An objective measure of auditory stream segregation based on molecular psychophysics

Daniel Oberfeld

Abstract Auditory stream segregation is an important paradigm in the study of auditory scene analysis. Performance-based measures of auditory stream segregation have received increasing use as a complement to subjective reports of streaming. For example, the sensitivity in discriminating a temporal shift imposed on one B tone in an ABA sequence consisting of A and B tones that differ in frequency is often used to infer the perceptual organization (one stream vs. two streams). Limitations of these measures are discussed here, and an alternative measure based on the combination of decision weights and sensitivity is suggested. In the experiment, for ABA and ABB sequences varying in tempo (fast/slow) and duration (long/short), the sensitivity (d') in the temporal shift discrimination task did not differ between fast and slow sequences, despite strong differences in perceptual organization. The decision weights assigned to within-stream and between-stream interonset intervals also deviated from the idealized pattern of near-exclusive reliance on between-stream information in the subjectively integrated case, and on within-stream information in the subjectively segregated case. However, an estimate of internal noise computed using a combination of the estimated decision weights and sensitivity differentiated between sequences that were predominantly perceived as integrated or segregated, with significantly higher internal noise estimates for the segregated case. Therefore, the method of using a combination of decision weights and sensitivity provides a measure of auditory stream segregation that overcomes some of the limitations of purely sensitivity-based measures.

D. Oberfeld (✉)
Department of Psychology, Section Experimental Psychology,
Johannes Gutenberg-Universität, 55099 Mainz, Germany
e-mail: oberfeld@uni-mainz.de

Subjective and objective measures of auditory stream segregation

The question of how humans and animals are able to structure the mixture of sound waves arriving at their ears into different auditory objects (e.g., a single speaker in a noisy environment), known as *auditory scene analysis* (Bregman, 1990) or the *cocktail party problem* (Cherry, 1953), is one of the current hot topics in psychoacoustics, experimental psychology, and the neurosciences (e.g., Carlyon, 2004; Ciocca, 2008; Griffiths & Warren, 2004; Moore & Gockel, 2012; Shamma, 2008; Shinn-Cunningham, 2008). Strong efforts have also been made to develop algorithms that mimic the capability of the auditory system to identify auditory sources or objects (computational auditory scene analysis; e.g., Bell & Sejnowski, 1995; Cooke & Ellis, 2001; Haykin & Chen, 2005; Wang & Brown, 1999). Since the pioneering work of Al Bregman (e.g., Bregman, 1990; Bregman & Campbell, 1971) and Leon van Noorden (1975), auditory stream integration or segregation is a paradigm very frequently used for studying auditory scene analysis. In experiments on auditory streaming, sequences of at least two different types of tones (e.g., high- and low-pitched) are presented. The conditions in which subjects perceive a temporal sequence of auditory events as either a single, serially integrated auditory stream ("fusion") or multiple, segregated streams ("fission") (Miller & Heise, 1950) have been studied rather extensively (for

recent reviews, see Moore & Gockel, 2012; Shamma & Micheyl, 2010). Modeling the psychophysical results still remains a challenge, as is evident in the broad range of models that have been proposed for the phenomena, with a recent focus on neuronal and neurophysiological models (e.g., Beauvois & Meddis, 1996; Elhilali, Ma, Micheyl, Oxenham, & Shamma, 2009; Ellis & Vercoe, 1992; Nelken & Bar-Yosef, 2008; Roelfsema, 2006; Rogers & Bregman, 1993; Wang & Chang, 2008; Winkler, Denham, & Nelken, 2009).

During the first decades of research, auditory stream segregation was primarily studied on the basis of *subjective* measures. For example, van Noorden (1977) presented ABAB . . . sequences, in which two different tones alternated, and listeners varied the difference in frequency, level, or another attribute between tone A and tone B so that they could just perceive two separate streams (e.g., an isochronous sequence of high tones and an isochronous sequence of low tones). Recently, so-called *objective* or *performance-based* measures of auditory streaming have gained importance (e.g., Carlyon et al., 2010; Cooper & Roberts, 2009; Micheyl, Carlyon, Cusack, & Moore, 2005; Micheyl & Oxenham, 2010). These measures are based on the assumption that specific auditory tasks are either facilitated or rendered more difficult by auditory stream segregation, and thus that the perceptual organization can be inferred from the observed *accuracy* or *sensitivity* in a specific task. For example, it was demonstrated that the perception of temporal position or temporal order is more difficult for tones belonging to separate streams than for tones belonging to a single stream (Bregman & Campbell, 1971; Warren, Obusek, Farmer, & Warren, 1969). Performance-based measures of this kind are assumed to be less prone to the effects of expectations or instructions than are the subjective measures. Beyond that, they can be used in subjects who cannot communicate the perceived organization (e.g., Ma, Micheyl, Yin, Oxenham, & Shamma, 2010). Therefore, performance-based measures represent an important complement to subjective indices of stream segregation.

Limitations of sensitivity-based measures of auditory stream segregation

One exemplary task that has frequently been used as a performance-based measure of auditory stream segregation (for a recent review, see Micheyl & Oxenham, 2010) involves detecting a temporal shift in the onset of one tone ("target") embedded in a longer sequence of tones. For instance, Jones, Jagacinski, Yee, Floyd, and Klapp (1995) presented an isochronous sequence of high-pitched tones at a fast rate (interonset interval [IOI] = 533 ms) together with a sequence of low tones with a slower rate (IOI = 800 ms). The listeners' task was to detect a temporal shift of one of the low tones. On no-shift trials, the onset of this tone was centered between the

onsets of the two neighboring high tones. The temporal shift could be detected, for instance, by comparing the IOI between the preceding high tone and the target tone to the IOI between the target and the following high tone. If both intervals were identical, then the target was not shifted. If the two intervals were unequal, then the target was either shifted forward in time (delayed onset) or shifted backward (early onset). The critical assumptions for inferring the perceptual organization from accuracy in this task are (a) that detecting the temporal shift would be more difficult if the to-be-compared temporal intervals are long (e.g., Friberg & Sundberg, 1995; Getty, 1975; Sorkin, Boggs, & Brady, 1982) and (b) that listeners would be unable to use between-stream information. If listeners perceived the sequence as being segregated, they should compare the interval beginning with the low tone preceding the target and ending with the target tone to the interval beginning with the target tone and ending with the next low tone. Thus, two within-stream intervals of about 800 ms would need to be compared. If, on the other hand, the sequence was perceived as being integrated, then the two between-stream intervals constituted by the target tone and its two neighboring high tones could be compared, which would be an interval of only about 260 ms. Because the gap duration difference limens (GDDLs) approximately follow Weber's law for IOI durations between 200 and 1,000 ms (Friberg & Sundberg, 1995; Getty, 1975; Grondin, 2012; Rammsayer, 2010; Sorkin et al., 1982), the accuracy for detecting the shift should be *higher* in the case of integrated perception than for segregated perception, simply because the compared IOIs would be shorter than in the segregated case.

Although variants of this task have been successfully used in several studies (Boehnke & Phillips, 2005; Elhilali et al., 2009; Vliegen, Moore, & Oxenham, 1999), in some situations a sensitivity-based measure of stream segregation must necessarily fail. For example, with the frequency separation between the high and the low tones held constant, the perception as integrated or segregated would critically depend on the tempo of the sequence (Bregman, 1990; van Noorden, 1977), with faster tempi promoting stream segregation. However, this fundamental effect of sequence tempo on perceptual organization cannot be assessed in terms of sensitivity in the temporal shift discrimination task. As the tempo increases, all IOI durations are reduced, and therefore the accuracy in temporal interval discrimination would also increase (Friberg & Sundberg, 1995; Getty, 1975; Sorkin et al., 1982). This increase in sensitivity with the tempo of the sequence would be in conflict with the decrease in sensitivity caused by the higher probability of perceiving the sequence as segregated at faster tempi, and the resulting difficulty in using between-stream information. For this reason, the dependence of auditory-stream segregation on the presentation rate cannot be studied with a sensitivity-based measure of the described type.

A related but less serious problem applies even to the study of effects of the frequency difference Δf between the tones in a sequence on the probability of stream segregation. The probability of stream segregation increases with Δf, and this is reflected by a decrease in accuracy in the shift detection task (e.g., Roberts, Glasberg, & Moore, 2008). However, even for pairs of sounds presented in isolation rather than embedded in a sequence, the accuracy for judging the temporal interval between the two sounds would decrease when the frequency separation was increased (e.g., Divenyi & Danner, 1977; Izumi, 1999; Kinney, 1961; Lister, Besing, & Koehnke, 2002). Thus, the effect of stream segregation on accuracy in the temporal shift detection task would be confounded by an effect of Δf on gap duration discrimination (Boehnke & Phillips, 2005). As a consequence, when accuracy in the temporal shift detection task is used as a measure of stream segregation, then the estimate of the frequency difference at which the sequences become segregated perceptually might be biased toward lower values of Δf.

A closer look on temporal shift discrimination in alternating sequences: Sensitivity, decision weights, and efficiency measures

The aim of this study was to develop and to evaluate a new performance-based measure of stream segregation that avoids some of the limitations of previous measures and provide complementary information about the effects of auditory stream segregation on performance in a temporal discrimination task. To this end, methods of "molecular psychophysics" (Green, 1964) were used to estimate *decision weights* (also termed *perceptual weights*) from the trial-by-trial data (cf. Ahumada, 2002; Ahumada & Lovell, 1971; Berg, 1989).[1] These decision weights provide a rather direct insight into the decision process and the information sources used by the listener, rather than just summarizing the average outcome of the decision process in terms of sensitivity, the latter being the approach used in "molar psychophysics" (Green, 1964). The use of decision weights as information about the performance in a temporal shift discrimination task is best explained by taking one condition from the present experiment as an example. Figure 1 shows an ABA type of sequence, which consists of an isochronous sequence of A tones with tone frequency f_A, and a slower isochronous sequence of B tones

---

[1] The term molecular psychophysics (Green, 1964) refers to trial-by-trial analyses that provide information about the relation between a stimulus feature (e.g., the duration of one IOI in an ABA sequence) and the response of the listener (e.g., "target early" vs. "target late"). These methods, for which the alternative terms perceptual weight analysis and psychophysical reverse correlation are used, typically impose random trial-by-trial variation on the stimulus components. Correlational or regression analyses are then used to estimate the impact of the variation of each individual stimulus feature on a behavioral or neural response (for a detailed explanation, see Oberfeld, 2008; Oberfeld & Plank, 2011).

with higher frequency f_B. If integrated, the sequence would be perceived as a "galloping" rhythm "ABA-ABA-ABA..." (van Noorden, 1975), where "- " represents a pause. If segregated, two separate isochronous sequences are be perceived ("A-A-A-A..." and "B—B—B—B..."). In the experiment presented here, the task was to identify the temporal position of the last B tone (target) as being either *early* (thick black rectangle in Fig. 1) or *late* (thick gray rectangle in Fig. 1). As is indicated by the small arrows in Fig. 1, the onsets of all tones except the target were randomly perturbed. On each trial and for each tone, a random temporal shift was imposed on the onset, by drawing independently from a normal distribution. In order to decide whether the target was presented early or late, the listeners could, for example, use the interonset interval IOI_{AB-T} (beginning with the A tone preceding the target and ending with target onset) or the following interval (IOI_{BA-T}, constituted by the target and its following A tone). Due to the random perturbation of the onsets, both IOIs will vary from trial to trial. However, if the target onset was delayed by Δt, then IOI_{AB-T} would on average become longer, and IOI_{BA-T} would become shorter. For this reason, the probability of responding that the late target had been presented should increase with increases in IOI_{AB-T}, and decrease with increases in IOI_{BA-T}. In other words, the listener should assign a positive decision weight to IOI_{AB-T} and a negative weight to IOI_{BA-T}. Note that using the temporal information from IOI_{AB-T} or IOI_{BA-T} would mean that the listener made use of *between-stream* information. Alternatively, the listener could base his or her decision on the duration of the *within-stream* interval IOI_{BB-T}. The longer this interval, the higher the probability that the target had been presented at the late temporal position.

Using methods of molecular psychophysics (e.g., Ahumada & Lovell, 1971; Berg, 1989), decision weights can be estimated from the trial-by-trial data by relating the response ("target early" or "target late") to the randomly varying IOIs. For example, if the listener used information provided by IOI_{BB-T}, this would be evident in the probability to respond "target late" increasing with increases in IOI_{BB-T}. If, however, IOI_{BB-T} was absolutely unimportant for the decision, the probability of a "target late" response would be *independent* of the trial-by-trial variation in IOI_{BB-T}. Thus, the decision weights represent a quantitative measure of the attention directed to the different within-stream and between-stream IOIs (Berg, 1990).

How should different perceptual organizations be reflected in the decision weights? In sensitivity-based measures of stream segregation, it is usually assumed that a listener will use between-stream information if the sequence is perceived as integrated, but exclusively or predominantly within-stream information if the stimuli are perceived as two separate streams. Now it is evident that the decision weights described above provide a direct measure for the use of these two types
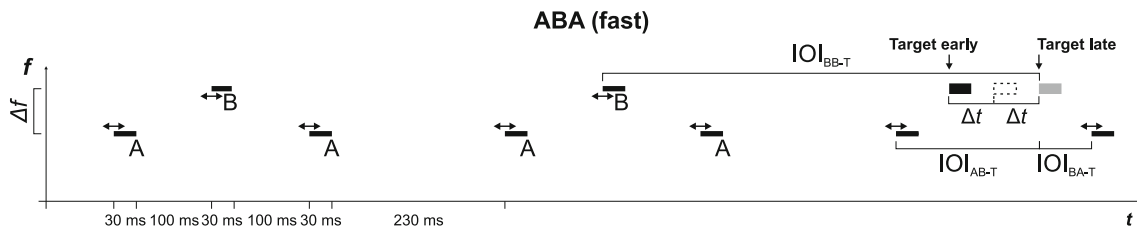
Fig. 1 ABA sequence. Schematic depiction of the ABA sequence. A fast isochronous sequence of A tones and a slower isochronous sequence of B tones were presented together. The frequency of the A tones was $\Delta f = 9$ semitones below the frequency of the B tones ($f_B = 800$ Hz). The task was to decide whether the target tone was presented *early* (black thick bar) or *late* (gray thick bar). As is indicated by the double-headed arrows, the onsets of all tones except the target tone were randomly perturbed. The brackets indicate interonset intervals (IOIs) for which decision weights were estimated. The ABA sequence was presented at two tempi. The figure depicts the fast tempo. For the slow tempo, all IOIs were multiplied by a factor of 2. The fast sequence was presented with two durations: a short sequence duration with three ABA triplets, as depicted, and a long sequence with 15 triplets. The slow ABA sequence contained seven triplets

of information. In the integrated case, according to the assumptions above, the decision variable should be dominated by the between-stream IOIs if they are shorter than the within-stream IOIs, which is the case for an ABA rhythm. In contrast, in the stream segregation case, the decision weights on the within-stream intervals should be much higher than the weights assigned to the between-stream intervals.

This reversal in the predicted relative weights on within- and between-stream IOIs is a specificity of the ABA type of sequences, owing to the fact that the within-stream IOIs including the target tone are considerably longer than the between-stream IOIs. Other rhythms do not exhibit this characteristic. To gain insight into the decision weights and sensitivity for a rhythm with similar durations of the within- and between-stream IOIs, an ABB sequence (Sussman, Wong, Horvath, Winkler, & Wang, 2007), displayed in Fig. 2, was additionally presented. In the segregated case, this type of sequences is perceived as an isochronous stream of low tones and a stream of pairs of high tones (cf. Fig. 2 in Sussman & Steinschneider, 2009). As in the ABA rhythm, the probability of perceiving the sequences as two streams increases with the frequency separation between the A and B tones (Sussman & Steinschneider, 2009). The effects of sequence tempo on stream segregation have not been studied systematically for the ABB type of sequence. In the present experiment, the target tone in the shift discrimination task was the penultimate B tone of the ABB sequence (see Fig. 2). Concerning the information provided by within-stream and between-stream

IOIs, the ABB sequences differ markedly from the ABA sequences. In the ABB sequence, the target is preceded by a between-stream interval (IOI_{AB-T}) and followed by a within-stream interval (IOI_{BB+1-T}) of the *same* duration as the between-stream interval. For this reason, in the integrated case the subject should place equal weights on the two IOIs.

At this point, it is important to note that the optimum weights maximizing the percentage of correct responses can be assumed to differ between a fast and a slow sequence, even in the absence of stream segregation. Imagine that for the fast and slow ABA sequence described above, the listener's task was to discriminate a temporal shift of the target tone onset of ±25 ms. For the slow presentation rate used in the experiment, the within-stream interval IOI_{BB-T} had a duration of 1,040 ms. What is the just-noticeable temporal shift for this base interval—that is, the gap duration difference limen (GDDL)? According to Friberg and Sundberg (1995), the Weber fraction (GDDL/IOI) can be expected to be 5% for a 1,040-ms IOI, so that the GDDL (just-noticeable temporal shift) should be $GDDL_{BB} = 1{,}040 \text{ ms} \cdot 0.05 = 52$ ms. The duration of the between-stream interval IOI_{AB-T} was 260 ms in the slow ABA sequence, corresponding again to a Weber fraction of 5% (Friberg & Sundberg, 1995), and $GDDL_{AB} = 260 \text{ ms} \cdot 0.05 = 13$ ms. Thus, the temporal shift of 25 ms should be easy to detect in the between-stream interval, but would be subliminal for the within-stream interval. As a consequence, listeners should assign a higher weight to IOI_{AB-T}. In contrast, for the *fast* ABA sequence presented in our experiment, the within-stream interval
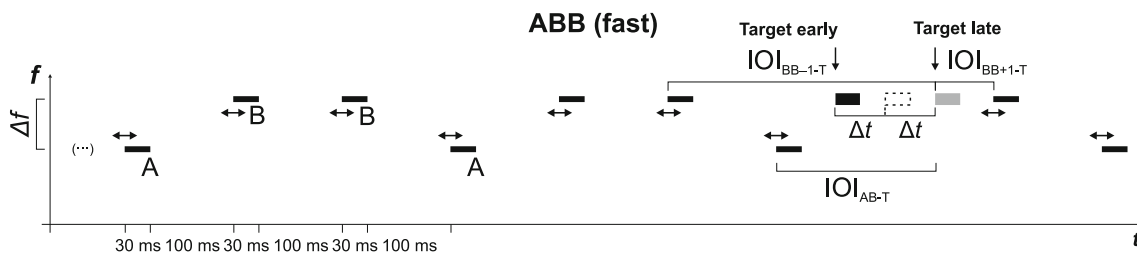

Fig. 2 ABB sequence. Schematic depiction of the fast ABB sequence. The target tone was the penultimate B tone, which could be presented either *early* (black thick bar) or *late* (gray thick bar). For the slow tempo, all IOIs were multiplied by a factor of 2

IOI_{BB-T} had a duration of 520 ms (Weber fraction = 5%, $GDDL_{BB}$ = 520 ms · 0.05 = 26 ms). The duration of the between-stream interval IOI_{AB-T} was 130 ms (Weber fraction = 9.6%, $GDDL_{BB}$ = 130 ms · 0.096 = 12.5 ms). Thus, the difference in temporal interval discrimination performance between the within-stream and between-stream intervals should be smaller for the fast than for the slow sequence, even *without* differences in sequential streaming. This example demonstrates that it may be difficult to unequivocally relate changes in decision weights caused by changing the tempo to differences in the perceived grouping, just as was discussed for sensitivity.

To address this issue, measures of observer efficiency (Berg, 1990; Tanner & Birdsall, 1958) were used to put the observed sensitivity and the decision weights into the context of the temporal resolution underlying performance in the shift discrimination task. On a more general level, the experiment took into account that two different factors could limit observers' performance in the shift discrimination task. First, the information about, for example, the duration of IOI_{AB-T} that was available at the decision stage might be inexact—for example, due to the inherent variability of the sensory system, which is often described and modeled as *internal noise* (Swets, Shipley, McKey, & Green, 1959). According to the assumptions underlying the sensitivity-based measures of stream segregation, the information about IOI_{AB-T} is degraded if the listener perceives the sequence as being two segregated streams. In other words, streaming is expected to cause an increase in internal noise. Second, the different sources of information available at the decision stage (i.e., the information about the different IOI durations that can be used to infer the temporal position of the target tone) could be combined in a suboptimal fashion (Swets et al., 1959). A reduction in sensitivity caused by sequential stream segregation could be caused by either of these factors alone, or by both. Therefore, sensitivity-based measures alone do not allow for deciding whether stream segregation increases internal noise, or results in an inappropriate integration of information, or both. A recent study by Richards, Carreira, and Shen (2012)—which appeared after the data collection for the present study had been completed—indicated that estimates of decision weights alone are also not sufficient for differentiating between sequences perceived as integrated or segregated. Their results indicated that between-stream IOIs received a significant weight even in some cases in which the sequence was clearly perceived as being segregated, due to a large frequency difference between the A and B tones. The present study takes the molecular psychophysics approach one step farther, by *combining* decision weights and estimates of sensitivity (see Berg, 2004, for an excellent explanation of these techniques). This allowed for computing efficiency measures that quantify the streaming-induced loss in sensitivity due to (a) increases in internal noise or (b) suboptimal decision weights (Berg, 1990). Importantly, the proposed efficiency measures for the first time take into account the differences in sensitivity in the absence of streaming, which above have been identified as potential confounds of the purely sensitivity-based measures of stream segregation. These analyses were possible because in the experiment, GDDLs were measured for isolated temporal intervals with the same duration and marker frequency as the IOIs presented in the ABA and ABB sequences.

Indeed, the results showed that the sensitivity in the temporal shift discrimination task did not differ between the fast and slow sequences, despite the fact that these sequences clearly differed in their perceptual organization (one stream vs. two streams, as revealed by subjective ratings). In contrast, an observer efficiency measure indexing the increase of internal noise relative to a situation without stream segregation was able to dissociate the fast from the slow sequences and to predict the subjective reports of the perceptual organization on an individual basis.

Method

The study comprised two experiments. In the main experiment, listeners judged the temporal position of one tone (the target) in ABA and ABB tone sequences (temporal shift discrimination). Decision weights for the different IOIs, as well as the sensitivity and subjective ratings of streaming, were obtained. In addition, GDDLs were measured for isolated temporal intervals with the same duration and marker frequency as the IOIs in the ABA and ABB sequences.

In order to select appropriate parameters for the main experiment, a pretest was conducted, which is described in detail in the Appendix.

Ethics statement

The experiments were conducted according to the principles expressed in the Declaration of Helsinki. All subjects participated voluntarily after providing informed written consent. They received partial course credit or were paid for their participation.

Apparatus

The stimuli were generated digitally, played back via two channels of an RME ADI/S D/A converter (f_s = 44.1 kHz, 24-bit resolution), attenuated by a TDT PA5 attenuator, buffered by a TDT HB7 headphone buffer, and presented diotically via Sennheiser HDA 200 circumaural headphones calibrated according to IEC 318 (1970). The experiment was conducted in a double-walled sound-insulated chamber. The listeners were tested individually.

Subjects

Ten subjects participated in the experiment. All of them reported normal hearing. The data of two subjects had to be excluded from the analysis, because of unusually high gap duration difference limens and very low sensitivity in the temporal shift discrimination task in all conditions. For the remaining eight subjects (seven female, one male; 20–29 years of age), detection thresholds measured by von Békésy tracking (Hartmann, 2005, pp. 132–133; von Békésy, 1947) were better than 17 dB HL between 125 Hz and 8 kHz, for both ears. Three of the subjects were research assistants at this lab and had also participated in the pretest (see the Appendix).

Stimuli, procedures, and conditions

Temporal shift discrimination task The stimuli were tone sequences consisting of pure tones with a duration of 30 ms (including 5-ms $\cos^2$ ramps), presented diotically with a sound pressure level of 60 dB SPL. Two different rhythms were presented: ABA and ABB sequences (see Figs. 1 and 2). Both consisted of A and B tones differing in frequency (f_B = 800 Hz; f_A = 475.68 Hz, Δf = 9 semitones). The frequency difference Δf was kept fixed, to avoid complications with effects of across-frequency gap discrimination (e.g., Kinney, 1961).

Both rhythms were presented at two different tempi (fast vs. slow), in order to vary the probability of perceiving the sequences as integrated or segregated (Bregman, 1990; van Noorden, 1975). Since stream segregation needs time to build up (Anstis & Saida, 1985; Bregman, 1978; Carlyon et al., 2010; Snyder, Alain, & Picton, 2006; Thompson, Carlyon, & Cusack, 2011), two different sequence durations (short and long) were also presented. This allowed for inducing a stronger tendency toward segregation in the longer sequence, while keeping the sequence parameters constant, which is an important feature because it avoids the effects of IOI duration and Δf on sensitivity in the temporal shift discrimination task. For the fast sequences, both durations were presented. For the slow sequence, only the long duration was used.

For the fast ABA sequence (see Fig. 1), the IOI between the first A tone of an ABA triplet and the following B tone was 130 ms, the IOI between the B tone and the last A tone of the triplet was also 130 ms, and the IOI between the last A tone and the first A tone of the following triplet was 260 ms. For the slow ABA sequence, all of the IOIs were multiplied by a factor of 2. The tone duration was constant in all conditions. The short ABA sequences comprised three ABA triplets (sequence duration = 1,330 ms). The fast long ABA sequence comprised 15 ABA triplets (sequence duration = 7,570 ms). The slow long ABA sequence comprised seven triplets (sequence duration = 6,790 ms). The perception of this type of

sequence as either integrated or segregated seems to stabilize after 10 s (Anstis & Saida, 1985). Slightly shorter durations were used to reduce the experimentation time and to avoid bistability (Pressnitzer & Hupé, 2006).

For the ABA rhythm, the target tone was the last B tone of the sequence. On each trial, it was presented either *early* (onset Δt before the midpoint between the onsets of the two A tones in the last triplet) or *late* (onset Δt after the midpoint between the onsets of the two A tones). On the basis of the results of Pretest 1 (see the Appendix), Δt was set to 26 ms for all listeners and conditions. The target tone was not marked by using a different duration or a synchronous visual signal, in order to avoid problems with stream resetting (Haywood & Roberts, 2010). To make the estimation of decision weights possible, the onsets of all tones except the target were randomly and independently perturbed. For each onset, a random temporal shift was drawn from a normal distribution with mean μ = 0 ms. The standard deviation of the distribution was set to σ = 20 ms for the fast and σ = 40 ms for the slow sequences. To avoid temporal overlap between neighboring tones, the random perturbations were restricted to 0 ms $\pm$ 2.5 · σ.

The task of the subjects was to decide whether the ultimate B tone was presented early (i.e., "A-B--A," where the dash denotes a silent gap) or late (i.e., "A--B-A"). In the first sessions, the listeners first received practice blocks without random perturbations of the tone onsets, in order to make the task clear. They then received practice blocks with random timing perturbations. The responses were indicated on a four-point rating scale, with the ordered response categories *Early–rather certain*, *Early–rather uncertain*, *Late–rather uncertain*, and *Late–rather certain*. This rating scale, including information about the confidence when giving the response, was used in order to be able to construct ROC curves for estimating the sensitivity. This method avoided the necessity to make potentially unjustified assumptions about the form of the ROC curve, which would be necessary if d' based on binary responses had been used as a measure of sensitivity (e.g., Macmillan, Rotello, & Miller, 2004; Swets, 1986b). Note that unlike in most previous experiments using a temporal shift discrimination task, the task here cannot be viewed as the detection of a temporal irregularity (e.g., Brochard, Drake, Botte, & McAdams, 1999) because, due to the random perturbations, all IOIs varied from trial to trial, not only the two IOIs within the last ABA triplet that contained the target.

For the fast ABB sequence (see Fig. 2), all of the IOIs between neighboring tones had a duration of 130 ms. For the slow ABB sequence, the IOIs were 260 ms long. Three ABB triplets were presented for the fast short ABB sequence, plus one appended A tone (sequence duration = 1,200 ms). The fast long ABB sequence comprised 20 ABB triplets plus one appended A tone (sequence duration = 7,830 ms). The slow long ABB sequence comprised ten triplets plus one A tone

(sequence duration = 7,830 ms). For the ABB rhythm, the target tone was the penultimate B tone of the sequence. On each trial, it was presented either early (onset Δt before the midpoint between the onsets of the two neighboring tones) or late (onset Δt after the midpoint between the onsets of the two neighboring tones). The tone duration, tone frequencies, sound pressure level, value of Δt, and standard deviation of the random timing perturbations were identical to those values in the ABA sequences.

In each experimental block, 25 trials were presented with a delayed target onset, 25 trials with an early target onset, and 50 no-shift trials ($\Delta t = 0$). Only a single sequence (i.e., Rhythm × Tempo × Duration combination) was presented per block. Trial-by-trial feedback was provided only in the practice blocks. In the experimental blocks, information was provided about the percentage of correct responses on the shift trials at the end of each block.

A given sequence may sometimes be perceived as integrated, and sometimes as segregated (Bregman, 1990). In addition, the perceptual organization may change during the presentation (Pressnitzer & Hupé, 2006), showing bistability. In order to be able to compare the performance-based measures of stream segregation to the perceptual organization, the listeners were asked to indicate after each trial whether they had perceived one or two streams at the *end* of the sequence, where the target tone was located. Thus, even on trials in which the perceptual organization may have switched between one stream and two streams during the presentation of the sequence, the responses of the subjects could be assumed to reflect the perceptual organization in the target triplet. The subject first listened to the sequence, then indicated the temporal position of the target tone on the four-point rating scale, and finally indicated whether at the end of the sequence he or she had perceived one or two streams. The next sequence started after an intertrial interval of 2 s.

Measurement of gap discrimination difference limens (GDDLs) As is explained in the Estimation of Decision Weights section below, for the ABA sequences, decision weights were estimated for three temporal intervals involving the target tone and its adjacent A and B tones (IOI_{AB-T}, IOI_{BA-T}, and IOI_{BB-T}; see Fig. 1). Two of these temporal intervals were marked by tones of different frequencies. For the ABB sequences, the analyzed IOIs were IOI_{AB-T}, IOI_{BB+1-T}, and IOI_{BB-1-T} (see Fig. 2). The GDDLs were measured for the corresponding IOI durations of 130, 260, 520, and 1,040 ms, and taking into account the potential effect of frequency differences between the two tones marking a temporal interval (Divenyi & Danner, 1977; Izumi, 1999; Kinney, 1961; Lister et al., 2002). These GDDLs were used to estimate the optimal decision weights in the absence of stream segregation—that is, when there was no difficulty in using between-stream IOIs. They were also used to compute measures of observer efficiency, which made it possible to distinguish between the roles of suboptimal decision weights and internal noise.

Just as in the ABA and ABB sequences, for the 130-ms and 260-ms base IOIs, the temporal gaps were marked either by BB pairs of tones (i.e., both tones had a frequency of 800 Hz) or by AB pairs (i.e., the first tone was presented at $f_A = 475.68$ Hz, and the second tone at $f_B = 800$ Hz). For the two longest base IOIs, all gaps were marked by BB pairs. As in previous studies (e.g., Divenyi & Danner, 1977), it was tacitly assumed that the GDDLs would not differ between AB and BA pairs. The sound pressure level and the durations of the tones were identical to the values used in the temporal shift discrimination task.

A one-interval, absolute identification task was applied (e.g., Green, von Gierke, & Hanna, 1986). For a given combination of base IOI and frequency difference between the first and second tones, on each trial the IOI was randomly selected from a normal distribution with a mean equal to the base IOI (e.g., $\mu = 130$ ms, corresponding to IOI_{AB-T} at the fast tempo). The SDs of the distribution were 20 ms for the 130-ms and 260-ms base IOIs, 40 ms for the 520-ms base IOI, and 100 ms for the 1,040-ms base IOI. On each trial, listeners classified the presented IOI as being either short or long (i.e., comparison with an implicit standard; Nachmias, 2006), using a four-point rating scale with the values *Short–rather certain*, *Short–rather uncertain*, *Long–rather uncertain*, and *Long–rather certain*. Visual trial-by-trial feedback was provided. Each experimental block presented 155 trials of only one Base IOI × Frequency Difference combination. For each listener, two blocks presenting 155 trials were run for each of the six Base IOI × Frequency Difference combinations, in separate sessions.

The first six trials per block were excluded from the data analysis. For each block, a cumulative-normal psychometric function (PMF) was fitted to the dichotomized responses (*early–rather certain* and *early–rather uncertain* vs. *late–rather uncertain* and *late–rather certain*), using a maximum-likelihood (ML) approach (e.g., Treutwein & Strasburger, 1999). The ML estimate of the SD parameter represents the spread of the PMF.

The GDDL was defined as half the difference between the 75% and 25% points on the PMF, GDDL = $(x_{.75} - x_{.25})/2$. For each block (155 trials) obtained in the experiment, the GDDL was computed from the ML estimate of the SD parameter of the PMF, using the relation GDDL = $0.67449 \cdot SD$, which applies to a cumulative-normal PMF.

Data analysis

Estimation of sensitivity in the temporal shift discrimination task For each experimental block, an ROC curve was constructed from the observed frequencies of the rating responses

on early-target and late-target trials (for details, see Macmillan & Creelman, 2005, chap. 3). The first five trials per block were excluded from the analysis. The area under the ROC curve (AUC), converted to d', was used as an index of sensitivity. AUC does not require strong assumptions about the internal distributions of "signal" and "noise" (e.g., Macmillan et al., 2004; Swets, 1986b). It corresponds to the proportion of correct responses obtained with the same stimuli in a forced choice task (e.g., Green & Moses, 1966; Iverson & Bamber, 1997), if bias-free responding can be assumed. To compute AUC, an ML procedure (Dorfman & Alf, 1969) was used for fitting a binormal model (Hanley, 1988).[2] For each block, AUC was computed from the ML estimates of slope and intercept of the ROC curve (Swets, 1979). The observed values of AUC were then transformed to d'. Given the correspondence between AUC and $P(C)$ in a forced choice task with unbiased responding, the relation $d'_{2I} = \sqrt{2} \; z(AUC)$ can be used, where $z(P)$ is the standard normal deviate corresponding to the proportion P (cf. Macmillan & Creelman, 2005, pp. 170–172; Swets, 1986b, Eq. 21). The advantage of using d' rather than AUC is that d' can be viewed as a linearization of the binomial quantity AUC, and d' is often found to be linearly related to stimulus magnitude (e.g., Buus & Florentine, 1991; Moore, Peters, & Glasberg, 1999). A linearization is also desirable because repeated measures analyses of variance (ANOVAs) are sensitive to departures from normality (Oberfeld & Franke, 2013).

Estimation of decision weights The decision weights representing the importance of particular IOIs for the decision in the temporal shift discrimination task were estimated from the trial-by-trial data using multiple logistic regression (Alexander & Lutfi, 2004; Dittrich & Oberfeld, 2009; Oberfeld, 2008; Pedersen & Ellermeier, 2008). For both rhythms and both sequence durations, decision weights were estimated for three IOIs involving the target tone. For the ABA sequences, these IOIs were (a) the between-stream interval ending with target onset (IOI_{AB-T}; see Fig. 1), (b) the between-stream interval beginning with the target onset (IOI_{BA-T}), and (c) the within-stream interval preceding the target (IOI_{BB-T}). For the ABB sequences, the analyzed IOIs were (a) the between-stream interval ending with target onset (IOI_{AB-T}; see Fig. 2), (b) the within-stream interval following the target (IOI_{BB+1-T}), and (c) the within-stream interval preceding the target (IOI_{BB-1-T}). The respective three IOIs were used as predictors/covariates in the multiple logistic regression model. Note that, in principle, all other IOIs might also contribute to the decision variable. For example, for the ABA rhythm, IOI_{BB-T} might be compared to previous B–B intervals in the sequence. However, the three intervals directly involving the target were

expected to be most important for the decision. In fact, additional analyses (not shown, due to lack of space) showed that the decision weights assigned to the intervals comprising tones from the preceding triplet were consistently lower than those for the three selected IOIs, and were nonsignificant for most listeners. Analyzing additional IOIs would also have increased the complexity of the analyses, and would have required more trials for weight estimation. However, in future experiments, additional sources of information can be included simply by adding additional IOIs to Eq. 1.

The ordered categorical rating responses (*Early–rather certain, Early–rather uncertain, Late–rather uncertain, Late–rather certain*) served as the dependent variable. The predictors were entered simultaneously. A proportional-odds model was used (McCullagh, 1980). The regression coefficients were taken as the decision weight estimates. For a given IOI, a regression coefficient equal to zero would mean that the IOI duration had no influence at all on the decision to judge the target position as being either early or late. A regression coefficient greater than zero would mean that the probability to respond that the *late* target had been presented increased with the duration of the given IOI. A regression coefficient smaller than zero would indicate the opposite relation between IOI duration and the probability to respond that the late target had been presented.

This analysis is based on a decision model assuming that listeners use a decision variable

$$D_j(\mathbf{IOI}) = \left(\sum_{i=1}^{k} w_i IOI_i\right) - c_j, \tag{1}$$

where IOI_i is the duration of a particular IOI (e.g., IOI_{AB-T}), k is the number of decision-relevant IOIs, $\mathbf{IOI}$ is the vector of IOIs, w_i is the perceptual weight assigned to IOI_i, and c_j is a constant representing the decision criterion for the jth of the four ordered response categories (cf. Agresti, 1989; Berg, 1989; Pedersen & Ellermeier, 2008). In other words, $D_j(\mathbf{IOI})$ is a weighted average of the different IOI durations.

Because of the four-category response variable Y, a proportional-odds model was assumed (McCullagh, 1980), according to which

$$P(Y \leq j) = \frac{e^{D_j(\mathbf{IOI})}}{1 + e^{D_j(\mathbf{IOI})}}, j = 1, \ldots, J-1, \tag{2}$$

where J is the number of ordered response categories. This model applies simultaneously to all $J-1$ cumulative probabilities, and it assumes identical effects of the predictors for all cumulative probabilities (Agresti, 1989).

---

[2] Swets (1986a, 1986b) demonstrated that empirical data are typically consistent with the assumption of Gaussian distributions. The binormal model has also been shown to be robust against violations of this assumption (Hanley, 1988).

A separate logistic regression model was fitted for each combination of subject and sequence (Rhythm × Tempo × Duration). Since the interest here was in the relative contributions of the three different IOIs to the decision rather than in the absolute magnitude of the regression coefficients, the weights w_i were normalized for each fitted model, such that the sum of their absolute values was unity (see Kortekaas, Buus, & Florentine, 2003), resulting in a set of relative decision weights for each listener and sequence.

Estimation of ideal weights On each trial, the three different IOIs analyzed in the weight estimation procedure provided information concerning the temporal position of the target tone (early or late). Which decision weights would an observer maximizing the proportion of correct responses assign to the different IOIs? Even without the potentially detrimental effects of stream segregation, the information concerning the correct response in the temporal shift discrimination task provided by a given IOI is reduced by the external variability due to the random perturbations of the tone onsets ("external noise"; e.g., Jesteadt, Nizami, & Schairer, 2003; Swets et al., 1959).[3] It is also compromised by internal variability, in the sense of *internal noise* (e.g., Swets et al., 1959). The optimal decision strategy would be to place the highest decision weight on the IOI providing the most reliable information about the temporal position of the target tone (integration model; e.g., Green, 1958). Thus, the optimal decision weights would depend on the individual sensitivity for temporal interval discrimination, and on the external variability. For this reason, GDDLs were measured for the different IOIs relevant in the shift discrimination task. These individual empirical GDDLs were used for determining the optimal decision weights for each subject, separately for each condition in the temporal shift discrimination task. The analysis assumed the absence of stream segregation—that is, no difficulty in using between-stream IOIs—apart from the effects of the frequency difference between the two tones marking an IOI (e.g., Divenyi & Danner, 1977). The latter effect is already included in our measurements of the GDDLs. It was also assumed that presenting an IOI embedded in a longer sequence of tones would not result in lower sensitivity for judging its duration than if the IOI were presented in isolation, as in the gap duration discrimination task. Thus, the ideal weights computed here reflect gap duration discrimination in the absence of effects caused by presenting longer sequences of tones.

Although in principle the ideal weights could be derived analytically (Berg, 1990; Oberfeld, Kuta, & Jesteadt, 2013), the correlation between, for example, IOI_{AB-T} and IOI_{BA-T}—caused by the temporal shifts of the target—makes it more

difficult to find a solution in closed form. Therefore, a Monte Carlo method was applied. Each of the $j = 1, …, k$ IOIs that were relevant for the decision (e.g., IOI_{AB-T}) was assumed to elicit a value X_{ij} on the internal continuum. These values were modeled as

$$X_{ij} = IOI_j + Z_{ij}; Z_{ij} \sim N(0, \sigma_{ij}), i = 1, …, n, \qquad (3)$$

where IOI_j is the duration of a specific IOI presented on a given trial (e.g., IOI_{AB-T}), Z_{ij} is a random variable representing the effect of additive internal noise, and i indexes the n different subjects. The Z_{ij} were independent and normally distributed with mean 0 ms and a standard deviation σ_{ij} selected on the basis of the individual GDDL for this particular interval. The value of σ_{ij} was set to the GDDL that had been measured for subject i in the temporal-interval discrimination task for an IOI with the same mean duration as IOI_j and with the same frequency difference between the two tones constituting the interval. As an example, consider a fast ABA sequence for which $IOI_1 = IOI_{AB-T}$ had a mean duration of 130 ms. If now for Subject 1 the GDDL estimated for an interval duration of 130 ms and for the two tones constituting the interval differing in frequency by nine semitones was 30 ms, then σ_{11} was set to 30 ms/0.67449, because for a cumulative normal PMF the DL, defined as half of the difference between the 75% and 25% points on the PMF, is just $0.67449 \cdot \sigma$, where σ is the standard deviation of the normal distribution representing internal noise. Due to the random perturbations, the IOIs were also normally distributed, $IOI_j \sim N(\mu_j, \sigma_j)$, where μ_j denotes the mean duration of IOI_j and σ_j its standard deviation.

For each combination of subject, rhythm, and tempo, 5,000 trials were simulated with the early target tone, and 5,000 trials with the late target. It was not necessary to distinguish between short and long sequences, because the three IOIs considered in the simulation were generated identically for both sequence durations. The randomly perturbed tone onsets were computed exactly as in the experiment, and the $k = 3$ random numbers Z_{ij} were recorded for each trial. In order to estimate the optimal decision weights for a particular listener and condition, a multiple logistic regression model was used, relating the temporal position of the target (early or late)—that is, the correct response—to the predictors X_{ij}. The resulting regression coefficients for the IOI_j would maximize the probability of a correct response, and therefore represent ideal weights. As for the estimated decision weights, the ideal weights were normalized so that the sum of their absolute values was 1.0.

Sessions The data were collected in a completely within-subjects design. Each listener participated in a total of 15 experimental sessions. In Session 1, audiometric hearing levels were measured bilaterally. Practice blocks for all

---

[3] The task can be viewed as a sample discrimination task (Berg & Robinson, 1987; Lutfi, 1989; Sorkin, Robinson, & Berg, 1987).

experimental conditions were presented in Sessions 1–3. In Session 4, GDDLs were measured. In Sessions 5–14, the temporal shift discrimination task was presented. Only one rhythm (ABA or ABB) was presented per session (in alternating order), to help the listeners adopt the optimal decision strategy for one particular type of sequence. In each session, one experimental block of 100 trials was presented for each sequence (fast long, fast short, and slow long), in random order. For each sequence (Rhythm × Tempo × Duration), five blocks (and thus a total of 500 trials) were obtained, in different sessions. In Session 15, the GDDLs were measured for a second time. Each session had a duration of about 60 min.

Results

Gap duration discrimination limens

Figure 3 shows the mean relative GDDLs (i.e., Weber fractions) as a function of base IOI and frequency separation. For the same-frequency condition and base IOIs of 130, 260, 520, and 1,040 ms, the average relative GDDLs were 7.5%, 6.5%, 5.1%, and 7.2%, respectively. These values are very similar to GDDLs reported previously (Friberg & Sundberg, 1995; Hirsh, Monahan, Grant, & Singh, 1990).[4] The higher relative GDDL at the shortest base IOI is also compatible with previous studies (Friberg & Sundberg, 1995; Hirsh et al., 1990; Rammsayer, 2010). However, the latter experiments showed an approximate compatibility with Weber's law at IOIs longer than 200 ms, whereas in the present data the Weber fraction was higher at the 1,040-ms than at the 520-ms base IOI. This pattern was also present for one subject in Matthews and Grondin (2012).

To test for an effect of the frequency difference between the two tones marking the temporal interval, a repeated measures ANOVA was conducted for the data obtained at the 130- and 260-ms base IOIs. As in previous studies (e.g., Divenyi & Danner, 1977; Hirsh et al., 1990), the relative GDDL was significantly higher if the two tones marking the temporal interval differed in frequency (i.e., A–B rather than A–A), $F(1, 7) = 13.28$, $p = .008$, Cohen's (1988) $d_z = 1.29$. Cohen defines values of 0.2, 0.5, and 0.8 as small, medium, and large effect sizes, respectively. Descriptively, the relative GDDL was higher at the 130-ms than at the 260-ms base IOI, again compatible with previous results (Friberg & Sundberg, 1995), but the effect of base IOI was not significant, $F(1, 7) = 2.57$,

---

[4] Friberg and Sundberg (1995) stated on p. 2562 that their GDDLs, computed from an adjustment task, should be multiplied by a factor of 2 in order to be comparable with GDDLs estimated using the method of constant stimuli.
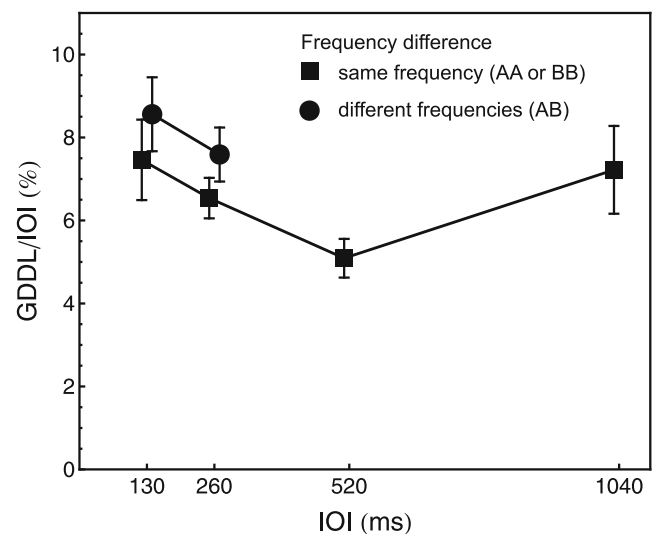
Fig. 3 Mean relative gap duration difference limens (GDDL/IOI as a percentage) as a function of the interonset interval (IOI) duration and the frequency separation between the two tones marking the temporal interval. Squares show same frequencies (A–A or B–B). Circles show different frequencies (A–B; Δf = 9 semitones). Error bars show ±1 *SEM* of the eight individual values

$p = .15$. The Base IOI × Frequency Difference interaction was also not significant, $F(1, 7) = 0.04$.

Subjective reports of the perceptual organization

Figure 4 shows the average proportions of trials on which the subjects reported perceiving one stream (integrated) at the end of the sequence, termed *P*(one stream). As expected, the slow sequences were predominantly perceived as integrated, and the fast long sequences as segregated. As intended, the short sequences were only seldom perceived as segregated. For the fast long ABB sequences, the subjects reported a higher amount of stream segregation than for the fast long ABA sequences.

In our experimental design, the factors Rhythm, Tempo, and Sequence Duration were not fully crossed, because slow but short sequences were not presented. For this reason, here and in the following discussion the data were analyzed with two separate repeated measures ANOVAs, one for the fast sequences (analyzing the effect of sequence duration), and one for the long sequences (analyzing the effect of tempo).

For the fast sequences, a repeated measures ANOVA with the within-subjects factors Rhythm (ABA or ABB) and Sequence Duration (short or long) showed a strong and significant effect of sequence duration on the proportions of trials perceived as integrated, $F(1, 7) = 73.89$, $p < .001$, $d_z = 3.04$, and a significant Rhythm × Sequence Duration interaction, $F(1, 7) = 10.35$, $p = .015$. The effect of rhythm was not significant, $F(1, 7) = 4.03$, $p = .085$. This analysis confirms the expected differences in subjective organization between the long and short fast sequences.
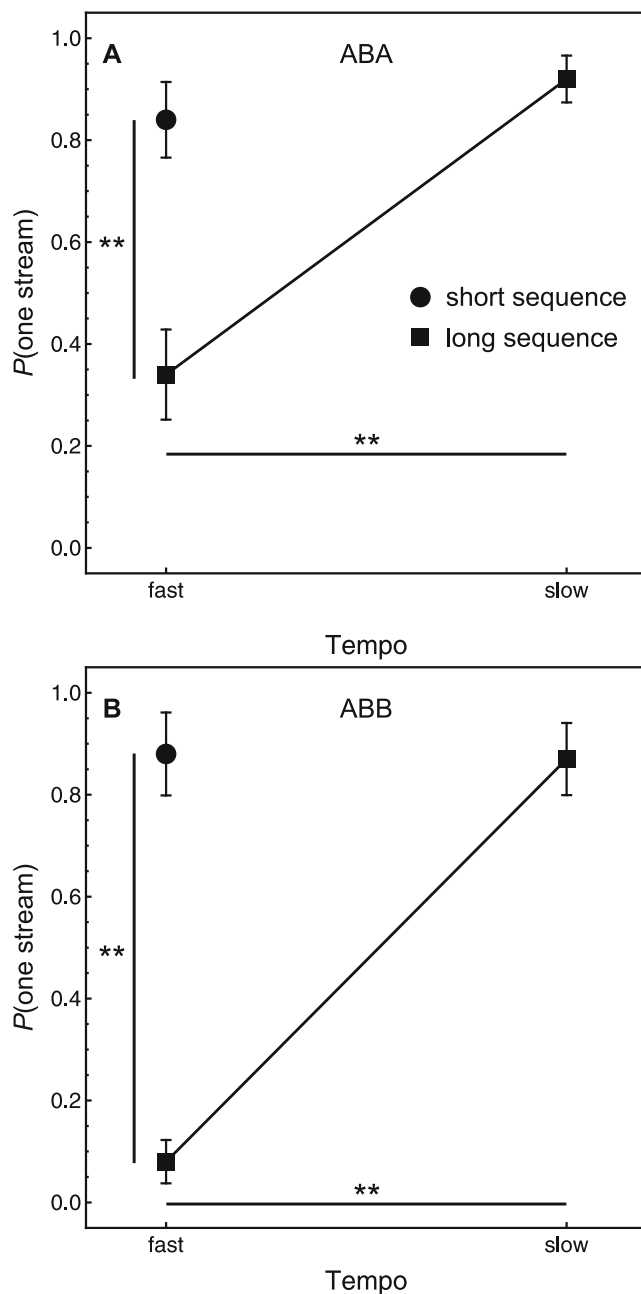
Fig. 4 Subjective reports of auditory stream segregation and integration. The figure shows the average proportion of trials on which the subjects reported to have perceived one stream (integrated), as a function of rhythm, tempo, and sequence duration. Upper panel: ABA rhythm. Lower panel: ABB rhythm. Squares are long sequences and circles are short sequences. Error bars show ±1 *SEM* of the eight individual values. Lines marked by asterisks denote significant pairwise differences

For the long sequences, a repeated measures ANOVA with the within-subjects factors Rhythm and Tempo showed a significant effect of tempo, $F(1, 7) = 99.29$, $p < .001$, $d_z = 3.52$. Thus, the variation in tempo had the expected effect on the probability of perceiving the sequences as integrated. The effect of rhythm was also significant, $F(1, 7) = 21.07$, $p = .003$, $d_z = 1.62$. On average, the ABB rhythm was perceived as

being segregated with a higher probability than the ABA rhythm, presumably due to the shorter IOIs in the ABB rhythm (Bregman, 1990). The Rhythm × Tempo interaction was nonsignificant, $p = .16$.

Inspection of the individual data showed very consistent effects of tempo and duration on the reported organization. For all listeners and both rhythms, the probability of perceiving the sequences as integrated was higher for the fast short than for the fast long sequences, and it was higher for the slow long than for the fast long sequences. The actual numerical probabilities of perceiving a given sequence as integrated differed between subjects, however.

Sensitivity in the temporal shift discrimination task

Figure 5 shows the average sensitivity (AUC converted to d') in the temporal shift discrimination task. For the fast sequences, a repeated measures ANOVA with the within-subjects factors Rhythm and Sequence Duration showed a significant effect of sequence duration on d', $F(1, 7) = 27.46$, $p = .001$, $d_z = 1.82$. The sensitivity was higher for the fast short than for the fast long sequence, for both rhythms and all listeners. The effect of rhythm and the Rhythm × Sequence Duration interaction were not significant (both ps > .9). Thus, for both rhythms the sensitivity in the temporal shift discrimination task reflected the expected difference in sensitivity due to stream segregation versus integration (cf. the subjective reports in Fig. 4). Note that the IOIs and the frequency difference between the A and B tones were constant in this analysis, only the durations of the sequences changed. Therefore, the observed difference in sensitivity cannot be attributed to confounding changes in the former two parameters.

Is it also possible to infer the differences in perceptual organization between the fast and slow long sequences from sensitivity? As was discussed in the introduction, the answer is *No* in this case, because the fast and slow sequences differed in terms of the IOI durations. As a consequence, differences in sensitivity between fast and slow sequences could be expected to reflect the effects of the IOI duration on the one hand, and of sequential streaming on the other hand. The former factor should result in higher sensitivity in the fast sequences, and the latter factor should result in higher sensitivity in the slow sequences. It is therefore not very surprising that the sensitivities did not differ between the fast long and slow long sequences (Fig. 5), whereas the subjective ratings (Fig. 4) show a clear difference in perceptual organization between these two conditions. A repeated measures ANOVA with the within-subjects factors Rhythm and Tempo showed no significant effect of tempo on sensitivity, $F(1, 7) = 0.106$, $p = .76$. The effect of rhythm and the interaction were also nonsignificant (both p values > .13).
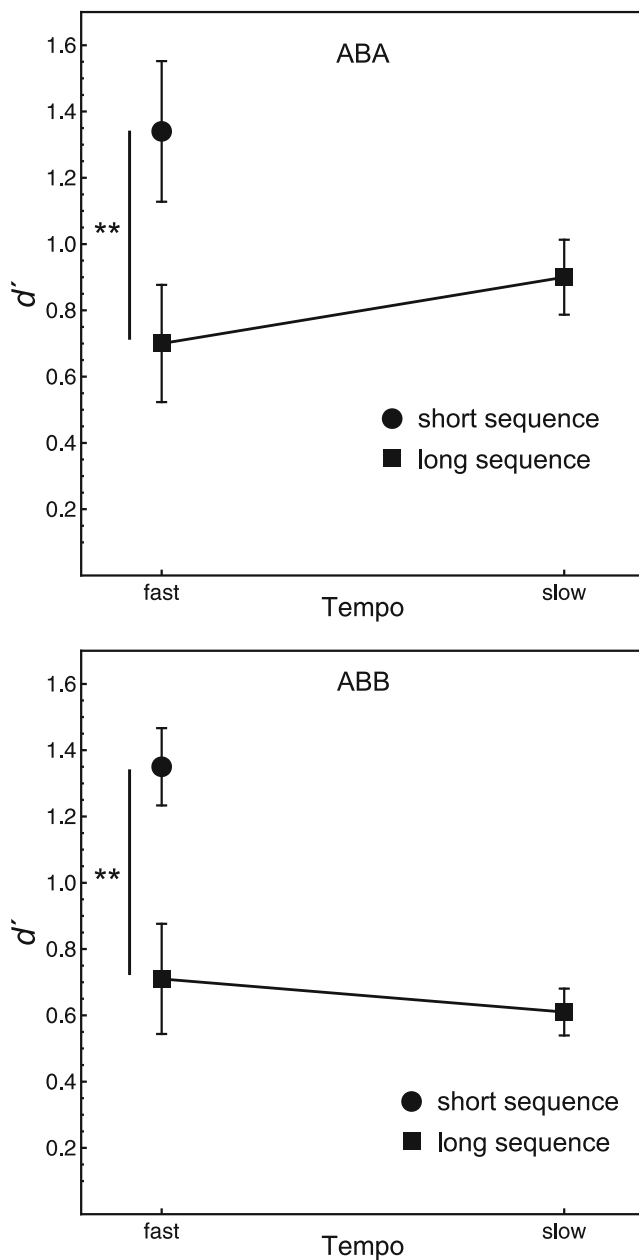
Fig. 5 Mean sensitivity (AUC transformed to *d'*) in the temporal shift discrimination task, as a function of rhythm, tempo, and sequence duration. Upper panel: ABA rhythm. Lower panel: ABB rhythm. Squares indicate long sequences, and circles indicate short sequences. Error bars show ±1 *SEM* of the eight individual values. Lines marked by asterisks denote significant pairwise differences

Fig. 6 ABA rhythm. Mean normalized decision weights for the three IOIs involving the target, as a function of tempo and sequence duration. Upper panel: Fast tempo. Lower panel: Slow tempo. Squares indicate long sequences; circles are short sequences; and open triangles are ideal decision weights in the absence of stream segregation, derived from the individual gap discrimination difference limens (see the text). Error bars show 95% confidence intervals

Decision weights

ABA rhythm Figure 6 shows the mean normalized decision weights for the ABA rhythm. It also depicts the average optimal decision weights estimated individually on the basis of the GDDLs measured during the experiment, and assuming the absence of stream segregation (i.e., no difficulty in using information from between-stream IOIs).
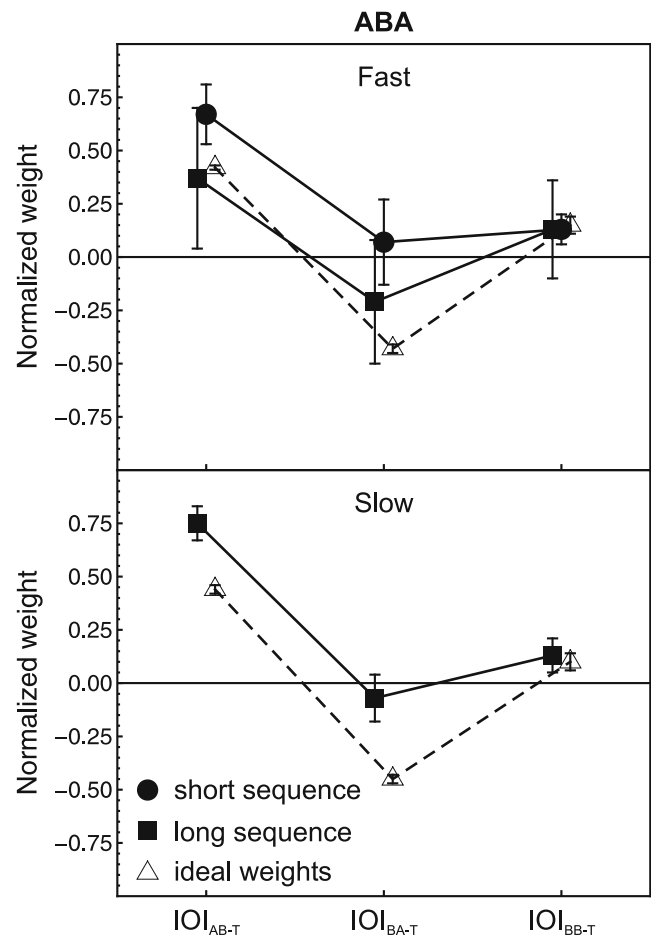
First, the decision weights for the fast ABA sequences (Fig. 6, upper panel) were analyzed. As expected, for the fast *short* sequences (circles in Fig. 6), which were predominantly perceived as integrated according to the subjective ratings (Fig. 4), the between-stream interval IOI_{AB-T} dominated the decision. As the confidence intervals (CIs) in Fig. 6 show, across subjects, the decision weights were also significantly different from 0 for the within-stream IOI_{BB-T} (six of the eight individual weights were significantly greater than 0, as revealed by the Wald CIs of the ML estimates). Thus, listeners also used within-stream information, and the observed weight was close to the ideal weight for IOI_{BB-T}. The between-stream interval IOI_{BA-T}, on average, did not receive a significant weight. This is surprising, because the fast short ABA sequences were mostly perceived as integrated, and the ideal weight indicates that IOI_{BA-T} would provide as reliable information as IOI_{AB-T} when streaming was assumed to be absent. A speculative explanation for the low weight on IOI_{BA-T} is

that judgments of temporal shift may be biased toward positive rather than negative decision weights. The weight on IOI_{AB-T} was higher than would have been optimal, and the weight on IOI_{BA-T} was much too small. Inspection of the individual weights revealed a considerable interindividual variability of the weight on IOI_{BA-T}. For example, one listener assigned a significantly negative weight to IOI_{BA-T}, whereas another listener assigned a significant positive weight to IOI_{BA-T}, which produces a systematic bias toward the incorrect response.

For the fast *long* sequences, which the subjects often reported to be perceived as segregated, the decision weight on IOI_{AB-T} was lower than for the short sequences, but still significantly greater than 0 (six of the eight individual weights were significantly greater than 0, whereas two listeners assigned nonsignificant negative weights). The weight on IOI_{BB-T} was identical in magnitude to the weight observed for the fast short sequences, but it was not significantly different from 0 (only three of the eight individual estimates were significantly different from 0, and one of these three weights was negative). This finding is incompatible with the hypothesis that listeners should rely mainly on *within-stream information* if they perceive the ABA rhythm as two separate streams. The weight on IOI_{BA-T} was negative for the fast long sequences, as would be the ideal weight, although across listeners the weight was not significantly different from 0. For four listeners, the weight on IOI_{BA-T} was close to the ideal value and was significant, for three listeners it was close to 0, and one listener assigned a significant positive weight. If one compares the observed weights to the ideal weights estimated under the assumption of no streaming, it is surprising that the observed weights were closer to the ideal weights for the long (segregated) than for the short (integrated) sequences.

The decision weights for the fast sequences were analyzed with a repeated measures ANOVA using a univariate approach and Huynh–Feldt correction to the degrees of freedom. The within-subjects factors were IOI (IOI_{AB-T}, IOI_{BA-T}, IOI_{BB-T}) and Sequence Duration. The effect of IOI was significant, $F(2, 14) = 12.95$, $p = .002$, $\tilde{\varepsilon} = .74$, $\eta^2_p = .65$, confirming the descriptive differences between the weights assigned to the three IOIs. Even more importantly, the IOI × Sequence Duration interaction was also significant, $F(2, 14) = 4.79$, $p = .028$, $\tilde{\varepsilon} = .95$, $\eta^2_p = .41$. Thus, the patterns of weights differed significantly between the two sequence durations.

For the *slow* long ABA sequence (Fig. 6, lower panel), the between-stream IOI_{AB-T} again received the highest weight, as would be expected for these sequences perceived predominantly as integrated. The average decision weight for the within-stream IOI_{BB-T} was also significantly higher than 0. Thus, listeners also used within-stream information. The weight assigned to IOI_{BA-T} was again nonsignificant (only two individual weights were significantly different from 0). Thus, the pattern of weights was generally similar to the

weights observed for the fast short sequences, which were also predominantly perceived as integrated.

An ANOVA analyzing the weights for the long sequences showed a significant effect of IOI, $F(2, 14) = 31.49$, $p < .001$, $\tilde{\varepsilon} = .96$, $\eta^2_p = .82$. Importantly, a marginally significant IOI × Tempo interaction emerged, $F(2, 14) = 3.37$, $p = .075$, $\tilde{\varepsilon} = .85$, $\eta^2_p = .33$: The weight on IOI_{AB-T} was higher for the slow long (i.e., integrated) than for the fast long (i.e., segregated) sequences. Thus, the decision weights indicate at least a tendency toward an effect of tempo that is also clearly evident in the subjective ratings (Fig. 4), but not in the analysis of sensitivity (Fig. 5).

As was mentioned above, the patterns of weights were rather similar for the two ABA sequences perceived as integrated on most trials. Therefore, as a post-hoc analysis, the weights assigned to the three different IOIs were compared between the fast short and slow long sequences (integrated), on the one hand, and the fast long sequence (segregated), on the other. For each listener and each of the three IOIs, the average weights in the slow long and fast short condition were computed. The resulting weights were contrasted with the weights for the fast long sequence. A repeated measures ANOVA with the factors IOI and Sequence Type (integrated vs. segregated) showed a significant IOI × Sequence Type interaction, $F(2, 14) = 3.88$, $p = .046$, $\tilde{\varepsilon} = 1.0$, $\eta^2_p = .36$.

Taken together, whereas the patterns of decision weights showed some significant differences between sequences predominantly perceived as integrated versus segregated, the decision weights for the ABA rhythm showed several important deviations from the often voiced hypothesis that listeners make near-exclusive use of within-stream IOIs when the sequences are perceived as segregated. On average, the decision weights on the within-stream intervals IOI_{BB-T} did not differ between the fast long sequence (often perceived as segregated), on the one hand, and the fast short and slow long sequences (predominantly perceived as integrated), on the other. The weights assigned to the between-stream IOI_{AB-T} and IOI_{BA-T} were even closer to the ideal weights, assuming absence of stream segregation, when the sequence was predominantly perceived as segregated. Additionally, inspection of the individual weights showed that some listeners used a qualitatively different decision strategy that systematically resulted in incorrect responses, assigning, for example, a positive weight to IOI_{BA-T}. Interindividual differences in decision strategies were also reported in previous studies (Lutfi & Liu, 2011; Oberfeld, 2009).

ABB rhythm Figure 7 shows the mean normalized decision weights for the ABB rhythm, together with the ideal weights assuming the absence of stream segregation. Recall that the within-stream interval IOI_{BB+1-T} was rather short and identical in duration to the between-stream IOI_{AB-T}. Thus, the listeners were expected to use information from this IOI in all
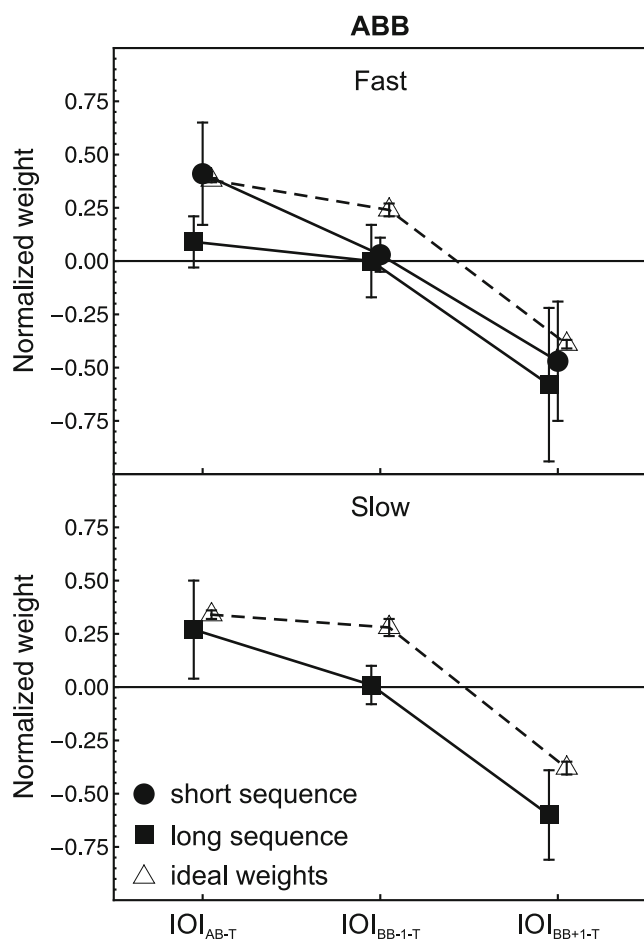
ABB

Fig. 7 ABB rhythm. Mean normalized decision weights for the three IOIs involving the target, as a function of tempo and sequence duration. Squares are fast long sequences; circles are fast short sequences; and open triangles are slow long sequences. Error bars show 95% confidence intervals

conditions, and that is exactly what the decision weights show. In fact, the weight assigned to IOI_{BB+1-T} was even higher than optimal. The within-stream interval preceding the target (IOI_{BB-1-T}) was twice as long as IOI_{BB+1-T}, and the ideal weight was lower in magnitude than for IOI_{BB+1-T} (see the triangles in Fig. 7). In fact, this IOI was virtually ignored in the decision. As expected, the decision weights for the between-stream IOI showed descriptive differences between the conditions. For the fast sequences, the weight on IOI_{AB-T} was significantly different from 0 for the short, but not for the long, sequence duration, which is exactly the expected pattern (i.e., between-stream information was not used in the condition in which the sequences were often perceived as segregated). However, a repeated measures ANOVA for the fast sequences showed a significant effect of IOI, $F(2, 14) = 26.67$, $p < .001$, $\tilde{\varepsilon} = .82$, $\eta_p^2 = .79$, confirming the differences in weights between the three IOIs, but no significant IOI × Sequence Duration interaction, $F(2, 14) = 2.49$, $p = .13$.

For the slow long ABB sequence, the CIs in the lower panel of Fig. 7 show that listeners used information from both within- and between-stream IOIs, as was the case for the ABA rhythm. Descriptively, the weight assigned to the between-stream interval IOI_{AB-T} was higher for the slow long than for the fast long sequence, which is compatible with the expected effect of stream segregation. On the other hand, the patterns of weights were rather similar between the slow long and fast short ABB sequences, which were both predominantly perceived as integrated. However, a repeated measures ANOVA showed no significant difference between the patterns of weights in the fast long and slow long conditions; the IOI × Tempo interaction was not significant, $F(2, 14) = 1.26$, $p = .31$.

Taken together, the observed decision weights deviated from the expected patterns in several ways. First, for the ABA rhythm, the data did not correspond to the idealized pattern of zero weights assigned to between-stream intervals in the segregated case often implicitly assumed for the temporal shift discrimination task. Second, for the ABB rhythm, clear differences in subjective organization (e.g., fast vs. slow long sequences) did not correspond to significantly different decision weights. Third, in the integrated cases, in which the listeners should have been able to use both within- and between-stream information, the observed weights deviated from the ideal weights derived from the GDDLs, although the latter weights would have maximized accuracy in this case. Finally, inspection of the individual data revealed considerable interindividual differences in the patterns of weights.

Efficiency measures: Disentangling nonoptimal decision weights and increases in internal noise

The above analyses showed that it is difficult to infer the perceptual organization of a sequence from the decision weights alone or from sensitivity alone. As was suggested in the introduction, this limitation might be overcome by combining the information gained from sensitivity and the decision weights (Berg, 1990). Three different efficiency measures were computed to quantify the loss in sensitivity in comparison to a reference sensitivity representing the optimal performance.

In the analysis, the upper reference for sensitivity was d'_{GDDL}, which denotes the sensitivity in the temporal shift discrimination task that an observer using the optimal set of decision weights and limited only by the finite sensitivity in judging the duration of a temporal interval would obtain. More specifically, as in the computation of the ideal weights above, d'_{GDDL} was determined under the assumption that the sensitivities for judging the intervals were equivalent to the measured GDDLs. Therefore, d'_{GDDL} represents the performance of a subject applying the optimal decision weights in the absence of stream segregation. This reference sensitivity d'_{GDDL} could in principle be computed analytically by

determining the expected value of the decision variable on $D(\mathbf{IOI})$ for trials with early and late target onsets, and then dividing the difference between those two values by the standard deviation of this difference (Green & Swets, 1966). However, as was noted above, finding an analytic solution was somewhat complicated for our experiment, due to the correlations between the predictors (IOIs). Therefore, a Monte Carlo method was applied for computing d'_{GDDL}, following the same rationale as for the simulations used for estimating the ideal weights (Eq. 3). In all, 5,000 trials were simulated with late target onsets and 5,000 trials with early target onsets, and for each trial the value on the internal continuum (i.e., the perceived IOI duration, which includes the external noise) was computed as the actual IOI duration plus a random variate (internal noise), with the standard deviation of the internal noise being selected to match the GDDLs measured for the respective type of IOI for a given listener. To estimate d'_{GDDL}, the individual ideal weights (see above) were used to compute the value of the decision variable $D(\mathbf{IOI})$ according to Eq. 1. The difference between the average value of $D(\mathbf{IOI})$ on trials with delayed target onsets (μ_{Late}) and on trials with early target onsets (μ_{Early}), divided by the common standard deviation ($\sigma = SD[D(\mathbf{IOI})]$) is, by definition, $d'_{\text{GDDL}} = (\mu_{\text{Late}} - \mu_{\text{Early}})/\sigma$. Note that our definition of d'_{GDDL} includes internal noise corresponding to the less-than-perfect gap duration discrimination performance for isolated temporal intervals (i.e., the measured GDDLs), rather than assuming an ideal observer without internal noise as, for example, in Berg (2004). Thus, d'_{GDDL} represents the sensitivity that a given listener could obtain if (a) his or her representation of the decision-relevant IOIs in the sequence was as precise as for the isolated temporal intervals, and (b)he or she applied the optimal decision weights. A similar type of analysis was used by Alexander and Lutfi (2008) and Oberfeld, Kuta, and Jesteadt (2013).

The molar (d') and molecular (decision weights) measures can now be combined in order to investigate which factor(s) affected the performance in the temporal shift discrimination task (Berg, 1990, 2004). First, imagine that for an ABA sequence the information about the durations of the three relevant IOIs available at the decision stage was as precise as if these IOIs had been presented in isolation. In other words, combining the IOIs into the ABA sequence did not increase the internal noise. Which sensitivity would now result if the listener applied the empirically estimated decision weights for the sequence, rather than the ideal weights? This sensitivity (d'_{wgt}) can be estimated using the same simulation method as above, but this time using the empirically observed rather than the ideal weights when computing the decision variable for the given listener and sequence. The weighting efficiency $\eta_{\text{wgt}} = (d'_{\text{wgt}}/d'_{\text{GDDL}})^2$ represents the loss in sensitivity caused by the suboptimal decision weights (Berg, 1990).

If the assertion were true that streaming caused only a change in the decision weights, but not an increase in internal noise, then d'_{wgt} should be equal to the observed sensitivity (d'_{obs}). In this case, the efficiency measure $\eta_{\text{noise}} = (d'_{\text{obs}}/d'_{\text{wgt}})^2$, representing an additional loss in efficiency due to increased internal noise (i.e., other factors besides applying suboptimal decision weights), should be 1.0 (Berg, 1990). If, however, stream segregation caused an increase in internal noise (i.e., made it difficult to use information from between-stream intervals), this would be indicated by values of η_{noise} smaller than 1.0. Finally, $\eta = (d'_{\text{obs}}/d'_{\text{GDDL}})^2 = \eta_{\text{wgt}} \eta_{\text{noise}}$ represents the overall loss in efficiency due to both factors. It is important to note that Berg (1989) showed that the relative decision weight estimates are unaffected by additive internal noise, which is of course a prerequisite for this analysis (cf. Oberfeld et al., 2013).

Using the two efficiency measures, η_{wgt} and η_{noise}, it was possible to analyze whether a higher probability of stream segregation (e.g., for a fast long as compared to a fast short sequence) resulted only in the adoption of suboptimal weights, but no increase in internal noise ($d'_{\text{wgt}} < d'_{\text{GDDL}}$ but $d'_{\text{obs}} = d'_{\text{wgt}}$; thus, $\eta_{\text{wgt}} < 1$ and $\eta_{\text{noise}} = 1$), or also in higher internal noise, as is assumed in the literature ($d'_{\text{obs}} < d'_{\text{wgt}}$, thus $\eta_{\text{noise}} < 1$). For one subject, d'_{wgt} and d'_{obs} were negative for the fast long ABB sequence, because she had assigned a negative weight to $\text{IOI}_{\text{BB}-1-\text{T}}$, systematically resulting in incorrect responses. This subject was excluded from the efficiency analyses.

The mean efficiency is displayed in Fig. 8. First, the ABA rhythm (upper panel) is discussed. The measure for total efficiency, η, scored considerably below 1.0 for each sequence type (Tempo × Duration), indicating that the observed sensitivity was smaller than was predicted from the GDDLs and the ideal decision weights derived from them. A repeated measures ANOVA showed a significant effect of sequence type (fast long, fast short, and slow long), $F(2, 14) = 9.37$, $p = .005$, $\tilde{\varepsilon} = .83$, $\eta_{\text{p}}^2 = .57$. The overall sensitivity was lowest for the fast long and highest for the slow long sequence. Pairwise comparisons between the three levels of sequence type were computed by means of separate paired-samples t-tests (nonpooled error terms; Keselman, 1994) and using Hochberg's (1988) sequentially acceptive step-up Bonferroni procedure, which controls the family-wise Type I error rate. At an α level of .05, the difference between η in the fast long and the fast short as well as the slow long sequences was significant.

The weighting efficiency η_{wgt} was also smaller than 1.0 for all sequence types, showing that even in the sequences predominantly perceived as integrated, the listeners did not apply the optimum weights, as was discussed above in the section on Decision Weights. However, η_{wgt} did not differ significantly between the three sequence types, $F(2, 14) = 1.51$, $p = .26$.
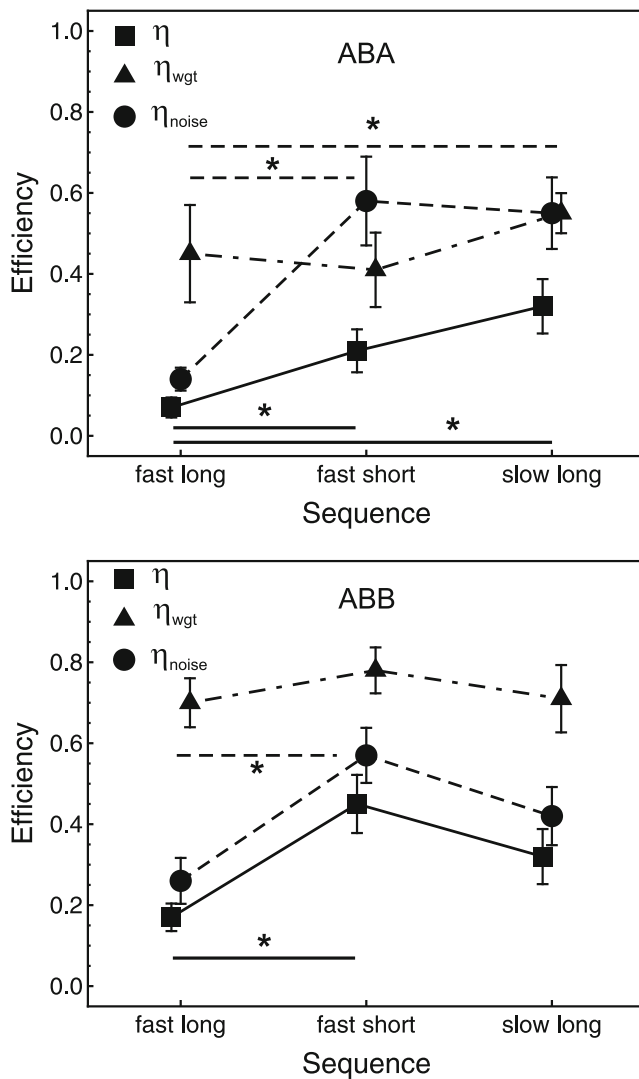
Fig. 8 Mean efficiencies for the three sequence conditions. Upper panel: ABA rhythm. Lower panel: ABB rhythm. Squares are η, triangles are η_{wgt}; circles are η_{noise}. Lines marked by asterisks denote significant pairwise differences. Error bars show ±1 *SEM*

In contrast, sequence type had a significant effect on η_{noise}, $F(2, 14) = 7.45$, $p = .012$, $\tilde{\varepsilon} = .78$, $\eta_p^2 = .52$. Pairwise comparisons indicated that η_{noise} was significantly lower for the fast long than for the fast short and slow long conditions. This pattern is compatible with the assumption that stream segregation—experienced most frequently in the fast long sequences—impaired the representation of the IOI durations (i.e., caused higher internal noise).

Taken together, for the ABA rhythm it can be concluded that the weighting efficiency was not significantly influenced by streaming. However, sequential stream segregation caused an increase in internal noise, and the measure η_{noise} clearly differentiated between the fast long sequence (predominantly perceived as segregated) and the other two types of sequence, which were perceived as integrated on the majority of trials (see Fig. 4). The overall efficiency η also differed between the

"integrated" and "segregated" sequences, but showed an additional nonsignificant difference between the fast short and the slow long sequence.

For the ABB rhythm, the analyses of the three efficiency measures showed patterns of effects similar to those for the ABA rhythm. A significant effect of sequence type on η and η_{noise} emerged, but not on η_{wgt} (for η, $F(2, 12) = 8.92$, $p = .004$, $\tilde{\varepsilon} = 1.0$, $\eta_p^2 = .60$; for η_{noise}, $F(2, 12) = 6.03$, $p = .015$, $\tilde{\varepsilon} = 1.0$, $\eta_p^2 = .50$; for η_{wgt}, $F(2, 12) = 2.20$, $p = .15$). For η and η_{noise} only, the pairwise comparison between the fast long and fast short sequences was significant, although descriptively the two measures were smaller for the fast long ("segregated") than for the slow long ("integrated") sequence.

It can thus be concluded that efficiency measures based on a combination of molar and molecular estimates reflect differences between integrated and segregated sequences that sensitivity measures like d' fail to show. The explanation for this finding is that efficiency, as computed here, corrects for the opposite effects that changing the tempo of a sequence should have on sensitivity. As was discussed above, the longer IOI durations in a slow sequence make the temporal shift discrimination task more difficult because the GDDLs are higher than for a fast sequence. This general difference in sensitivity between the two tempi is reflected in the reference sensitivity d'_{GDDL}. On the other hand, using the between-stream IOIs should be easier in a slow than in a fast sequence, because the latter is more frequently perceived as segregated, and our analyses indeed provide evidence for increased internal noise in the segregated case.

The above analyses showed that in the mean data, η and η_{noise} differed between sequences predominantly perceived as integrated and those predominantly perceived as segregated. Does this relation apply at the individual level? More precisely, was the probability of perceiving a sequence as integrated, P(one stream), correlated with any of the efficiency measures? To answer this question, the data were analyzed using random-effects models with a random intercept and slope, taking into account the repeated measures structure of the data. The variance–covariance matrix of the random effects was specified as being of the "unstructured" (UN) type, corresponding to a random coefficient model (Wolfinger, 1996). The degrees of freedom were computed according to the method by Kenward and Roger (1997). For the regression of η_{noise} on P(one stream), the population estimates of the slope of the regression line were $\beta_{ABA} = 0.42$ ($SE = 0.24$, two-tailed $p = .13$) and $\beta_{ABB} = 0.25$ ($SE = 0.10$, $p = .033$) for the ABA and ABB rhythm, respectively. The nonsignificant regression coefficient for the ABA rhythm seemed to be due to a single outlying value of η_{noise} in the vicinity of 1.0. Excluding this data point (1 out of 23) from the analysis resulted in a significant regression coefficient, $\beta_{ABA} = 0.58$, $SE = 0.15$, $p = .002$. These results show that when a sequence was perceived as integrated on a high proportion of trials, then the estimated

value of η_{noise} also tended to be high. The coefficient of determination computed according to Edwards, Muller, Wolfinger, Qaqish, and Schabenberger (2008) was higher for the ABA rhythm ($R^2_\beta = .51$) than for the ABA rhythm ($R^2_\beta = .37$). The weighting efficiency η_{wgt} was not systematically related to P(one stream): For both rhythms, the regression coefficient was not significantly different from 0 (p values > .17). These results confirm that the increase in internal noise quantified by η_{noise} represents a performance-based measure of stream segregation, whereas the pattern of decision weights indexed by η_{wgt} cannot be used to differentiate between sequences perceived as integrated or segregated. Because the total efficiency $\eta = \eta_{wgt} \eta_{noise}$ encompasses η_{noise}, η was also correlated with P(one stream). The population estimates of the slope of the regression line were $\beta_{ABA} = 0.36$ ($SE = 0.10$, $p = .004$, $R^2_\beta = .48$) and $\beta_{ABB} = 0.23$ ($SE = 0.08$, $p = .029$, $R^2_\beta = .48$) for the ABA and ABB rhythm, respectively. Without the outlying value of η in the ABA condition, β_{ABA} was 0.34 ($SE = 0.11$, $p = .007$).

Discussion

Here, auditory stream segregation was studied in ABA and ABB sequences, using a combination of subjective ratings, sensitivity in a temporal shift discrimination task, decision weights in the latter task, and efficiency measures. This combination of different methods and measures provided an unprecedented "microscopic" look on the effects of stream segregation on performance in the shift discrimination task, and revealed several interesting issues.

As expected, the sequence tempo and the sequence duration had a strong effect on the perceptual organization of the sequences (integrated versus segregated). Although the sensitivities in the shift discrimination task differed between short and long fast sequences, reflecting the difference in perceived organization, the sensitivity did not differ between fast and slow sequences, despite clearly differing perceptual organizations. This dissociation between the perception as integrated and segregated on the one hand and sensitivity in the shift discrimination task on the other hand was expected because a change in sequence tempo alters the sensitivity for temporal gap discrimination (e.g., Friberg & Sundberg, 1995), which forms the basis of performance in the shift discrimination task. Thus, as expected our data demonstrate a serious limitation of sensitivity-based measures of stream segregation because the fundamental effect of presentation rate on the perceptual organization (van Noorden, 1975) is not reflect by these measures.

Concerning the relation between the patterns of decision weights and the perceptual organization, our data revealed pronounced deviations from the idealized decision strategies often explicitly or implicitly assumed for the shift discrimination task. It is typically presumed that it is difficult or impossible to use between-stream information if a sequence is perceived as segregated. Therefore, the decision weights, which represent a direct measure for the use of different sources of information in the shift discrimination task, should be near zero for between-stream IOIs in the segregated case. However, in the ABA sequences were studied, one of the between-stream IOIs received a significant weight in the fast long sequences that were predominantly perceived as segregated (see Fig. 6). Additionally, in sequences predominantly perceived as integrated the decision weights differed from the optimum weights that would have maximized the accuracy in the absence of stream segregation. It can therefore be concluded that listeners do not always apply the idealized pattern of decisions weights typically assumed when using sensitivity-based measures for stream segregation. As was discussed earlier, similar results have been reported in a recent study by Richards, Carreira, and Shen (2012), in which listeners detected a temporal shift on one A tone in an ABAB sequence, and the frequency difference between A and B tones was varied. Using likelihood-ratio tests, the authors compared the goodness of fit of a model containing as predictors the onset and offset times of both the A and the B tones (full model) and the goodness of fit of a model containing as predictors only the A-tone on- and offsets (restricted model). Although model comparisons indicated that for Δfs of 17 semitones or greater information about the onsets of the tones belonging to the stream not containing the target did not have a substantial influence on the decision for the majority of listeners, information from the nontarget stream received a significant weight in several cases. Additionally, the likelihood-ratio test favors models with fewer degrees of freedom—that is, fewer predictors (Agresti, 2002). Therefore, even if the test indicates that the full model did not provide a significantly better fit than the restricted model, this does not necessarily show that the B tones were absolutely unimportant for the decision. On a more general level, the results of the present study indicate that the decision weights alone cannot be used for differentiating between sequences perceived as integrated or segregated.

The present study took the molecular psychophysics approach one step further by combining decision weights and estimates of sensitivity. One of the most important features of the present study is that GDDLs were also measured for isolated pairs of tones representing the different IOIs in the ABA and ABB sequences assumed to be utilized in the temporal shift discrimination task. This made it possible to estimate ideal decision weights in the absence of stream segregation, on an individual basis. The GDDLs also played a crucial role in the computation of efficiency measures that were used to disentangle effects of stream segregation on internal noise and the weighting patterns. These analyses were based on the concept that two factors can contribute to limitations in sensitivity (Swets et al., 1959). Stream segregation

could reduce the precision of the information about the between-stream IOIs available at the decision stage, as it is assumed when using sensitivity-based measures of stream segregation. An impairment in performance could also be due to nonoptimal decision strategies—that is, suboptimal decision weights. Equally important, the applied efficiency measures put the observed sensitivity into the context of the temporal resolution underlying performance in the shift discrimination task. The reference sensitivity was computed on the basis of the individually measured GDDLs and therefore reflected the fact that even in the absence of streaming the sensitivity differs between a fast and a slow sequence, as discussed above. In fact, our analyses showed that two of the efficiency measures were capable of discriminating between sequences perceived as integrated and those perceived as segregated. In particular, the measure η_{noise} that estimates a loss in sensitivity attributable to internal noise rather than a suboptimal decision strategy was significantly lower for the fast long sequences (segregated) than for the fast short or slow long sequences (integrated). Additionally, regression analyses showed that η_{noise} is systematically related to the probability of perceiving a sequence as integrated. Therefore, the conclusion is that η_{noise} is a useful performance-based measure of stream segregation that avoids some of the limitations of using sensitivity alone or decision weights alone. On a more general level, the difference in η_{noise} between segregated and integrated sequences is compatible with the assumption that stream segregation causes an increase in internal noise in the sense that it renders between-stream information less precise or even unusable. This assumption is also the basis of sensitivity-based measures of streaming. In principle, the difference in η_{noise} between segregated and integrated sequences might also be caused by less precise within-stream information in the segregated as compared to the integrated case. It remains for future research to show whether it is possible to obtain separate estimates of the internal noise associated with within-stream and between-stream information. These separate estimates would make it possible to quantify which proportions of the change in η_{noise} can be attributed to less precise within-stream or between-stream information. It should also be noted that in the individual data the ordering of the probabilities of perceiving a sequence as integrated was not always reflected by the ordering of the values of η_{noise}. Additional experiments are needed to show whether this is merely due to imprecise estimates caused by the rather small number of trials used, or whether additional factors need to be considered here.

In the preceding discussion, the internal noise components associated with each IOI were conceived as being "early" or "sensory"; that is, they appear prior to integration (cf. Eq. 3). An increase in an additional "central" internal noise source located at or after integration (cf. Durlach, Braida, & Ito, 1986) might also have contributed to the observed difference in η_{noise} between integrated and segregated sequences. A "central" noise source would equally affect the information available from the within-stream and the between-stream IOIs. More specifically, if the central noise dominated the "sensory" noise, then the listeners should assign approximately uniform weights to all IOIs, regardless of the internal noise SDs (indexed for example by the GDDLs) effective for the different IOIs (see Oberfeld et al., 2013, for a detailed discussion). Thus, the observed differences in the decision weights render it unlikely that central noise dominates the sensory noise in the shift-discrimination task. However, as discussed above, it remains for future research to obtain separate estimates of the different potential internal noise components.

The present analysis was restricted to the three IOIs most adjacent to the target. It would be interesting to investigate the extent to which additional IOIs contribute to the decision in future studies. This can be achieved simply by adding additional IOIs to Eq. 1. Such analyses may provide insight into questions like whether listeners consider IOIs from earlier parts of the sequence, using for example all previous A-B intervals to form a representation of the average duration of this interval and compare it to the IOI_{AB-T} in the target triplet. If the data show that this was the case, and if the weights assigned to previous triplets are higher for the longer sequences, then the extended model including the additional IOIs could be used to answer the question whether the estimated increase in internal noise in the present analyses was partly due to the inclusion of only the three IOIs from the target triplet.

Another interesting generalization of the methodological approach developed here could be to incorporate potential nonlinear transformations of the IOIs. To this end, the IOI_i terms in Eq. 1 can be replaced by $f(IOI_i)$, where $f()$ is a monotonic function (cf. Richards, 2002). Whether this will provide a better fit of the data than the linear model used in this article is an empirical question, but does not render the general approach invalid.

Importantly, the method here is not restricted to the temporal shift discrimination task used as an example in the present experiment, but can be applied to any task in which the performance can be assumed to depend on the perceptual organization of the sequence—as, for example, frequency discrimination (Ma et al., 2010). The methodology outlined in this article should also be useful for other

domains than auditory stream segregation. For example, in an experiment on object-based visual attention (Chen, 2012; Kahneman & Henik, 1981), the methods described in this article could be used to quantify the amount of selective attention to the target in different conditions, by measuring the decision weights assigned to target and distractor elements. In addition, a potential increase in internal noise caused by the distractor elements could be identified and quantified by the efficiency analyses.

In summary, using exactly the same task as in an experiment aimed at a sensitivity-based objective measure of auditory streaming, a rich set of behavioral measures of streaming can be obtained, overcoming limitations of measures based only on sensitivity. In terms of the experimental method, it is only necessary to impose random variations on the IOIs, and to additionally measure the sensitivity for the stimulus elements underlying the performance in the studied task, as for example gap duration discrimination limens for the different IOIs constituting an ABA sequence. This will make it possible to estimate (a) sensitivity, (b) decision weights, and (c) a measure of internal noise, using standard data analysis techniques like logistic regression. It should be noted, however, that in order to obtain precise estimates of the decision weights on an individual level, the proposed method requires a somewhat higher number of trials than does an experiment measuring only the sensitivity. The multimeasure molecular psychophysics approach applied in the present study offered a detailed insight into effects of stream segregation on the performance in a temporal shift discrimination task. In particular, the method provides an observer efficiency measure indexing the increase of internal noise, as compared to a situation without stream segregation. Unlike estimates of sensitivity, this measure was able to dissociate fast from slow ABA and ABB sequences. These sequences clearly differed in their perceptual organization (one vs. two streams, as revealed by subjective ratings), but not in terms of sensitivity in the temporal shift discrimination task, demonstrating the limitations of sensitivity as an objective measure of streaming.

Appendix: Pretest

The pretest was completed before the main experiment started. The purpose of this experiment was to select appropriate parameter values for the temporal shift (Δt) of the target tone.

Pretest: Selection of a value of Δt for the temporal shift discrimination task

For the random variation in the tone onsets to have an effect on the decision, the temporal shift should not be too easy to detect. If Δt were so large that subjects always detect the shift (i.e., 100% correct responses are obtained), then the random timing perturbations would have no effect at all on the decision, and an estimation of decision weights would not be possible. Although, as Berg (2004) noted, the optimal level of performance for estimating decision weights has yet to be identified, previous studies have successfully used $P(C)$ around .7 (e.g., Oberfeld & Plank, 2011).

Method

In the pretest, seven listeners (six female, one male; 19–28 years of age) participated voluntarily. All reported normal hearing. The same temporal shift discrimination task as in the main experiment was presented, except that no no-shift trials were presented, and no ratings of the perceptual organization were required. Listeners indicated their confidence of "target early" or "target late" on the four-point rating scale also used in the main experiment. After one practice session, all listeners participated in six experimental sessions. Only one rhythm (ABA or ABB) and one value of Δt (15, 30, or 49 ms) was presented per session, and one block of 110 trials for each of the three combinations of tempo and sequence duration was obtained in each session. The early and the late temporal position of the target tone were presented on 55 trials each (in random order). The sensitivity was estimated in terms of AUC converted to d', as in the main experiment (see the section on "Sensitivity in the temporal shift discrimination task").

Results

To analyze the relation between Δt and sensitivity, the average d' was computed per listener, condition, and value of Δt. Figure A1 shows the average d' across listeners as a function of Δt, for the different conditions. The lines are the best-fitting linear population regression lines, which were estimated using a subject-specific, random-effects model approach (SAS PROC MIXED; cf. Burton, Gurrin, & Sly, 1998; Liang & Zeger, 1993) accounting for the repeated measures structure of
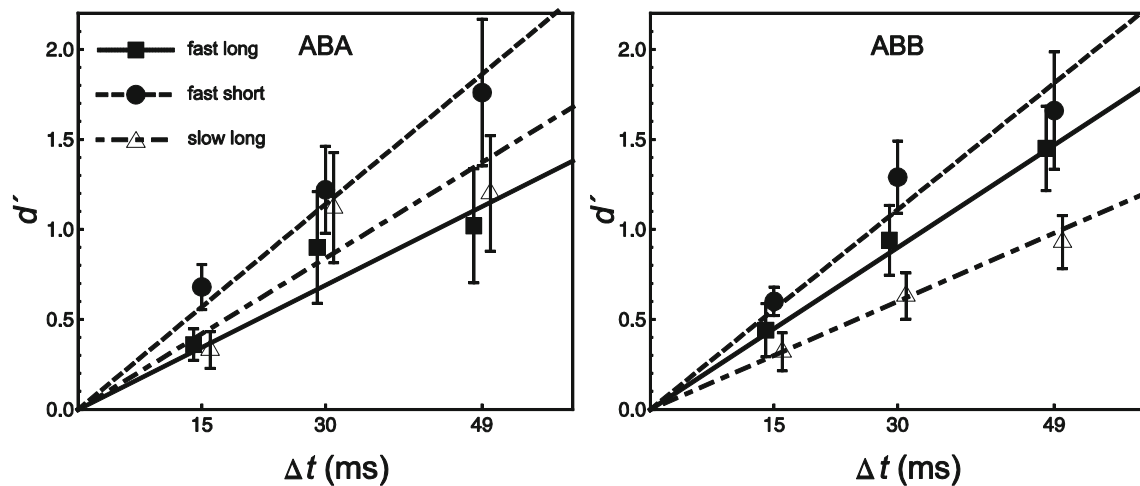
Fig. 9 Pretest: Mean sensitivity (AUC transformed to *d*′) in the temporal shift discrimination task, as a function of Δ*t*, rhythm, tempo, and sequence duration. Lines indicate the estimated population regression lines through the origin. Squares and solid lines show the fast-tempo, long sequence.

Circles and dashed lines show the fast-tempo, short sequence. Triangles and dot-dashed lines show the slow-tempo, long sequence. Error bars show ±1 standard error of the mean (*SEM*) of the seven individual values

the data (each subject received three different values of Δ*t*). In this approach, the slope is assumed to vary from subject to subject, and the correlation structure is modeled by treating the subjects as a random sample from a population of all subjects. Linear regression through the origin was used because, by definition, *d*′ = 0 for Δ*t* = 0. The variance–covariance matrix was specified as being of the "unstructured" (UN) type; that is, no restrictions were imposed on the variances and covariances (Wolfinger, 1993). The degrees of freedom were computed according to the method by Kenward and Roger (1997), which performs favorably relative to other methods (e.g., Fouladi & Shieh, 2004).

The population estimates of the slopes in the different conditions are displayed in Table A1. Because the slopes did not differ too extremely between conditions, a common value of Δ*t* = 26 ms was selected for all conditions of the main experiment, so that the average expected AUC was .70, corresponding to an average expected *d*′ = 0.76. Across

conditions, the expected AUC ranged between .64 and .76. In the main experiment, the same value of Δ*t* = 26 ms was used for all subjects and all experimental conditions. Using a fixed value of Δ*t* has the advantage that the sensitivities can be compared directly between conditions.

Table 1 Pretest: Estimated population regression slopes for the relation between Δ*t* (in milliseconds) and *d*′ in the temporal shift discrimination task

Rhythm	Tempo	Duration	Slope	SE
ABA	Fast	Long	0.023[*]	0.0074
		Short	0.038[*]	0.0080
	Slow	Long	0.028[*]	0.0073
ABB	Fast	Long	0.030[*]	0.0052
		Short	0.037[*]	0.0063
	Slow	Long	0.020[*]	0.0033

Linear regression through the origin, subject-specific random-effects model. [*] Estimated population slope significantly different from 0, *p* < .05, two-tailed.

References

Agresti, A. (1989). Tutorial on modeling ordered categorical response data. *Psychological Bulletin, 105,* 290–301.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: Wiley.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*(1), 8. doi:10.1167/2.1.8. 121–131.

Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49,* 1751–1756.

Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *Journal of the Acoustical Society of America, 116,* 2234–2247. doi:10.1121/1.1784437

Alexander, J. M., & Lutfi, R. A. (2008). Sample discrimination of frequency by hearing-impaired and normal-hearing listeners. *Journal of the Acoustical Society of America, 123,* 241–253. doi: 10.1121/1.2816415

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology. Human Perception and Performance, 11,* 257–271. doi:10.1037/0096-1523.11.3.257

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America, 99,* 2270–2280.

Bell, A. J., & Sejnowski, T. J. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation, 7,* 1129–1159. doi:10.1162/neco.1995.7.6.1129

Berg, B. G. (1989). Analysis of weights in multiple observation tasks. *Journal of the Acoustical Society of America, 86,* 1743–1746.

Berg, B. G. (1990). Observer efficiency and weights in a multiple obser-
vation task. *Journal of the Acoustical Society of America, 88,* 149–
158. doi:10.1121/1.399962

Berg, B. G. (2004). A molecular description of profile analysis: Decision
weights and internal noise. *Journal of the Acoustical Society of
America, 115,* 822–829. doi:10.1121/1.1639904

Berg, B. G., & Robinson, D. E. (1987). Multiple observations and internal
noise. *Journal of the Acoustical Society of America, 81,* S33–S33.

Boehnke, S. E., & Phillips, D. P. (2005). The relation between auditory
temporal interval processing and sequential stream segregation ex-
amined with stimulus laterality differences. *Perception &
Psychophysics, 67,* 1088–1101.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of
Experimental Psychology. Human Perception and Performance, 4,*
380–387. doi:10.1037/0096-1523.4.3.380

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organi-
zation of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segre-
gation and perception of order in rapid sequences of tones. *Journal
of Experimental Psychology, 89,* 244–249.

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual
organization of complex auditory sequences: Effect of number of
simultaneous subsequences and frequency separation. *Journal of
Experimental Psychology: Human Perception and Performance,
25,* 1742–1759.

Burton, P., Gurrin, L., & Sly, P. (1998). Extending the simple linear
regression model to account for correlated responses: An introduc-
tion to generalized estimating equations and multi-level mixed
modelling. *Statistics in Medicine, 17,* 1261–1291.

Buus, S., & Florentine, M. (1991). Psychometric functions for level
discrimination. *Journal of the Acoustical Society of America, 90,*
1371–1380.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in
Cognitive Sciences, 8,* 465–471. doi:10.1016/j.tics.2004.08.008

Carlyon, R. P., Thompson, S. K., Heinrich, A., Pulvermüller, F., Davis,
M. H., Shtyrov, Y., & Johnsrude, I. S. (2010). Objective measures of
auditory scene analysis. In E. A. Lopez-Poveda, A. R. Palmer, & R.
Meddis (Eds.), *Advances in auditory research: Physiology, psycho-
physics and models* (pp. 507–519). Berlin, Germany: Springer.

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention,
Perception, & Psychophysics, 74,* 784–802. doi:10.3758/s13414-
012-0322-z

Cherry, E. C. (1953). Some experiments on the recognition of speech,
with one and with two ears. *Journal of the Acoustical Society of
America, 25,* 975–979.

Ciocca, V. (2008). The auditory organization of complex sounds.
Frontiers in Bioscience, 13, 148–169. doi:10.2741/2666

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*
(2nd ed.). Hillsdale, NJ: Erlbaum.

Cooke, M., & Ellis, D. P. W. (2001). The auditory organization of speech
and other sources in listeners and computational models. *Speech
Communication, 35,* 141–177. doi:10.1016/s0167-6393(00)00078-9

Cooper, H. R., & Roberts, B. (2009). Auditory stream segregation in
cochlear implant listeners: Measures based on temporal discrimina-
tion and interleaved melody recognition. *Journal of the Acoustical
Society of America, 126,* 1975–1987. doi:10.1121/1.3203210

Dittrich, K., & Oberfeld, D. (2009). A comparison of the temporal
weighting of annoyance and loudness. *Journal of the Acoustical
Society of America, 126,* 3168–3178.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals
marked by brief acoustic pulses of various intensities and spectra.
Perception & Psychophysics, 21, 125–142.

Dorfman, D. D., & Alf, E. (1969). Maximum-likelihood estimation of
parameters of signal-detection theory and determination of confidence
intervals—Rating-method data. *Journal of Mathematical Psychology,
6,* 487–496.

Durlach, N. I., Braida, L. D., & Ito, Y. (1986). Towards a model for
discrimination of broadband signals. *Journal of the Acoustical
Society of America, 80,* 63–72. doi:10.1121/1.394084

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., &
Schabenberger, O. (2008). An R^2 statistic for fixed effects in the
linear mixed model. *Statistics in Medicine, 27,* 6137–6157. doi:10.
1002/sim.3429

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., & Shamma, S. A.
(2009). Temporal coherence in the perceptual organization and
cortical representation of auditory scenes. *Neuron, 61,* 317–329.
doi:10.1016/j.neuron.2008.12.005

Ellis, D. P. W., & Vercoe, B. L. (1992). A perceptual representation of
sound for source separation. *Journal of the Acoustical Society of
America, 91,* 2334. doi:10.1121/1.403499

Fouladi, R. T., & Shieh, Y. Y. (2004). A comparison of two general
approaches to mixed model longitudinal analyses under small sam-
ple size conditions. *Communications in Statistics: Simulation and
Computation, 33,* 807–824.

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic,
isochronous sequence. *Journal of the Acoustical Society of America,
98,* 2524–2531.

Getty, D. J. (1975). Discrimination of short temporal intervals: A com-
parison of two models. *Perception & Psychophysics, 18,* 1–8.

Green, D. M. (1958). Detection of multiple component signals in noise.
Journal of the Acoustical Society of America, 30, 904–911.

Green, D. M. (1964). Consistency of auditory detection judgments.
Psychological Review, 71, 392–407.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two
recognition measures of short-term memory. *Psychological
Bulletin, 66,* 228–234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and
psychophysics.* New York, NY: Wiley.

Green, D. M., von Gierke, S. M., & Hanna, T. E. (1986). Absolute
intensity discrimination using two responses. *Journal of the
Acoustical Society of America, 79,* 760–766.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object?
Nature Reviews. Neuroscience, 5, 887–892. doi:10.1038/Nrn1538

Grondin, S. (2012). Violation of the scalar property for time perception
between 1 and 2 s: Evidence from interval discrimination, reproduc-
tion, and categorization. *Journal of Experimental Psychology.
Human Perception and Performance, 38,* 880–890. doi:10.1037/
a0027188

Hanley, J. A. (1988). The robustness of the binormal assumptions used in
fitting ROC curves. *Medical Decision Making, 8,* 197–203.

Hartmann, W. M. (2005). *Signals, sound, and sensation* (5th ed.). New
York, NY: Springer.

Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural
Computation, 17,* 1875–1902. doi:10.1162/0899766054322964

Haywood, N. R., & Roberts, B. (2010). Build-up of the tendency to
segregate auditory streams: Resetting effects evoked by a single
deviant tone. *Journal of the Acoustical Society of America, 128,*
3019–3031. doi:10.1121/1.3488675

Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies
in auditory timing: 1. Simple patterns. *Perception & Psychophysics,
47,* 215–226.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of
significance. *Biometrika, 75,* 800–802.

IEC 318. (1970). *An IEC artificial ear, of the wide band type, for the
calibration of earphones used in audiometry.* Geneva, Switzerland:
International Electrotechnical Commission.

Iverson, G., & Bamber, D. (1997). The generalized area theorem in signal
detection theory. In R. D. Luce & A. A. J. Marley (Eds.), *Choice,
decision, and measurement: essays in honor of R. Duncan Luce* (pp.
301–318). Mahwah, NJ: Erlbaum.

Izumi, A. (1999). The effect of marker frequency disparity on the dis-
crimination of gap duration in monkeys. *Perception, 28,* 437–444.

Jesteadt, W., Nizami, L., & Schairer, K. S. (2003). A measure of internal noise based on sample discrimination. *Journal of the Acoustical Society of America, 114,* 2147–2157. doi:10.1121/1.1610456

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology. Human Perception and Performance, 21,* 293–307.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53,* 983–997.

Keselman, H. J. (1994). Stepwise and simultaneous multiple comparison procedures of repeated measures' means. *Journal of Educational Statistics, 19,* 127–162.

Kinney, J. A. S. (1961). Discrimination in auditory and visual patterns. *American Journal of Psychology, 74,* 529–541. doi:10.2307/1419663

Kortekaas, R., Buus, S., & Florentine, M. (2003). Perceptual weights in auditory level discrimination. *Journal of the Acoustical Society of America, 113,* 3306–3322. doi:10.1121/1.1570441

Liang, K. Y., & Zeger, S. L. (1993). Regression analysis for correlated data. *Annual Review of Public Health, 14,* 43–68.

Lister, J., Besing, J., & Koehnke, J. (2002). Effects of age and frequency disparity on gap discrimination. *Journal of the Acoustical Society of America, 111,* 2793–2800. doi:10.1121/1.1476685

Lutfi, R. A. (1989). Informational processing of complex sound. I: Intensity discrimination. *Journal of the Acoustical Society of America, 86,* 934–944.

Lutfi, R. A., & Liu, C. J. (2011). Target enhancement and noise cancellation in the identification of a rudimentary sound source in noise. *Journal of the Acoustical Society of America, 129,* EL52–EL56. doi:10.1121/1.3533000

Ma, L., Micheyl, C., Yin, P. B., Oxenham, A. J., & Shamma, S. A. (2010). Behavioral measures of auditory streaming in ferrets (*Mustela putorius*). *Journal of Comparative Psychology, 124,* 317–330. doi:10.1037/a0018273

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Macmillan, N. A., Rotello, C. M., & Miller, J. O. (2004). The sampling distributions of Gaussian ROC statistics. *Perception & Psychophysics, 66,* 406–421.

Matthews, W. J., & Grondin, S. (2012). On the replication of Kristofferson's (1980) quantal timing for duration discrimination: Some learning but no quanta and not much of a Weber constant. *Attention, Perception, & Psychophysics, 74,* 1056–1072. doi:10.3758/s13414-012-0282-3

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, 42,* 109–142.

Micheyl, C., Carlyon, R. P., Cusack, R., & Moore, B. C. J. (2005). Performance measures of auditory organization. In D. Pressnitzer, S. McAdams, A. DeCheveigné, & L. Collet (Eds.), *Auditory signal processing: Physiology, psychoacoustics, and models* (pp. 203–211). New York, NY: Springer.

Micheyl, C., & Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *Journal of the Association for Research in Otolaryngology, 11,* 709–724. doi:10.1007/s10162-010-0227-2

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America, 22,* 637–638.

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B, 367,* 919–931. doi:10.1098/rstb.2011.0355

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1999). Effects of frequency and duration on psychometric functions for detection of

increments and decrements in sinusoids in noise. *Journal of the Acoustical Society of America, 106,* 3539–3552. doi:10.1121/1.428207

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research, 46,* 2456–2464. doi:10.1016/j.visres.2006.01.029

Nelken, I., & Bar-Yosef, O. (2008). Neurons and objects: The case of auditory cortex. *Frontiers in Neuroscience, 2,* 107–113. doi:10.3389/neuro.01.009.2008

Oberfeld, D. (2008). Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? *Canadian Journal of Experimental Psychology, 62,* 24–32. doi:10.1037/1196-1961.62.1.24

Oberfeld, D. (2009). The decision process in forward-masked intensity discrimination: Evidence from molecular analyses. *Journal of the Acoustical Society of America, 125,* 294–303. doi:10.1121/1.3021296

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and non-normal data. *Behavior Research Methods, 45,* 792–812. doi:10.3758/s13428-012-0281-2

Oberfeld, D., Kuta, M., & Jesteadt, W. (2013). Factors limiting performance in a multitone intensity-discrimination task: Disentangling nonoptimal decision weights and increased internal noise. *PLoS ONE, 8,* e79830. doi:10.1371/journal.pone.0079830

Oberfeld, D., & Plank, T. (2011). The temporal weighting of loudness: Effects of the level profile. *Attention, Perception, & Psychophysics, 73,* 189–208. doi:10.3758/s13414-010-0011-8

Pedersen, B., & Ellermeier, W. (2008). Temporal weights in the level discrimination of time-varying sounds. *Journal of the Acoustical Society of America, 123,* 963–972. doi:10.1121/1.2822883

Pressnitzer, D., & Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Current Biology, 16,* 1351–1357. doi:10.1016/j.cub.2006.05.054

Rammsayer, T. H. (2010). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Attention, Perception, & Psychophysics, 72,* 1591–1600. doi:10.3758/APP.72.6.1591

Richards, V. M. (2002). Effects of a limited class of nonlinearities on estimates of relative weights. *Journal of the Acoustical Society of America, 111,* 1012–1017.

Richards, V. M., Carreira, E. M., & Shen, Y. (2012). Toward an objective measure for a "stream segregation" task. *Journal of the Acoustical Society of America, 131,* EL8–EL13. doi:10.1121/1.3664107

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2008). Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. *Journal of Experimental Psychology. Human Perception and Performance, 34,* 992–1006. doi:10.1037/0096-1523.34.4.992

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29,* 203–227. doi:10.1146/annurev.neuro.29.051605.112939

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics, 53,* 179–189.

Shamma, S. (2008). On the emergence and awareness of auditory objects. *PLoS Biology, 6,* 1141–1143. doi:10.1371/journal.pbio.0060155

Shamma, S., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology, 20,* 361–366. doi:10.1016/j.conb.2010.03.009

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12,* 182–186. doi:10.1016/j.tics.2008.02.003

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience, 18,* 1–13. doi:10.1162/089892906775250021

Sorkin, R. D., Boggs, G. J., & Brady, S. L. (1982). Discrimination of temporal jitter in patterned sequences of tones. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 46–57.

Sorkin, R. D., Robinson, D. E., & Berg, B. G. (1987). A detection theory method for the analysis of visual and auditory displays. In *Proceedings of the 31st Annual Meeting of the Human Factors Society* (pp. 1184–1188). Santa Monica, CA: Human Factors Society.

Sussman, E., & Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia, 47,* 771–785. doi:10.1016/j.neuropsychologia.2008.12.007

Sussman, E., Wong, R., Horvath, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research, 225,* 117–127. doi:10.1016/j.heares.2006.12.013

Swets, J. A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology, 14,* 109–121.

Swets, J. A. (1986a). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Swets, J. A. (1986b). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin, 99,* 100–117.

Swets, J. A., Shipley, E. F., McKey, M. J., & Green, D. M. (1959). Multiple observations of signals in noise. *Journal of the Acoustical Society of America, 31,* 514–521.

Tanner, W. P., & Birdsall, T. G. (1958). Definitions of *d′* and eta as psychophysical measures. *Journal of the Acoustical Society of America, 30,* 922–928.

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,* 1253–1262. doi:10.1037/a0021925

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics, 61,* 87–106.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished PhD dissertation, Eindhoven University of Technology.

van Noorden, L. P. A. S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America, 61,* 1041–1045.

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. *Journal of the Acoustical Society of America, 106,* 938–945.

von Békésy, G. (1947). A new audiometer. *Acta Oto-Laryngologica, 35,* 411–422.

Wang, D. L., & Brown, G. J. (1999). Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks, 10,* 684–697.

Wang, D. L., & Chang, P. (2008). An oscillatory correlation model of auditory streaming. *Cognitive Neurodynamics, 2,* 7–19. doi:10.1007/s11571-007-9035-8

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science, 164,* 586–587.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences, 13,* 532–540. doi:10.1016/j.tics.2009.09.003

Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics: Simulation and Computation, 22,* 1079–1106.

Wolfinger, R. D. (1996). Heterogeneous variance: Covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1,* 205–230.