., Carlyon, R.P. 1998. Effects of temporal fringes on fundamental-frequency discrimination. J. Acoust. Soc. Am. 104, 3006-3018.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J. 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219, 36-47.

Moore, B.C.J. 1977. Effects of relative phase of the components on the pitch of three-component complex tones. In: Evans, E.F., Wilson, J.P., (Eds.), Psychophysics and Physiology of Hearing. Academic Press, London. pp. 349-358.

Moore, B.C.J. 2003. An Introduction to the Psychology of Hearing, 5th Ed. Emerald, Bingley, UK.

Moore, B.C.J., Gockel, H. 2002. Factors influencing sequential stream segregation. Acta Acust. United Acust. 88, 320-333.

Moore, B.C.J., Moore, G.A. 2003. Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. Hear. Res. 182, 153-163.

Moore, B.C.J., Gockel, H. 2011. Resolvability of components in complex tones and implications for theories of pitch perception. Hear. Res. 276, 88-97.

Moore, B.C.J., Huss, M., Vickers, D.A., Glasberg, B.R., Alcántara, J.I. 2000. A test for the diagnosis of dead regions in the cochlea. Br. J. Audiol. 34, 205-224.

Oxenham, A.J., Micheyl, C., Keebler, M.V. 2009. Can temporal fine structure represent the fundamental frequency of unresolved harmonics? J. Acoust. Soc. Am. 125, 2189-2199.

Paredes-Gallardo, A., Madsen, S.M.K., Dau, T., Marozeau, J. 2018. The role of temporal cues in voluntary stream segregation for cochlear implant users. Trends. Hear. 22.

Roberts, B., Glasberg, B.R., Moore, B.C.J. 2002. Primitive stream segregation of tone sequences without differences in F0 or passband. J. Acoust. Soc. Am. 112, 2074-2085.

Rose, M.M., Moore, B.C.J. 1997. Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 102, 1768-1778.

Ruggles, D.R., Freyman, R.L., Oxenham, A.J. 2014. Influence of musical training on understanding voiced and whispered speech in noise. PLoS One 9, e86980

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A. 2005. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat. Neurosci. 8, 1241-1247.

Shackleton, T.M., Carlyon, R.P. 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J. Acoust. Soc. Am. 95, 3529-3540.

Shearer, D.E., Molis, M.R., Bennett, K.O., Leek, M.R. 2018. Auditory stream segregation of iterated rippled noises by normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 143, 378-387.

Stainsby, T.H., Moore, B.C.J., Medland, P.J., Glasberg, B.R. 2004. Sequential streaming and effective level differences due to phase-spectrum manipulations. J. Acoust. Soc. Am. 115, 1665-1673.

van Noorden, L.P.A.S. 1975. Temporal coherence in the perception of tone sequences. Ph.D. Thesis, Eindhoven University of Technology.

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

Vliegen, J., Moore, B.C.J., 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.

Wang, J., Baer, T., Glasberg, B.R., Stone, M.A., Ye, D.T., Moore, B.C.J. 2012. Pitch perception of concurrent harmonic tones with overlapping spectra. J. Acoust. Soc. Am. 132, 339-356.

Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J. 2011. The Aligned Rank Transform for nonparametric factorial analysis using only ANOVA procedures, Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). NewYork: ACM Press, Vancouver, British Columbia. pp. 143-146.

Figure captions:

**Fig. 1.** Schematic illustration of the stimuli. A) Illustration of the ABA-ABA sequences used. A and B tones with a small difference in F0 (ΔF0) are likely to be perceived as one stream (integration; upper panel) whereas A and B tones with large ΔF0 are likely to be perceived as two streams (segregation; lower panel). B) Schematic spectra of complex tones. Tones were initially generated with many harmonics, and were bandpass filtered between 2 and 4 kHz. The harmonic rank was varied by varying the F0. Examples are shown of tones with low harmonics (F0 = 500 Hz, left) and high harmonics (F0 = 150 Hz, right).

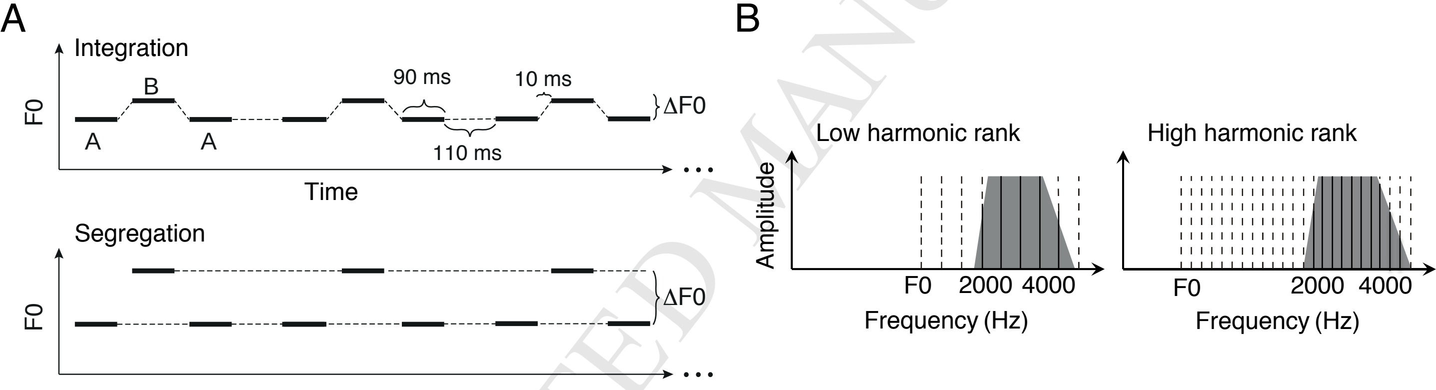
**Fig. 2**. Percentage of time the sequences were indicated as being perceived as two streams for each listener. The main boxes show the percentages for the complex tones, plotted as a function of the F0 of Tone A, with ΔF0 as parameter. The smaller panels to the right show results obtained with pure tones. Different symbols refer to different frequency differences or F0 differences between the A and B tones. Error bars indicate ±1 SE across trials.

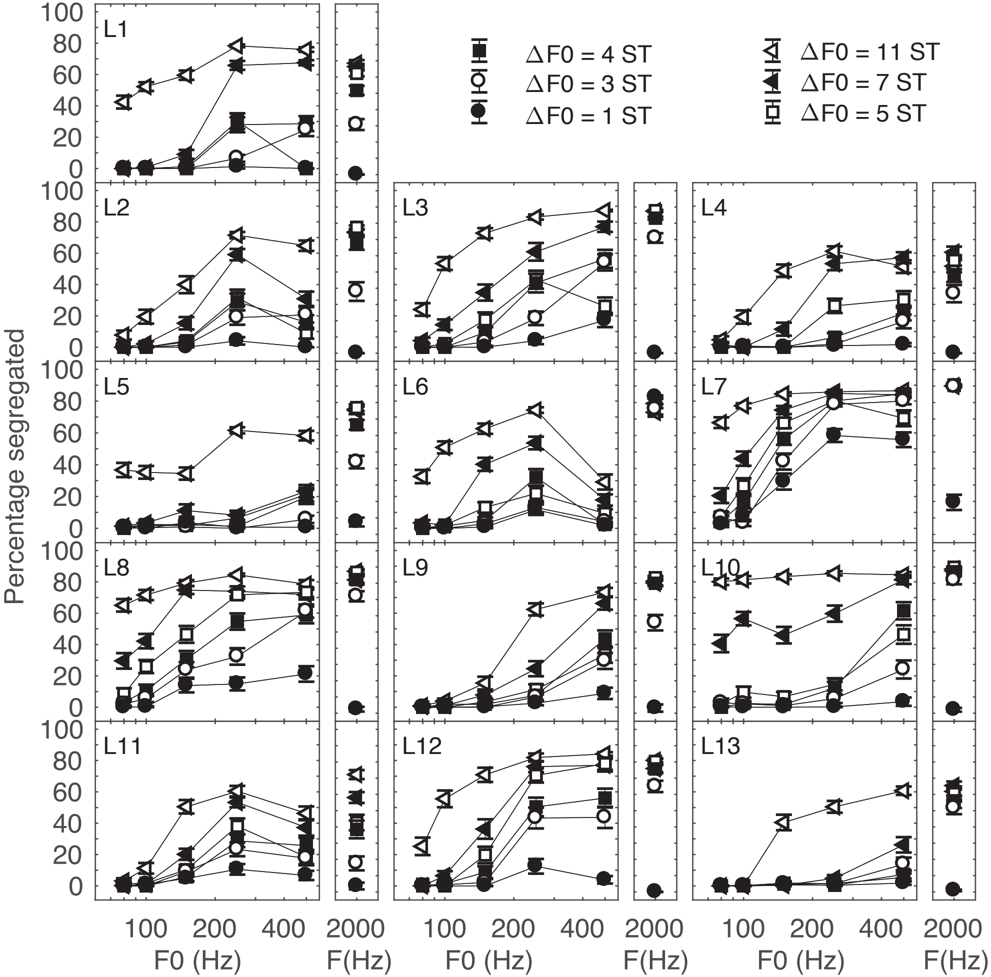
**Fig. 3.** As Fig. 2, but showing the mean across listeners. Error bars indicate ±1 standard error.

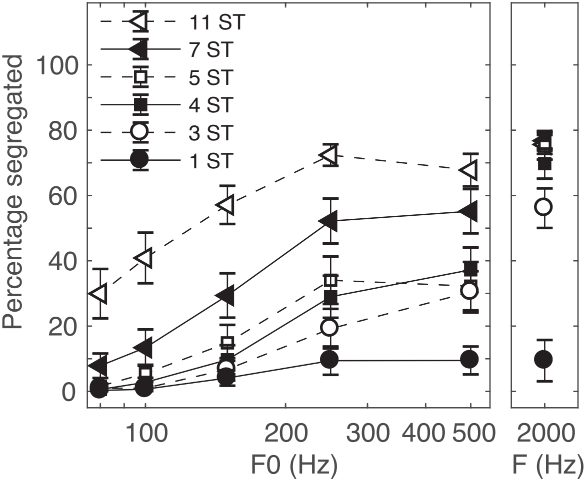
**Fig. 4.** Percentage of trials indicated as being heard as two streams after a given number of seconds for the condition with ΔF0 = 7 and F0 = 250 Hz. Each line shows results for one listener.

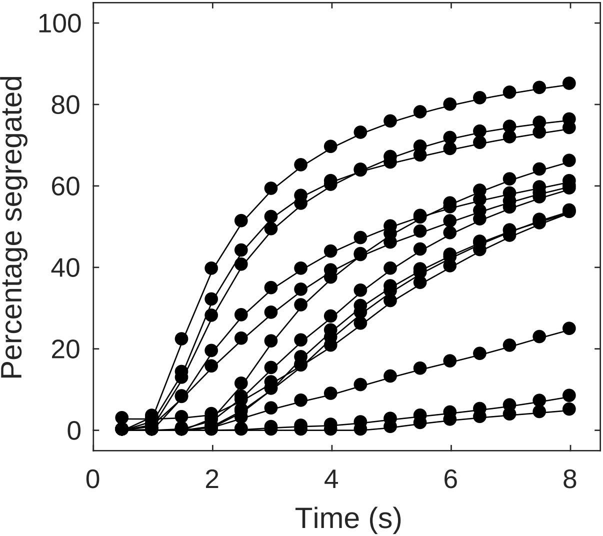
**Fig. 5.** Frequency difference limens for pure tones (right) and F0DLs for complex tones added in sine phase (filled diamonds) and in random phase (open circles) (left), plotted as a function of the reference F0. Error bars indicate ±1 SE.

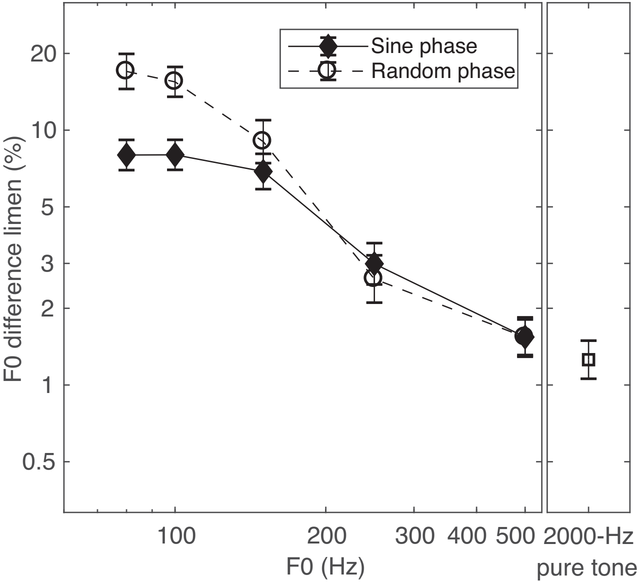
**Fig. 6.** Scatter plots showing the relation between stream segregation and F0DLs. Left panel: the mean percentage segregation for each A-tone F0 (averaged across ΔF0s and across listeners) is plotted against the mean F0DL (across listeners) obtained for the same F0. Right panel: the mean percentage segregation score for each listener (averaged across all conditions) is plotted against the mean F0DL for that listener (averaged across all F0s).

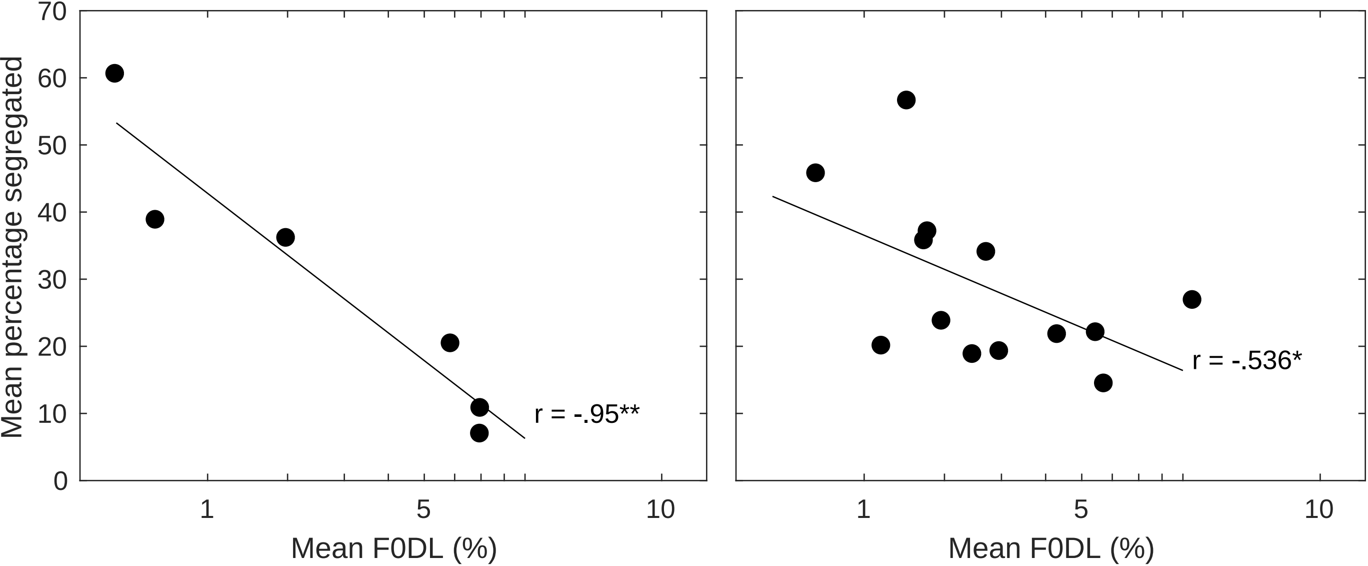

A
Integration
F0
B
A
A
90 ms
10 ms
110 ms
ΔF0
Time
Segregation
F0
ΔF0
B
Low harmonic rank
Amplitude
F0
2000
4000
Frequency (Hz)
High harmonic rank
F0
2000
4000
Frequency (Hz)


Percentage segregated
0
20
40
60
80
100
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
L13
ΔF0 = 4 ST
ΔF0 = 3 ST
ΔF0 = 1 ST
ΔF0 = 11 ST
ΔF0 = 7 ST
ΔF0 = 5 ST
100
200
400
F0 (Hz)
2000
F(Hz)


11 ST
7 ST
5 ST
4 ST
3 ST
1 ST
Percentage segregated
0
20
40
60
80
100
100
200
300
400
500
F0 (Hz)
2000
F (Hz)


Percentage segregated
100
80
60
40
20
0
Time (s)
0
2
4
6
8


Sine phase
Random phase
F0 difference limen (%)
20
10
5
3
2
1
0.5
100
200
300
400
500
F0 (Hz)
2000-Hz
pure tone


70
60
50
40
30
20
10
0
Mean percentage segregated
r = -.95**
1
5
10
Mean F0DL (%)
r = -.536*
1
5
10
Mean F0DL (%)

- Sequential stream segregation and fundamental frequency difference limens (F0DLs) were measured for pure tones and complex tones with varying harmonic rank
- Segregation increased with decreasing harmonic ranks
- The effect of harmonic rank was continuous and progressive and even small differences in harmonic rank led to differences in segregation.
- Correlations were found between stream segregation and F0DLs supporting the idea that the extent of stream segregation depends on the salience of the perceptual difference between sounds