Toward an objective measure for a "stream segregation" task

Virginia M. Richards, Eva Maria Carreira, and Yi Shen
Department of Cognitive Sciences, University of California, Irvine, 3151 Social Science Plaza, Irvine, California 92697-5100
v.m.richards@uci.edu, ecarreir@uci.edu, shen.yi@uci.edu

Abstract: A procedure to estimate the relative contribution of "A" and "B" tones for a stream-segregation task is described. Listeners detected a delay in the penultimate A tone in an A-B-A-B sequence of tones. For small A-B frequency separations, for most listeners, classification models based on both the A and B tones were superior to models based on just the A tones. For large frequency separations, models based on just the A tones were superior, indicating the A and B tones were segregated. The results also revealed individual differences in the strategies adopted to complete the task.

PACS numbers: 43.66.Lj, 43.66.Mk [QJF]
Date Received: September 19, 2011 **Date Accepted:** October 28, 2011

1. Introduction

The stream segregation paradigm (e.g., Bregman and Campbell, 1971; van Noorden, 1977) has emerged as a significant paradigm to study the perceptual organization of sound sources that unfold sequentially (for recent reviews, see Moore and Gockel, 2002; Snyder and Alain, 2007). In a generic experiment, the stimulus is composed of two alternating sounds, A and B. A temporal sequence ABABABA… may be perceived as forming a single "galloping" stream, ABABA…, forming two distinct streams, AAA… and BBB…, or drifting between the two organizations. When the A and B sounds are tones, the frequency separation between the tones, the time between tone onsets, as well as the listener's motivations contribute to whether one or two streams are perceived. It is of interest, therefore, to provide a measure of whether listeners rely on just the A tones, the B tones, or both the A and B tones, when completing a stream-segregation task. Such a measure is proposed here, and an initial data set is analyzed to demonstrate the proposed method. Additionally, potential limitations of the measure are noted.

Recently, Micheyl and Oxenham (2010) proposed an experimental paradigm to objectively assess stream segregation using behavioral methods. The temporal delay imposed on one of the A tones in an …ABA_ABA… sequence was detected. The detectability of changes in rhythm or tempo of one or more tones has been studied in the past (e.g., Vliegen *et al.*, 1999; Cusack and Roberts, 2000), reflecting a report by Bregman and Campbell (1971), who observed listeners did not make temporal judgments across tonal streams, only within streams. Micheyl and Oxenham strategically applied temporal variation in the timing of the tones (on the order of tens and hundreds of milliseconds) so that in separate conditions, performance would be facilitated or degraded by the perceptual integration/segregation of the A and B tones (Jones *et al.*, 2002). They showed that under their manipulations the thresholds co-varied with listeners' reports as to whether they heard one or two streams.

In the current experiment a sequence of …ABAB… tones were played, and the task was to detect whether the penultimate A tone was delayed. Unlike the Micheyl and Oxenham approach, however, stimulus manipulations were not applied to encourage listeners to adopt one strategy or another. Rather, the goal was to *unveil* the

strategies adopted by listeners when the stimuli and instructions were relatively neutral. This was achieved by estimating relative decision weights for the onsets and offsets of the A and B tones of the sequence, an approach similar to that described by Lutfi and Liu (2011) to estimate the segregation of concurrent sounds. If the listener relied on the timing of the A tones and not the segregated B tones, then the relative weights would be concentrated within the A stream containing the delayed signal. On the other hand, if the listener relied on both the A and B tones, the relative weights would be spread across both the A and B tones. To distinguish between these alternatives, a statistical criterion was adopted to choose between a reduced model (reliance on just the A tones) or a full model (reliance on both the A and B tones).

2. Methods

2.1 Psychophysical methods

Each stimulus was composed of 11 tones, six A tones and five B tones, as shown in Fig. 1. The frequency of the A tones was fixed at 1000 Hz, and the frequency of the B tones was changed across conditions: The frequencies of the B tones were 2, 17, 25, or 32 semitones higher than the frequency of the A tones (Δf), although for two listeners a finer gradation was tested. Each tone was 90 ms in duration, and the time between sequential tones was 60 ms. The first two A tones were presented without an intervening B tone to encourage the perception of segregated streams (Vliegen *et al.*, 1999). The signal to be detected was a fixed delay in the penultimate A tone and a yes/no procedure was used.

To obtain estimates of the relative weights, random variation was introduced along the same dimension as the detection task by applying temporal perturbations to the times of the onsets and offsets of the final three A tones and the final four B tones. The perturbations were independently drawn from a normal distribution with a standard deviation of 5 ms and were coded as positive or negative depending on whether the onset or offset was later or earlier in time relative to the nominal onsets and offsets.

The stimuli were digitally generated using a sampling frequency of 48 000 Hz on a PC, which also controlled the experimental procedure and data collection through custom-written software in MATLAB (The MathWorks). The stimuli were presented diotically to the listeners via a 24-bit soundcard (Envy24 PCI audio controller, VIA Technologies), a programmable attenuator and headphone buffer (PA4 and HB6, Tucker-Davis Technologies) and a Sennheiser HD410 SL headset. Each stimulus presentation was followed by visual feedback as to the correct response. The experiment was conducted in a double-walled, sound-attenuated booth.

Six listeners participated, three male and three female, whose age ranged from 19 to 29 years. Two of the listeners, L5 and L6, were highly experienced musical performers (e.g., minimally equivalent to an undergraduate degree in performance), one of whom (L6) was a percussionist. Five of the six listeners had absolute thresholds of

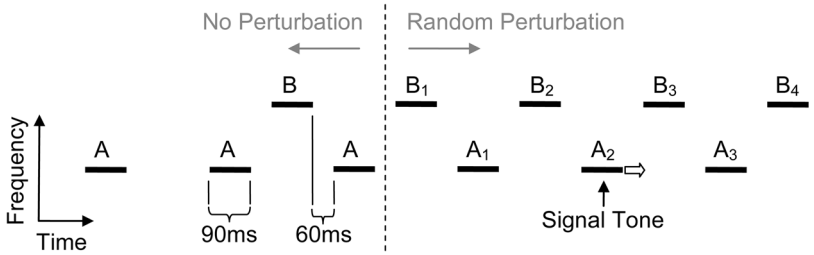

Fig. 1. The stimuli were composed of six A tones and five B tones. The onsets and offsets of the tones plotted to the right of the dashed line were perturbed; those plotted to the left were unaltered. The signal tone, A_2, was delayed (arrow; signal) or not.

20 dB HL or better between 250 and 8000 Hz. Listener 5 had absolute thresholds of 30–35 dB HL above 4000 Hz in the left ear.

Listeners became acquainted with the task by adjusting the frequencies of the B tones for stimuli similar to those used in the experiment and practicing the task using several values of Δf. For data collection, different values of Δf were tested in random order across listeners. Before data collection began for any one value of Δf, the signal delay was adjusted for each listener to maintain approximately 75% correct responses. Typically 5 or 10 30-trial blocks were required to choose the appropriate signal delay and ensure stable performance. Once the signal delay to be tested was established, data were collected in sets of five 30-trial blocks with breaks between sets. In nearly all instances 750 trials were analyzed for each listener and condition. For L1–L3 and L6, after completing the experiment, an additional data set (five blocks of 30 trials; two sets for L1) was collected *without* perturbation unless performance levels reached 100% correct (L2). For L4–L5, these no-perturbation sets were interspersed with data collection with perturbation.

2.2 Analysis

The data were analyzed using two logistic regressions (*glmfit,* MATLAB). The binary dependent variables were listeners' responses and the independent variables were the perturbation values. For the *full*, or A&B model, which assumed that information in the starting and ending times of A and B tones were incorporated into a listeners' decisions, a total of 14 coefficients, or relative weights, plus a bias term, were estimated, six associated with the last three A tones and eight associated with the last four B tones. For the *restricted*, or A model, which assumed that only information in the starting and ending times of the A tones were incorporated into the listeners' decisions, a total of six relative weights (plus a bias term) were estimated. Perturbation values from both signal and no-signal trials were used in the analysis, but the signal delay was not included. The resulting relative weights and associated statistical analyses were used in three separate ways for all the conditions and listeners.

First, the deviance values for the full and restricted models were differenced to form a χ^2 statistic with 8 DOF (the change in the number of degrees of freedom, full minus restricted models). If the two models differed at a level of $P < 0.05$ or less (a criterion χ^2 value of 15.5), it was taken as support for the full model; otherwise the restricted model was favored. Second, for whichever model was best supported by the initial χ^2 test, the Hosmer Lemeshow χ^2, (χ^2_{HL}), was used to indicate the quality of the fit to the model (Dobson, 2002). Data were separated into 10 categories based on the predicted probabilities of indicating yes (and the mirror probability of indicating no), and the summed values treated as frequencies. Next, χ^2_{HL} was estimated with the number of categories minus 2, or $10 - 2 = 8$, DOF. If this value exceeded the criterion value of 15.5, it was taken as evidence that the fit was poor; i.e., the data and the predictions were discernibly different. In a third analysis, the relative weights were examined. Relative weights with values more than two standard errors of the estimate removed from zero were deemed significant and ultimately reported.

3. Results and discussion

For each listener, the magnitude of the signal delays ultimately tested were largely independent of the values of Δf tested (the variation observed was not consistent across listeners). For L1–L6, the averaged delays tested were 27.5, 26.0, 24.4, 16.3, 11.0, and 14.4 ms, respectively. Note that smaller delays were tested for the musicians L5 and L6. On average, the resulting index of sensitivity (d') was 1.2 units (standard deviation = 0.2). For L2, sensitivity was far superior without perturbations than with perturbations; her performance was at or near 100% correct without perturbations. For L1, there was no dependence of sensitivity on whether or not perturbations were applied. For the remaining listeners, L3–L6, removing perturbations led to values of d' approximately 0.75 units higher than when perturbations were present. Thus, for all

but L2, it seems reasonable to suggest that the addition of the perturbations did not greatly alter listeners' decision processes. In addition to estimating values of d', the decision criteria, c, were also estimated. Four of the six listeners (L1–L3; L6) were strongly biased to indicate the signal was not present. Across values of Δf, when perturbations were present, the averaged values of c were 0.48, 0.21, 0.32, 0.05, 0.09, and 0.25 for L1–L6, respectively. Without perturbations, the trends were similar, except that estimates were not available for L2.

Table 1 summarizes the primary results regarding whether listeners relied on both the A and B tones or just the A tones as a function of Δf. For each listener, two columns are presented. The first column indicates the preferred model and the difference in deviance, full versus restricted model, is indicated in parenthesis. The second column indicates the quality of fit for the preferred model as χ^2_{HL}. For four of the listeners, as the frequency separation between the A and B tones increased, the full model gave way to the restricted model. For L3 and L4, however, there was no statistical evidence that the B tones were incorporated into their decisions even when the A and B tones were separated by two semitones (1000 and 1221 Hz, respectively). With regard to the two musicians (L5 and L6), these summary data show that they integrated information from the B tones into their decision for wider frequency separations than did the other listeners. The regression model was reasonably successful; the differences between the predicted and expected values were statistically different (asterisks in Table 1) for only three of the 26 fits, although for an additional three fits the value of χ^2_{HL} was near the $p = 0.05$ criterion (i.e., greater than 14). Overall, the current method appeared to provide a reasonable means of discovering which tones, A and B, or just A, listeners incorporated into their decision processes.

Table 2 lists the values of the relative weights for each listener for three of the Δf values tested, 2, 17, and 32 semitones. Only relative weights with values at least two standard errors of the estimates above or below zero are shown. Subscripts of "s" and "e" indicate the relative weights associated with the "start" and "end" (onsets and offsets) of the tones. A positive weight suggests that when a late onset/offset (positive perturbation) is present, listeners tended to respond "signal." Negative weights imply that early onsets/offsets (negative perturbation) were associated "signal" responses. For a simple contrast, the relative weights would be approximately equal in magnitude but opposite in sign.

Table 1. Preferred regression model for six listeners (L1–L6) listed as a function of frequency separation (Δf). For each listener, the left column shows the preferred model, based on just the A tones (restricted model) or both the A&B tones (full model) and the differences of deviances are indicated in parentheses. When the difference is larger than $\chi^2(8) = 15.5$, the full model is preferred over the restricted model ($p < 0.05$). The right column shows the Hosmer Lemeshow statistic, χ^2_{HL}, indicating the goodness of fit for the preferred model. Asterisks indicate significant deviations between the model prediction and the data ($p < 0.05$). Listener L3 was tested using 9 semitones and L6 was tested using 10 semitones.

Δf	L1		L2		L3		L4		L5		L6	
(semitones)	Model	χ^2_{HL}	Model	χ^2_{HL}	Model	χ^2_{HL}	Model	χ^2_{HL}	Model	χ^2_{HL}	Model	χ^2_{HL}
2	A&B (16.5)	6.6	A&B (20.3)	5.9	A (14.5)	8.0	A (11.3)	7.4	A&B (88.4)	14.7	A&B (26.7)	11.1
9/10					A (4.5)	9.1					A&B (34.7)	19.2*
17	A&B (19.1)	14.8	A (11.7)	7.5	A (11.6)	8.3	A (13.5)	20.3*	A&B (22.3)	14.8	A&B (17.8)	9.5
25	A (10.2)	6.4	A (14.4)	7.7	A (7.3)	10.3	A (4.4)	8.2	A&B (16.3)	9.6	A&B (27.7)	8.8
32	A (5.9)	7.1	A (11.7)	19.2*	A (6.9)	7.2	A (3.4)	6.7	A (14.5)	9.5	A (9.2)	11.6

Table 2. Relative weights more than two standard errors of the estimate from zero are listed for four B tones (B_1–B_4) and three A tones (A_1–A_3) and for Δf values of 2, 17, and 32 semitones. The tones are numbered according to Fig. 1. The subscripts "s" and "e" indicate the start and the end of a tone, respectively. The signal tone (A_2) is bracketed by two vertical lines.

	Δf (semitones)	B_{1s}	B_{1e}	A_{1s}	A_{1e}	B_{2s}	B_{2e}	A_{2s}	A_{2e}	B_{3s}	B_{3e}	A_{3s}	A_{3e}	B_{4s}	B_{4e}
L1	2	–	–	–	–	–36	–	–	–	–	–	–	−39	–	−47
	17	–	–	−35	–	–	–	59	–	–	–	−46	–	−49	−33
	32	–	–	−33	–	–	–	94	−32	–	–	–	–	–	–
L2	2	–	–	−42	−45	–	−33	66	–	40	–	–	–	–	–
	17	–	–	–	–	–	–	54	–	–	–	–	−31	–	–
	32	26	–	−40	–	–	–	68	−30	–	–	–	−38	–	–
L3	2	–	–	−32	–	34	–	39	–	–	–	–	–	–	–
	17	–	–	–	–	–	–	55	–	–	–	–	–	–	–
	32	–	–	–	–	–	–	61	–	–	–	–	–	–	–
L4	2	–	–	–	–	–	–	64	–	–	–	–	–	–	–
	17	–	–	–	–	–	–	62	–	–	–	–	–	−33	–
	32	–	–	−77	32	–	–	85	–	–	–	–	–	–	–
L5	2	–	–	–	–	–	–	81	–	–	–	68	−40	−119	–
	17	–	–	30	–	–	–	153	–	–	–	−107	−68	−52	–
	32	–	–	–	–	–	–	113	36	37	–	−120	−70	–	–
L6	2	–	–	–	–	65	–	148	–	–	–	−49	–	−36	–
	17	–	–	–	–	–	–	71	–	–	−48	−39	–	–	–
	32	–	–	–	–	–	–	–	191	–	–	−147	–	–	–

The most obvious feature of Table 2 is that listeners relied on tone A_2 to make their decision—as was appropriate given that a delay to tone A_2 was the signal to be detected. Moreover, A_{2s} was particularly important, which is consistent with models of rhythm perception which posit that onsets are crucial (e.g., Krumhansl, 2000; Jones *et al.*, 2002). A second feature to note is that whether or not the B tones have coefficients listed in Table 2 does not wholly predict whether the full or restricted model was favored (Table 1). Overall (including measurements shown in Table 1 but not Table 2), on 23% of the measurements the "significant" relative weights were not in agreement with the statistical model (e.g., Table 1 indicated model A alone, whereas Table 2 indicates a non-zero relative weight for one of the Bs, etc). While the nested statistical model might miss minor aspects of the listeners' decision processes, it provides a reasonable first-order summary. Third, it is apparent that the musicians differ from the non-musicians in their dependence on the onset of tone A_3 (A_{3s}) in forming their detection decisions.

To further explore differences between the musician and non-musician listeners, the efficiency with which listeners used their decision rules was estimated as suggested by Berg (1990). Briefly, values of d' estimated from the data and the d' predicted by the full model based on 2000 simulated trials (d'_w) were calculated. The "noise efficiency" η, was then calculated as (d'/d'_w).[2] This value indicates the degree to which listeners were efficient in using their relative weights, assuming the relative weights were perfectly estimated. On average, η was approximately twice as large for the musicians than for the non-musicians: 0.39 versus 0.20.

To summarize, the proposed approach associated with the χ^2 statistic derived from a logistic regression provided a reasonable description of the listeners' decision strategies in the sense that it reasonably separated dependencies on just A or both A and B tones for the psychophysical task. Moreover, the approach was successful even

though relatively few trials were obtained. To provide an example of an advantage of current procedure, the results indicate that the "musicians" relied on both the A and B in making their decisions even for frequency separations expected to yield separate A and B streams percepts (e.g., for values of Δf 17 semitones and larger). Current procedures, whether subjective or objective (e.g., psychometric functions), do not directly provide this information regarding decision processes.

There are several limitations to the approach as well. Although decisions regarding the full versus restricted model could be expressed statistically, the use of a single criterion (Table 1) is not wholly satisfactory. As an example, for L2, and to a lesser degree L3 and L4, the differences in the deviances, full minus restricted model, changed gradually. Potentially, these listeners altered strategies throughout data collection when intermediate values of Δf were tested. Variations in strategy would be difficult to discern unless the experimental hypothesis was explicit with regard to when the strategies would shift. An additional limitation is the use of a nested statistical model—restricting the analysis to just A or A&B models. There might be value in considering a more complex, but also more nuanced, model. A third limitation of the proposed approach is that the method depends on the number of degrees of freedom associated with the A versus B tones. In the current experiment, there were more relative weights associated with the B tones than A tones, meaning there was pressure on the procedure to choose the restricted model (only A tones) because this provided a relatively larger change in the number of degrees of freedom at relatively little cost. Simulations suggested that this bias in the procedure has the potential to be a significant problem when the number of trials is small. For these reasons, it is important to examine the χ^2 values and the values of the relative weights rather than simply forming a decision based on p values.

Acknowledgments

This work was supported by Grant No. R21 DC010058 from NIDCD. We thank Theodore S. Lin for assistance in data collection and Dr. Bruce G. Berg for helpful discussions regarding this work. The authors thank three anonymous reviewers for helpful comments on an earlier version of this manuscript.

References and links

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Bregman, A. S., and Campbell, J. (**1971**). "Primary auditory stream segregation and perception of order in rapid sequence of tones," J. Exp. Psych. **89**, 244–249.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Dobson, A. J. (**2002**). *An Introduction to Generalized Linear Models*, 2nd ed. (Chapman and Hall/CRC, Boca Raton, FL).

Jones, M. R., Moynihan, H., MacKenzie, N., and Puente, J. (**2002**). "Temporal aspects of stimulus-driven attending in dynamic arrays," Psych. Sci. **13**, 313–319.

Krumhansl, C. L. (**2000**). "Rhythm and pitch in music cognition," Psych. Bul. **126**, 159–179.

Lutfi, R. A., and Liu, C.-J. (**2011**). "A method of evaluating the relation between sound source segregation and masking," J. Acoust. Soc. Am., **129**, EL34–38.

Micheyl, C., and Oxenham, A. J. (**2010**). "Objective and subjective psychophysical measures of audtiroy stream integration and segregation," J. Assn. Res. Otol. **11**, 709–724.

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segeregation," Acta Acust. Acust. **88**, 320–332.

Snyder, J. S., and Alain, C. (**2007**). "Toward a nurophysiological theory of auditory stream segregation," Psych Bul. **133**, 780–799.

van Noorden, L. P. A. S. (**1977**). "Minimal differences of level and frequency for perceptual fission of tone sequences ABAB," J. Acoust. Soc. Am. **61**, 1041–1045.

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.