Perceptual organization and stability of auditory streaming for pure tones and /ba/ stimuli

Samantha J Gustafson

Department of Communication Sciences and Disorders, University of Utah, 390 South 1530
East, Salt Lake City, Utah 84112
Samantha.gustafson@utah.edu

John Grose & Emily Buss

Department of Otolaryngology-Head and Neck Surgery, University of North Carolina, 170
Manning Drive, Chapel Hill, North Carolina 27599
john_grose@med.unc.edu
emily_buss@med.unc.edu

[1]

Abstract

The dynamics of auditory stream segregation were evaluated using repeating triplets composed of pure tones or the syllable /ba/. Stimuli differed in frequency (tones) or fundamental frequency (speech) by 4, 6, 8, or 10 semitones, and the standard frequency was either 250 Hz (tones and speech) or 400 Hz (tones). Twenty normal-hearing adults participated. For both tones and speech, a two-stream percept became more likely as frequency separation increased. Perceptual organization for speech tended to be more integrated and less stable compared to tones. Results suggest that prior data patterns observed with tones in this paradigm may generalize to speech stimuli.

Keywords: auditory streaming, speech, perceptual bistability

1. Introduction

Successful communication in competitive listening environments, such as a crowded restaurant, requires the use of auditory scene analysis to disambiguate the combined auditory information. This analysis allows the listener to group together successive sounds that share acoustic properties to form auditory objects or 'streams' (Bregman, 1990). In ambiguous listening conditions, where sounds from different sources are perceptually similar, adults experience a delay in the formation of a segregated percept, as well as instability of that percept, when compared to listening conditions where sounds are perceptually distinct (Deike et al., 2015; Thompson et al., 2011). The prolonged build-up of auditory stream segregation is known to affect speech recognition in adults with normal hearing and hearing loss (Best et al., 2008, 2018), but little is known about auditory scene stability as it relates to speech perception. The combination of perceptual build-up of stream segregation and auditory scene stability might prove useful in characterizing challenges experienced by populations with known difficulties understanding speech in complex acoustic environments (e.g., children, older listeners, and listeners with hearing loss).

It has long been argued that, at the initial encounter with the auditory environment, the auditory scene is perceived as a single, albeit complex, auditory object (Anstis & Saida, 1985; Bregman, 1978; Micheyl et al., 2005). Over time, the listener is thought to collect evidence about statistical regularities of the auditory scene (Bregman, 1978) and adapt to the stimuli present in the combined signal (Anstis & Saida, 1985; Micheyl et al., 2005), such that segregation of the component streams builds up over time. The time required to segregate streams within an auditory scene is faster for sounds that are perceptually distinct (e.g., large frequency differences) than for those that are perceptually similar (Anstis & Saida, 1985; Carlyon et al., 2001; Cusack et al., 2004; van Noorden, 1975). This process is often studied using a paradigm of repeating tone triplets, where a pair of tones (A and B) is presented in the sequence ABA_ABA_ABA_, etc. Listeners are asked whether they hear a single stream with a galloping rhythm or two separate streams. The classic result is that stimuli are heard as one stream if A and B are close in frequency, but the galloping percept is lost and the tones separate into two streams when the frequencies of A and B are more disparate (Anstis & Saida, 1985; van Noorden, 1975). Therefore, the dominant perception (one vs two streams) is systematically related to stimulus features (Carlyon et al., 2001; Cusack et al., 2004).

Contrary to the long-standing view that stream segregation builds up over time, Deike and colleagues (2012) found no evidence of a single-object initial perception. Instead, they argue that perceptual build-up to segregation only occurs for ambiguous stimuli. When presented with an ambiguous auditory scene, listeners often report changes in their percept over time even when stimulus parameters are stable (Deike et al., 2015; Denham et al., 2012). These seemingly unprovoked changes in scene analysis have been described extensively for visual perception (see Alais & Blake, 2005) but are less well understood for the alternating, or bi-stable, auditory scene (Hupé & Pressnitzer, 2012; Pressnitzer & Hupé, 2006; Rankin et al., 2015). Compared to stable percepts, bistability has been characterized in terms of a more prolonged delay to initial perceptual decision (initial decision time), shorter duration of this first percept (first-percept inertia), and more frequent switches between the two visual/auditory images/streams (switch-rate; Deike et al., 2015; Hupé & Rubin, 2003; Pressnitzer & Hupé, 2006). Similar to the dominant one- vs two-stream perception discussed above, the listener’s initial perception (first-percept bias) is related to stimulus characteristics such as frequency separation. That is, larger frequency separations between streams are associated with a bias toward an initial two-stream percept (Deike et al., 2012).

One challenge in generalizing published results from psychophysical studies of auditory streaming (e.g., the ABA_ paradigm) to real-world listening is that psychophysical studies have typically used pure tones (e.g., Cusack et al., 2004) or harmonic tone complexes (e.g., Deike et al., 2012) rather than natural speech. Some previous work has evaluated the effect of fundamental frequency (F0) on streaming of vowel and consonant-vowel stimuli (e.g., David et al., 2017; Gaudrain et al., 2012). Results of these two studies are generally consistent with the idea that differences in F0 support segregation; however, these studies used several speech tokens in contrast to the repeating stimuli of the classic, tonal ABA_ paradigm. Furthermore, tasks used in this previous work relied on cues that were either distributed across streams or contained within a single stream, promoting either intentional integration or segregation of auditory streams, respectively. Given the differences between prior studies of streaming using tones vs. speech, it is unclear the extent to which auditory streaming of tones in the ABA_ paradigm generalize to speech.

Another difference between previous research on streaming for pure tones and for speech is that tonal stimuli used in auditory streaming paradigms are typically higher in frequency than the F0s of speech (Hillenbrand et al., 1995). Therefore, one objective of the current research was to determine whether auditory streaming in the ABA_ paradigm depends on the standard frequency (i.e., Tone B). The present study compared results for pure tones with a 400-Hz standard, as used previously by others (Carlyon et al., 2001; Cusack et al., 2004), to results obtained with a 250-Hz standard, which corresponds to a typical F0 for a female talker (Hillenbrand et al., 1995). Although it is unclear whether frequency affects streaming in undirected listening, previous results indicate no effect of standard frequency when participants are asked to maintain auditory stream segregation (Rose & Moore, 2000). Based on this result, we predicted no effect of standard frequency on pure-tone streaming with undirected listening.

Another objective of this research was to directly compare auditory streaming patterns obtained using pure-tone and speech tokens. We are unaware of any study that has directly compared auditory streaming in the repeating triplet ABA_ paradigm using these different stimulus types. To that end, we compared auditory streaming performance for the 250-Hz standard tone and a recording of /ba/ produced by a female talker with a standard F0 of 250 Hz. Because similarity across multiple spectral and temporal features is known to reduce segregation (Cusack & Roberts, 2004; Moore & Gockel, 2012), we hypothesized that the effect of frequency separation on auditory stream segregation would be smaller for speech stimuli when compared to pure tone stimuli. Derived variables used to better understand streaming over time for these two stimuli included build-up of perceptual stream segregation (i.e., probability of segregated percept over time), and measures of auditory scene stability (i.e., initial decision time, first-percept bias, first-percept inertia, and switch-rate).

2. Methods

Listeners were young adults (18-31 yrs, mean = 24.8 yrs, n=20) with pure-tone thresholds 20 dB HL or less from 250-8000 Hz, bilaterally. Participants reported no formal musical training and all but two were native speakers of English.

Tones or speech syllables were presented using an ABA_ pattern. The standard B stimulus was either a 400-Hz tone, a 250-Hz tone, or the speech syllable /ba/ with an F0 of 250

Hz. Condition names reflect this B standard. Stimulus frequency was parametrically manipulated for A stimuli. Those frequencies were approximately 4, 6, 8, or 10 semitones (ST) above the B stimulus frequency (314, 358, 402, and 465 Hz for the 250-Hz standard; 502, 572, 643, and 744 Hz for the 400-Hz standard). The speech F0 was manipulated in Praat (Boersma & Weenink, 2016). Each tone burst or speech syllable was 120 ms in duration, with 20-ms raised-cosine attack and decay ramps. The silent epoch between ABA triplets, denoted as _, was also 120 ms. The ABA_ pattern was repeated 40 times to create an approximately 20-s sequence that was expected to be perceived as a single stream (galloping rhythm) or two separate streams presented at the same time. Examples of the 4-ST and 10-ST speech conditions are provided as supplementary matrial.[1] Effects of standard frequency on auditory streaming were evaluated by comparing results for the 400-Hz tone and 250-Hz tone conditions, and effects of stimulus type on auditory streaming were evaluated by comparing results for the 250-Hz tone and the speech conditions.

A custom Matlab (MathWorks) script was used to control this experiment – sequences were routed through a real-time processor (RP2, TDT) and presented to the left ear via a Sennheiser HD25 headphone at 65 dB SPL. Prior to each 20-s run, participants heard three stimuli, each representing streams that they might hear during the run. Each of these example stimuli were presented along with graphics illustrating either (1) a single stream, represented by a galloping horse (ABA_), or (2) two segregated streams, one represented by a frog (_B_ _), and the other represented by a pair of birds (A_A_).[2] To ensure that the left-to-right order of cartoon illustrations on the screen did not create a bias toward a specific perception, thirteen participants were tested with the horse on the left side of the screen, and seven were tested with the horse on the right. During testing, participants used the touch-screen interface to select whether they heard one stream (horse) or two streams (frog and birds), making additional selections whenever the percept changed. To characterize changes in perception over time, participant responses were sampled once every 0.48 s, at the end of each ABA_ sequence. All data were initially coded as 'undecided' prior to the participant's initial response. After the first response, the absence of a response at any given sample time point was taken as evidence of no change since the previous response. Percept reports (undecided, one stream, two streams) and the corresponding response times were saved for later analysis.

Participants completed 10 runs per frequency separation (4, 6, 8, or 10 ST between A and B) for each stimulus condition (400-Hz tone, 250-Hz tone, and speech with a 250-Hz F0). Data were collected in six blocks (two per stimulus condition); each block included five runs at each of the four frequency separations, all for the same stimulus condition. The order of blocks and the order of runs within a block were randomized. Testing was completed in a double-walled sound booth in one visit lasting approximately 1.5 hours. All procedures were approved by the Biomedical IRB at the University of North Carolina at Chapel Hill. Data were analyzed in the R programming environment (R Core Team, 2013). A generalized linear mixed-effects model (*lme4;* Bates et al., 2015) was used to evaluate changes in perception over time. All reported p-values are two-tailed. A bootstrapping data resampling procedure with replacement (n = 1000) was used to visualize the effects of frequency separation and stimulus condition on derived variables.

3. Results

Figure 1 shows the median probability of a two-stream response as a function of time for each frequency separation and stimulus condition. Shaded areas represent the 25th to the 75th percentiles, and the size of the symbols represent the number of responses obtained across participants at each point in time. Note that the symbol size varies the most over the first five seconds, representing the time span over which initial perceptions were reported. For direct comparison, the median values for all three stimuli are plotted together in the right-most column. This plot indicates a positive association between a two-stream percept and spectral separation of the A and B streams, and broadly similar results for the Tone 400 and Tone 250 stimuli. The two-stream percept at large spectral separations was less consistent for the Speech than the Tone 250 stimuli.

These observations were evaluated statistically using a generalized linear mixed-effects model with a binomial link function. Because two contrasts were of interest – Tone 250 vs Tone 400 and Tone 250 vs Speech – a single model was constructed with Tone 250 stimulus condition acting as the reference condition. The initial model included fixed effects of time point (in 0.48-s bins), frequency separation (in ST), stimulus condition, and all associated two- and three-way interactions; run number was also included as a fixed effect, and there was a random intercept for participant. The three-way interaction was not significant (p = 0.504), so that effect was dropped from the final model. Results of this analysis are reported in Table 1. All effects in this model were significant apart from the non-significant main effect of Tone 400 (p = 0.863) and non-significant interaction between time point and Tone 400 (p = 0.886). The only significant effect including Tone 400 stimuli was the interaction with frequency separation (p = 0.001); this reflects a trend for more two-stream reports with the Tone 400 than the Tone 250 stimulus at small frequency separations. In contrast, all three effects including the Speech stimulus were significant, including the main effect, the interaction with time point, and the interaction with frequency separation ($p \leq 0.007$). Compared to the Tone 250 stimulus, these effects reflect fewer two-stream reports for the Speech stimulus, particularly for the large frequency separations and at later time points. The significant effect of run number ($p < 0.001$) indicates a reduction in the number of two-stream percepts over the course of the experiment.

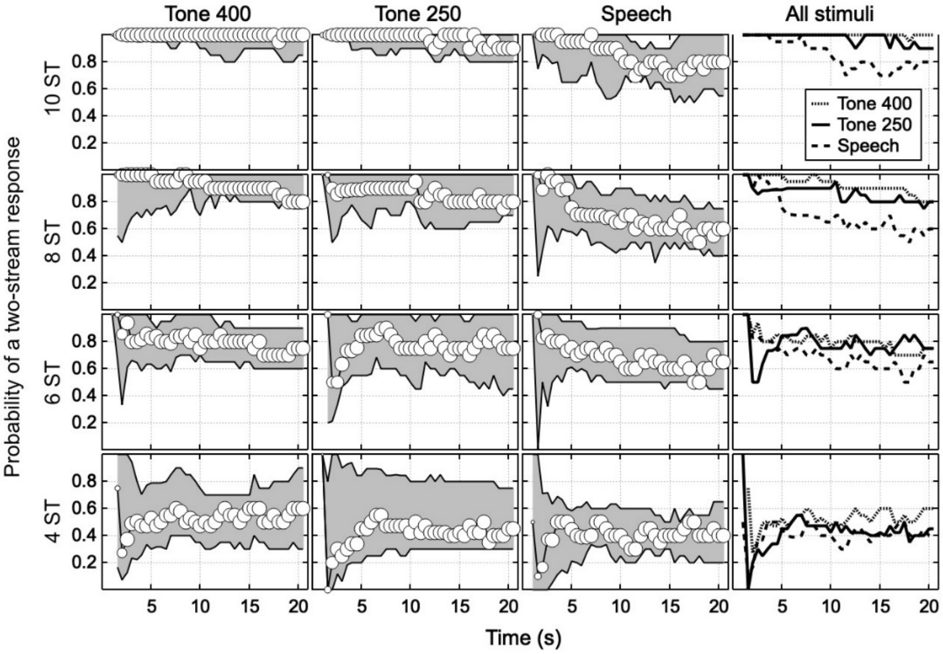

Figure 1. Median probability of a two-stream response as a function of time for each stimulus condition (Columns 1-3) and frequency separation condition (Rows 1-4). Shaded regions represent 25th and 75th quartiles. Symbol size is representative of the number of responses obtained across participants at each time point. Column 4 shows median probabilities overlaid for the three stimulus conditions.

Table 1. Output of a linear mixed-effects model with a binomial link function. The probability of a two-stream percept was the dependent variable. There was a random intercept for each subject.

	Estimate	Std. Error	*z* value	*p*
Intercept	-1.57	0.282	-5.58	<0.001
Time point	0.015	0.003	5.45	<0.001
Frequency separation	0.49	0.012	41.6	<0.001
Stimulus condition: Tone 400	0.014	0.083	0.17	0.863
Stimulus condition: Speech	0.50	0.078	6.38	<0.001
Run number	-0.021	0.003	-7.08	<0.001
Time point x Frequency separation	-0.004	<0.001	-9.47	<0.001
Time point x Tone 400	<0.001	0.002	-0.14	0.886
Time point x Speech	-0.005	0.002	-2.71	0.007
Frequency separation x Tone 400	0.036	0.011	3.33	0.001
Frequency separation x Speech	-0.15	0.010	-15.39	<0.001

Ninety-five percent confidence intervals derived using a bootstrap resampling procedure were computed for initial decision time, first-percept bias, first-percept inertia, and switch-rate. Figure 2 shows estimated median and 95% confidence intervals for the Tone 250 and Speech stimulus conditions at each frequency separation. In general, increases in frequency separation were associated with shorter initial decision time, increased bias for an initial two-stream percept, and longer first-percept inertia. Number of perceptual switches varied nonmonotonically across frequency separation, with higher switch-rates recorded for intermediate frequency separations. Comparisons of confidence intervals reveal that Speech stimuli elicited shorter initial percept durations and more perceptual switches than Tone 250 stimuli. First-percept bias toward reporting a two-stream initial percept was less likely for Speech stimuli than Tone 250 stimuli at larger frequency separations. Confidence intervals for these four measures of auditory scene stability overlapped at every frequency separation for the Tone 250 and Tone 400 stimulus conditions (not shown).

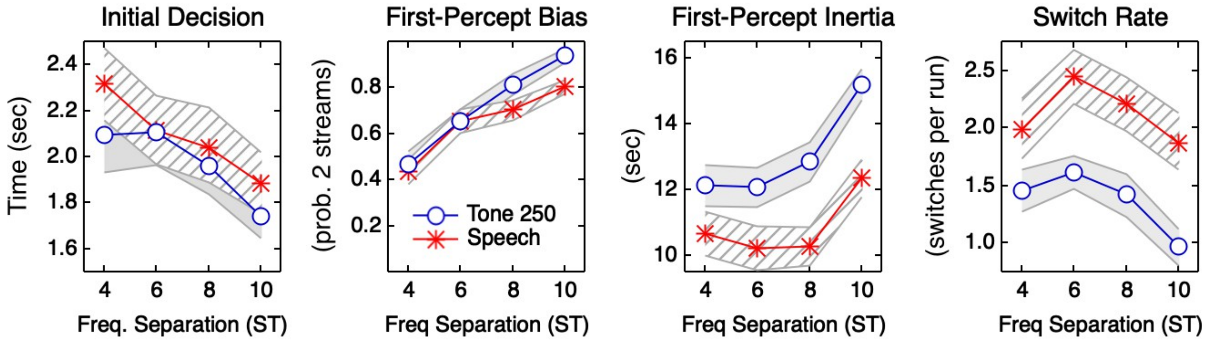

Figure 2. Median and 95% confidence intervals estimated based on bootstrapped resampling of initial decision time, first-percept bias, first-percept inertia, and switch-rate for each frequency separation in the Tone 250 and Speech stimulus conditions. (Color online)

4. Discussion

The purpose of this study was to evaluate the effect of standard frequency for pure-tone stimuli (400 Hz vs 250 Hz) and stimulus type (pure tone vs speech syllable) on the dynamics of auditory stream segregation. Within these two stimulus comparisons, we assessed the effect of stimulus frequency separation on traditional characterization of the build-up of perceptual stream segregation (i.e., probability of a two-stream percept over time) and on measures of auditory scene stability (i.e., initial decision time, first-percept bias, first-percept inertia, and switch-rate).

Auditory stream stability and perceptual change over time was largely consistent when the stimulus was a pure tone with a standard frequency of 400 or 250 Hz. For both standard frequencies, as frequency separation increased, the time to initial perceptual decision was reduced, the likelihood of a two-stream percept increased, the duration of the first percept lengthened, and fewer perceptual switches were reported. There was a modest difference between Tone 250 and Tone 400 stimuli with respect to the effect of frequency separation, although that could be a main effect “masked” by ceiling effects (i.e., segregation approaching 100% at 10 ST). Taken together, these results suggest similar patterns of auditory stream stability and perceptual change over time despite differences in standard frequency, consistent with previous data from a task that encouraged directed listening (Rose & Moore, 2000).

The tone and speech stimuli with a 250-Hz standard produced broadly similar results, but there were several notable differences. For both stimuli, increased frequency separation was associated with reductions in the time to an initial perceptual decision, increases in the likelihood of an initially segregated perceptual organization, and greater persistence of the initial percept. Systematic changes in the number of perceptual switches across frequency were observed for both stimulus conditions. Speech stimuli were more likely to be heard as one stream initially at large frequency separations, and auditory organization was less stable over time when compared to tone stimuli. Additionally, patterns of perception over time varied between speech and tone stimuli, as results show an increased tendency for speech stimuli to be integrated into a single stream over time when compared to tonal stimuli. This is particularly apparent in the 10, 8, and 6 ST conditions (see Figure 1).

Novel findings of the current study support the idea that some but not all aspects of the traditional psychophysical findings obtained with tones in the ABA_ paradigm generalize to speech stimuli. Together, these results suggest that streaming of speech stimuli may be more ambiguous than tone stimuli with comparable frequency separation. This increased ambiguity with speech stimuli is a potentially important area for future research, as previous studies have suggested that perceptual ambiguity measured using a streaming task might be associated with the higher task demands required to disambiguate the uncertain auditory scene (Deike et al., 2015; Dolležal et al., 2014).

Our finding that speech stimuli are more likely to be heard as one stream than tone stimuli could also be attributed to the particular speech stimuli that were used. Because both A and B speech tokens used here were produced by one talker, features of that talker's voice (e.g., timbre) could have introduced a bias for a one-stream percept. This would be consistent with results of Cusack and Roberts (2000), who found that differences in timbre between tokens facilitate stream segregation. Future research is needed using speech tokens recorded from different talkers to evaluate if the speech-specific ambiguity and tendency to integrate speech streams over time persists under multi-talker listening conditions.

Evaluation of stream segregation has historically assumed that the listener always begins with an integrated percept, and that the perception of segregation builds up over time (Anstis & Saida, 1985; Bregman, 1978; Micheyl et al., 2005). This assumption was challenged by Deike (2012) who showed that initial perceptions of one or two streams were dependent upon frequency separation. Our findings align with those of Deike (2012), in that there was a strong effect of frequency separation on first-percept bias and a tendency for the initially-segregated percepts to be integrated into a single stream over time. In the case of the initially-segregated percepts at large frequency separations, it is possible that participants in this study did, in fact, experience a brief, one-stream percept initially, but then experienced a switch to a segregated perceptual organization prior to making their initial response. If this were the case, earlier initial decision times might be expected when reporting a single-stream perception in comparison with those obtained for reports of a two-stream percept, as the listener would be primed to report a one-stream percept if a single-stream bias was present. Supplemental data analysis failed to find evidence for this pattern of results; instead, we found that initial decision times were fastest for single-stream percepts when the frequency separation was small (e.g., 4 ST) and for two-stream percepts when the frequency separation was large (e.g., 10 ST).[3]

5. Conclusion

This study represents an intermediate step between the classical paradigm of repeating triplets and previous work that used natural speech stimuli to improve our understanding of the perceptual organization and stability of the auditory scene. Here, we evaluated the effects of standard frequency and stimulus type on auditory stream segregation by comparing results for pure tones at either 400 Hz or 250 Hz, and by comparing results for a 250-Hz pure tone with speech stimuli having a 250-Hz F0, respectively. The effect of pure-tone standard frequency on measures of auditory scene stability and perceptual change over time was negligible. The effect of frequency separation was generally similar for tonal and speech stimuli, but there was tendency for a more integrated and less-stable perceptual organization for speech stimuli. Additional research is needed to determine if aspects of perceptual organization evaluated with this paradigm are related to a listener’s ability to understand speech in complex, multi-source listening environments.

Acknowledgements

We thank Katheryn Sobon for her help with recruitment and data collection. This work was funded by NIH NIDCD R01 DC014460. Statistical help was supported by the National Center for Advancing Translational Sciences of the NIH under Award Number UL1TR002538.

Textual footnotes

1. See supplementary material at [URL will be inserted by AIP] for example stimuli and spectrograms illustrating the 4-ST and 10-ST speech conditions.
2. This graphical interface was designed for use with children in future studies.
3. See supplementary material at [URL will be inserted by AIP] for details of this analysis.

References and links

Alais, D., & Blake, R. (2005). *Binocular Rivalry*. MIT Press.

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(3), 257.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4.' *Convergence*, *12*(1), 2.

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*(35), 13174–13178.

Best, V., Swaminathan, J., Kopčo, N., Roverud, E., & Shinn-Cunningham, B. (2018). A "Buildup" of Speech Intelligibility in Listeners With Normal Hearing and Hearing Loss. *Trends in Hearing*, *22*, 2331216518807519.

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer [Computer program]. Version 6.0*. retrieved from http://www.praat.org/

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 380.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115.

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 643.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112–1120. https://doi.org/10.3758/BF03212092

Cusack, R., & Roberts, B. (2004). Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. *Hearing Research*, *193*(1), 95–104. https://doi.org/10.1016/j.heares.2004.03.009

David, M., Lavandier, M., Grimault, N., & Oxenham, A. J. (2017). Sequential stream segregation of voiced and unvoiced speech sounds based on fundamental frequency. *Hearing Research*, *344*, 235–243.

Deike, S., Heil, P., Böckmann-Barthel, M., & Brechmann, A. (2012). The build-up of auditory stream segregation: A different perspective. *Frontiers in Psychology*, *3*, 461.

Deike, S., Heil, P., Böckmann-Barthel, M., & Brechmann, A. (2015). Decision making and ambiguity in auditory stream segregation. *Frontiers in Neuroscience*, *9*, 266.

Denham, S. L., Bendixen, A., Mill, R., Tóth, D., Wennekers, T., Coath, M., B\Hohm, T., Szalardy, O., & Winkler, I. (2012). Characterising switching behaviour in perceptual multi-stability. *Journal of Neuroscience Methods*, *210*(1), 79–92.

Dolležal, L.-V., Brechmann, A., Klump, G. M., & Deike, S. (2014). Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00119

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J.-C. (2012). The relationship between concurrent speech segregation, pitch-based streaming of vowel sequences, and frequency selectivity. *Acta Acustica United with Acustica*, *98*(2), 317–327.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Hupé, J.-M., & Pressnitzer, D. (2012). The initial phase of auditory and visual scene analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 942–953.

Hupé, J.-M., & Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: A fresh look at plaids. *Vision Research*, *43*(5), 531–548.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. *Neuron*, *48*(1), 139–148. https://doi.org/10.1016/j.neuron.2005.08.039

Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 919–931.

Pressnitzer, D., & Hupé, J.-M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology*, *16*(13), 1351–1357.

R Core Team. (2013). *R: A language and environment for statistical computing*.

Rankin, J., Sussman, E., & Rinzel, J. (2015). Neuromechanistic Model of Auditory Bistability. *PLOS Computational Biology*, *11*(11), e1004555. https://doi.org/10.1371/journal.pcbi.1004555

Rose, M. M., & Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, *108*(3), 1209–1214.

Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1253.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Vol. 3). Institute for Perceptual Research Eindhoven, The Netherlands.

Supplementary Material.

Initial decision times for one- and two-stream percepts as a function of frequency separation

To explore the possibility of a brief, initial one-stream percept that preceded the initial decision, we used a multilevel mixed-effects linear regression model to quantitatively evaluate the effect of first-percept bias on initial decision time. Table 1 shows the model summary for the linear mixed-effects model with initial decision time as the dependent variable. Fixed effects included first-percept bias, frequency separation, stimulus condition, run number, and the interaction between first-percept bias and frequency separation. Results of this model indicate significant main effects of first-percept bias, frequency separation, and run number. The significant effect of run number indicates a reduction in initial decision times over the course of the experiment. A significant interaction was found between first-percept bias and frequency separation. Table 2 illustrates the nature of this interaction. Initial decision times were faster for one-stream percepts when the frequency separation was small (e.g., 4 ST) and for two-stream percepts when the frequency separation was large (e.g., 10 ST).

Table 1. Output of a linear mixed-effects model with initial decision time as the dependent variable. There was a random intercept for each subject.

	Estimate	Std. Error	*t* value	*p*
Intercept	1.72	0.34	5.14	<0.001
First-percept bias	0.54	0.17	3.16	0.002
Frequency separation	0.11	0.047	2.43	0.015
Stimulus condition: Tone 400	0.041	0.057	0.71	0.48
Stimulus condition: Speech	0.090	0.058	1.57	0.12
Run number	-0.035	0.008	-4.27	<0.001
First-percept bias x Frequency separation	-0.10	0.026	-3.96	<0.001

Table 2. Marginal effects for initial decision times in sec by first-percept bias and frequency separation. The standard error of each estimate is reported in parentheses. These effects are evaluated for Tone-250 and the mean value of run (5.5).

	First-percept bias	
Frequency separation	One-stream	Two-stream
4 ST	2.11	2.25
	(0.17)	(0.17)
6 ST	2.14	2.06
	(0.17)	(0.16)
8 ST	2.16	1.88
	(0.17)	(0.16)
10 ST	2.18	1.69
	(0.19)	(0.16)