Journal of New Music Research
2007, Vol. 36, No. 4, pp. 313 – 336

Perceiving and Distinguishing Simple Timespan Ratios without Metric Reinforcement

Benjamin Carson

University of California, USA

1. Overview

This study is an investigation of whether we can perceive proportionally consonant timespan ratios as instances of rhythmic regularity, particularly when they are not reinforced by traditional meter. "Consonant ratio" refers here to any small-sum ratio like 1:1, 1:2, or 1:3, affecting a successive pair of interonset intervals (hereafter "timespans"); as opposed to larger-sum ratios like 3:4, 5:9, and 4:11. I will refer to rhythms with an abundance of consonant ratios as "simple," and to those with a lack of such consonance as "complex." (With these terms, I do not mean anything inherent about perception or experience; they are intended only as descriptions of a rhythm's quantifiable features.) My basic goal is to know how ratio characteristics will affect our sense of regularity in a musical passage, independent of the influence of meter. But "metric hierarchy" and "timespan ratio characteristics" are conceptually entangled, and that poses a difficult challenge. The present study therefore aims to sort through these practical and conceptual entanglements, outline an experimental approach to the problem, and discuss the data that we have collected in one version of that experimental approach.

An ability to distinguish timespan ratio characteristics in the absence of meter would carry important implications for the study of *expectancy* – a key concept in most recent research in rhythm perception. Expectancy – a perceiver's expectation of events in the future, based on patterns in the past – is often attributed to both proportional simplicity and metric hierarchy (Lerdahl & Jackendoff, 1983; Jones & Boltz, 1989; Desain 1992), but the two factors are not normally teased apart. In some theoretical approaches, the distinction of regular from irregular rhythm depends on whether events establish consonant metric patterns as cues for the expectation of a larger structure.[1] Following Desain's (1992) study, a number of recent experiments have confirmed that expectancy is a distinct and important outcome of metric regularity (Large & Jones, 1999; Timmers et al., 2000). Hasty's (1997) concept of durational *projection* further extends this notion, beyond the obvious level of meter, and into middleground aspects of musical time.

In these teleological models of rhythm, however, an opposing question arises: can there also be "individuated", or "metrically independent" objects of temporal experience, that determine anything about the way we hear rhythm? An unmetered timespan ratio is made of only three events, bracketing two timespans; the percept of that ratio is absent until the second timespan has been completed, at the moment of the third event. If a timespan ratio has any distinct quality – for example, its perceived simplicity or complexity – then that quality could be considered a property of a single timepoint. Thus, if we are capable of distinguishing a "simple" additive timespan ratio from a complex one, then we will have to account for that distinction in an out-of-time perceptual mode. In that case it may be necessary to supplement our expectancy-based theories of rhythm with some notion of a reducible, retrospective, and "collapsible" temporality.

[1] See for example Eugene Narmour's (2000) "Implication-Realization" model; also David Temperley's (2001, 35–38) discussion of the notion of meter in Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music*.

Correspondence: Benjamin Carson, Music Center, Faculty Services, University of California, 1156 High St., Santa Cruz, CA 95064, USA. E-mail: benja.carson@gmail.com

DOI: 10.1080/09298210802015898 © 2007 Taylor & Francis

It is important to note that these two models of rhythm perception – "expectancy-producing" models and "individuating" or "qualitative" models – are interdependent. The philosopher Gilles Deleuze has clarified this interdependency in relation to a number of relational experiences of time, including developmental narratives in psychology and history, as well as aesthetic narratives in film and music. In *Difference and Repetition* (1994, originally 1968), Deleuze theorizes experiences of time and structure through an interaction between "difference in itself" and "repetition for itself." "Repetition for itself" describes objects that distribute on linear and metric continua; examples of this would include periodic features of time, like musical meter. Such repetitions transform our experience of "difference in itself" – for example, the qualitative distinction of ratio consonance – into integrated continuities; difference effectively becomes similarity (Deleuze 1994 [1968], 28–32, 70–71). Deleuze's (1994 [1968], 168–221) argument turns the contrast between continuity and difference back on itself, so that syntheses of these two models of thought can account for a variety of historical philosophies of meaning and identity.

My immediate interest here is more modest, but nevertheless reaches toward the solution of a similar problem: how conceptual structures of repetition, like meter, affect the larger possibilities of difference perception in music. An important starting point in the investigation is the common-sense premise that our expectation of continuity is strengthened when a rhythm repeatedly juxtaposes and repeats identical timespans (repetitions "for themselves") in small-sum ratios like 1:2, 1:3, and 1:5. Conversely, when non-consonant juxtapositions obscure a basic timespan's repetition (as in ♪ ♪. or ♪. ♩, 2:3 or 3:4), the underlying denomination of the ratio is more difficult to infer. Expectancy, in other words, is weakened by complex timespan ratios, just as our experience of difference "in itself" – individuation of features out of time – is hypothetically strengthened.

The missing question in that intuitive relationship, however, is *metric hierarchy*. Metric hierarchy is the condition in which events separated by time are linked, or made equivalent, according to their positions in relation to repeated timespan proportions on more than one scale of periodicity. This feature is potentially separate from the nature of the timespan proportions themselves. A few recent studies have begun to clarify this separation, by carefully distinguishing the perception of meter from the perception of rhythm. Repp et al. (2005) demonstrate that meter plays a significantly lesser role than timespan proportion in the accuracy of listeners' performance of "uneven" rhythms (based on successions of timespans of lengths 2 and 3, at various tempi). Keller and Burnham (2005) show that "non-metrical conditions" present a significant detriment to the task of distinguishing multipart or texturally complex rhythmic stimuli. In earlier investigations, the inference of a single timespan unit (Idson & Massaro, 1976; Handel, 1993), or of repeated larger timespans or timespan groups (Povel & Essens, 1985; Desain, 1992) – even when not always made explicit by events – yielded an important inferential "clock", against which irregular or complex timespan pairs will produce perceptual dissonances. The scale at which similar event-groups repeat (such as the duration of a typical bar of metered music) can act as a comparative mechanism that trumps the question of proportionality: for example, the difference between the timespan ratios 4:5 and 4:4 (1:1) is much more salient when the ratio itself is a repeated sequential feature, as in an *ostinato* or ground bass, than when it occurs in an isolated circumstance. (Technically, the sequential repetition of a timespan ratio 4:5 produces a 9:9 (1:1) relationship, one structural level above the surface.)

How then, does the presence of simple or complex proportions, completely ungoverned by meter or repetition, influence our impressions of complex rhythmic textures? The question is important for at least two reasons. First, composers have long associated complex rhythmic proportions with complex musical experience. This association often motivates us to require performers to think outside of traditional metric hierarchies, where scores utilize nested tuplets, abstract metric modulations, or other non-standard rhythm notation, in order to reject or avoid binary and ternary timespan divisions. But empirical studies have not formalized the conditions under which such complexities of rhythm produce viscerally ambiguous or complex musical experiences.

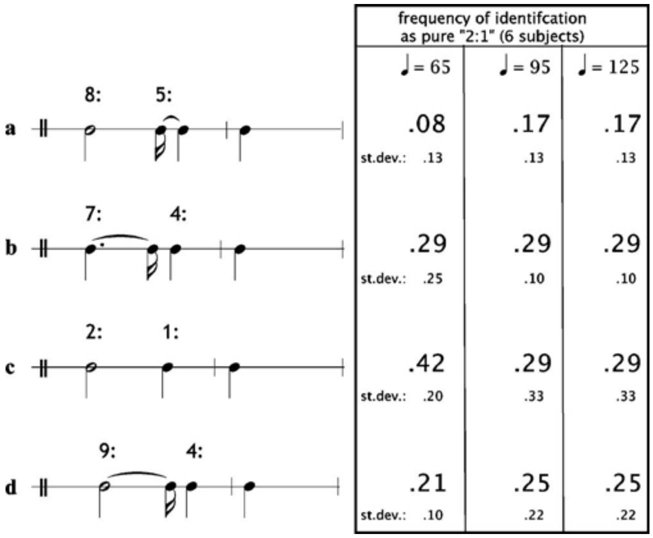
Secondly, the question of how we perceive unreinforced temporal structures carries with it some more general implications for how listeners engage musical time. At least since Ernst Kurth's 1917 *Grundlagen des linearen Kontrapunkts* ("Foundations of Linear Counterpoint"), theorists have idealized a dynamically unfolding and diachronic aspect of our psychological engagement of musical form. Kurth argues (perhaps more from a sense of desire for potent metaphors than from any empirical observation) that the truly resonant and persistent features of musical expression are not the events themselves, but a mysterious additional force that acts upon them to "counteract the autonomous significance of the individual tone" or individuated structure (Kurth [1917] 1991, p. 60). Like Kurth, Leonard Meyer holds that the musical composition is not a thing but "a process which gives rise to a dynamic experience" (1956, p. 54). Meyer likewise advocates a model of music perception in which "we are constantly revising our opinions... in the light of present events" (1956, p. 49), with continuously evolving predictions playing a dominating role in the apprehension of musical structure. The term expectancy comes to mind in these important precursors to Narmour's "implication-realization" model.

But psychological studies of rhythm have been much less likely to explore any of the ways in which synchronic time structures coalesce and persist in a rhythmic consciousness outside of musical time.

1.1 Preliminary test

To help clarify the immediate terms of our experiment, I designed a preliminary test of our ability to distinguish solitary timespan pairs. Figure 1 shows a set of four three-note rhythmic stimuli. Although the rhythms are proportionally similar – all approximately 2:1 – they are very different in complexity as I have defined it. A brief study[2] of these rhythms involved six musicians, choosing which of the four ametric stimuli expressed a "real 2:1" ratio. At best, the stimuli were difficult to distinguish: at the slowest tempo, mean reports only marginally favoured 2:1, with no significant distinction from its nearest alternatives;[3] at higher tempos, the "real 2:1" stimulus was not favoured by any significant margin.

These results suggest, at first glance, that listeners rely heavily on some kind of continuously reinforced meter in order to distinguish simple timespan ratios from complex ones. But the absence of metric hierarchy from these stimuli may not have been the only relevant factor. Listening to timespan pairs out of context is an unusual activity; in this test, participants' assessments of the stimuli powerfully incorporate variables like minutely fluctuating attention, priming "effects", and understandable confusions of short-term memory, that might have transposed listeners' experiences of one stimulus onto another.



frequency of identifcation as pure "2:1" (6 subjects)	♩= 65	♩= 95	♩= 125
a	.08	.17	.17
st.dev.:	.13	.13	.13
b	.29	.29	.29
st.dev.:	.25	.10	.10
c	.42	.29	.29
st.dev.:	.20	.33	.33
d	.21	.25	.25
st.dev.:	.10	.22	.22

Fig. 1. Brief "ametric" (unrepeated) stimuli, in which listeners failed to make reliable distinctions between complex and simple timespan pairs that were proportionally similar.

2. Background concepts for the main experiment

A more thorough investigation of this question requires a combination of methods from experimental psychology and music composition. To control against the problems described above, I needed a type of stimulus that produces more abundant timespan proportions, so that the characteristic of simplicity or complexity could persist through a more lengthy (and more conventionally musical) experience. At the same time, the stimuli must not be repetitive or metrically hierarchical. To meet these criteria, I composed long monodic (non-polyphonic) event sequences that contained simple timespan ratios interleaved with complex ones. I then adjusted pitch- and loudness-features of the events in order to emphasize and/or inhibit perception of the select interleaved group. The compositions' consonant features – although selectively emphasized by pitch or intensity differences – were not reinforced by significant repetition at larger structural levels; i.e. meter would not play a role in our perception of them as consonant. Yet, in comparison to the stimuli in our preliminary experiment, they were more persistent and melodic, so that participants might experience and assess them in a manner that resembles our everyday musical experiences.

The experiment described here is necessarily complex, because the phenomenon I am attempting to isolate depends on a variety of factors. The conditions under which these interleaved event streams would segregate, or coalesce, into separate percepts, exposing the rhythm's consonant features, could include any number of dimensional distinctions. Bregman (1990) has thoroughly

[2] Six musicians heard a recorded MIDI piano instrument on the specified rhythms, played back through iTunes on an Apple Powerbook G4 connected to Sennheiser HD465 headphones. The 3-note rhythms were heard in twelve sets, each containing all four stimuli, in various pseudo-random orders, with sets separated by four seconds of silence. "Tempo" refers to the perminute frequency of 4 × the subdivision. Listeners could hear a repetition of ach set of four stimuli one time (only) at will, and at their own pace. With each set, listeners were asked to identify, at their own pace, their best estimate of which was the pure "2:1" rhythm, indicating their response with pencil and paper on multiple-choice survey sheets. Order of presentation did not significantly increase the likelihood of selection for any of the rhythms ($F(3,23) = 2.44$, $p < 0.05$).

[3] At ♩ = 95, 8:5 was significantly disadvantaged ($F(3,11) = 3.6577$, $p < 0.05$), while at the highest tempo, no significant discrimination occurred for any stimulus. By contrast, when the same group of listeners heard the patterns in a repeated loop (with four iterations of each timespan pair), most reports were nearly error-free (at ♩ = 65: $F(3,11) = 257.44$, $p < 0.0001$, at ♩ = 95: $F(3,11) = 83.33$, $p < 0.0001$, at ♩ = 125: $F(3,11) = 91.53$, $p < 0.001$, Tukey HSD yielded significant comparisons (all $p < 0.01$) paired with the 2:1 stimulus only).

examined those conditions, which include pitch-proximity and intensity – the primary independent variables of the present study – as well as cues like timbre, harmonicity, and Gestalt principles of ''common fate'' or co-evolution. (The latter possibly including features like meter and timespan consonance.) To better communicate the nature of the variables involved in the investigation, some general principles and concepts will first need clarification.

2.1 ''Ordered proportionality'' versus ''perceived meter'' in rhythm perception

We understand a rhythm to be an event-group, distributed in time, whose identity is qualified by ordered proportions among intervals in that distribution. Thus, when ordered sets of proportions in two sequences are identical, the sequences are said to share a rhythmic identity, even if their metric organizations are different.

As suggested above, our practical experience of rhythm also involves events and relationships that are bound to their metric contexts. *Rubato* (variation of tempo from one beat to the next) and *rallentando* or *accelerando* (incremental tempo change over time) might significantly alter the actual ''clock time'' proportionality of a rhythm, without compromising the rhythm's identity (see for example Povel & Essens, 1985; Essens, 1995; Timmers et al., 2000; Eck, 2001; Ashley, 2002). Nevertheless, this kind of identity can also be explained in the terms of the ''ordered proportionality'' principle, as long as we assess proportions in terms of a unit of musical meter that is perceived invariantly, even when the unit's clock-time length fluctuates. Though the actual timing of events is affected by rubato, their metric timing should be invariant. A converse situation should also be noted: if ''clock-time'' spans in rhythm X and Y are identical, *though their metric contexts differ*, the identity shared by X and Y might not be appreciated. Thus, in the common practice of rhythm in meter, strict timespan proportions cannot always offer the final word in statements about what gives a rhythm its identity. As a rule, then: *perceived meter can both guard rhythmic identity against transformations in clock-time, and obfuscate rhythmic identity in spite of clock-time invariance.*

To illustrate a related phenomenon, more directly important to the present study, Figure 2 shows three examples containing different rhythmic percepts, potentially shared by a single time-point succession. The examples demonstrate, first, the potential for rhythmic proportions to inhibit perception of a metric structure, rather than reinforce it. A simple stream of sixteen notes (Figure 2(a)) may be made complex by an irregular distribution of segregated notes among its members. Likewise, a rhythm that appears irregular (Figure 2(b)) may contain simpler or more familiar percepts through hidden patterns involving compound timespans (Figure 2(c)). The distribution of a subset of events (numbered *2*, *4*, and *6*) in Figure 2(c) juxtaposes consonant timespans (22 units and 11 units, 2:1), so that, *if that subset has a perceptually salient group identity*, a listener might experience timespans of length 11 as a kind of background pulse that the rhythm as a whole would otherwise not produce. Thus, when viewed strictly according to timespan proportions, a single rhythm has the capacity to differ considerably from one level of its structure to another.

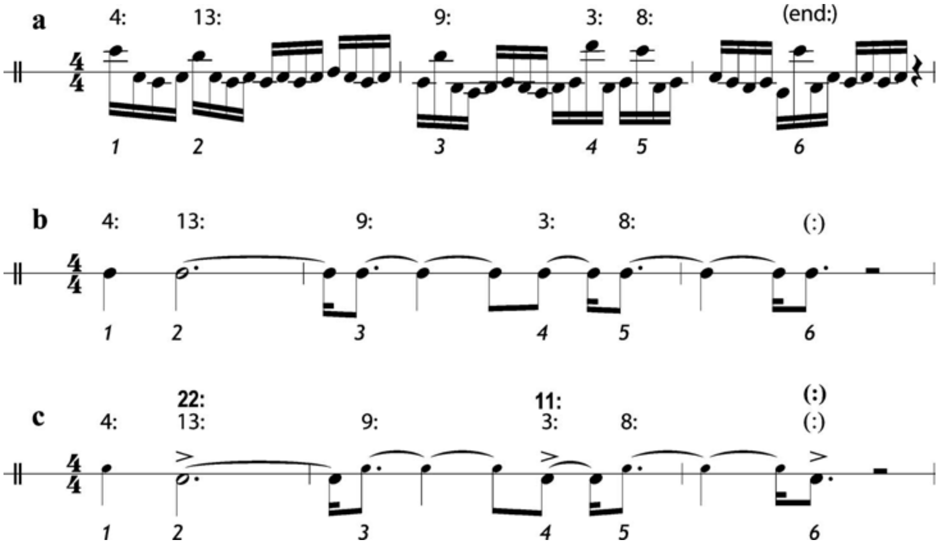


Fig. 2. Rhythms with potentially divergent percepts: (a) irregularly-spaced events in the context of a simple 16th-note pulse, (b) the same irregular pattern without a pulsed context, and (c) potential regularity in a smaller subset of the irregular pattern, distributed in a simple 2:1 (22:11) proportion of adjacent timespans.

3. Basic features of the experimental stimuli

The present experiment tested listeners' perceptions of stimuli resembling the example of Figure 2(c) above. At the surface, each stimulus consisted of highly irregular and unpulsed event sequences, excluding any instances of 1:1, 1:2, 1:3, 1:4, 2:3, or 1:5. Likewise the distinctive subset of notes in each example – the target rhythm – was always consonant, so that all proportions on this level would be either 1:1, 1:2, or 1:3. It is important to reiterate that the simple "target rhythm" and the irregular whole of the example are compositionally interdependent: the whole must be planned so that combinations of timespans in its series will yield a hidden structure of regularity. Rather than merely existing side-by-side, the irregular rhythm completely contains (and perhaps conceals) the rhythmically regular subset among its members. Dowling (1973) has constructed similar "compound" stimuli (for different purposes), which laid groundwork for the present experimental design. In Dowling's landmark study participants listened to stream-combinations and were tested for their reaction to the presence of familiar stimuli among concealed streams. Our study embeds target stimuli in a single distinguishable stream; however, the stream distinctions are formed among successions of non-simultaneous, non-overlapping single notes. This facilitates the possible closure of distance between streams so that independent stream identities might vanish completely, as in the example of Figure 2(b) above. Preserving a possible ambivalence between segregation and non-segregation is crucial to the present study.

There were eight compositions in all, separable into four types: A, B, C, and D. In A and B, the target rhythm was distinguished by loudness (determined by MIDI velocity, and measured in phons), while in C and D the target rhythm was distinguished by the mean pitch distance (in semitones) between the melody's pair of implied voices. In order to confirm perception of unreinforced consonant ratios, I hoped to observe a positive correlation between the target rhythm's streaming distance and the confidence of listeners' perceptions of pulse.

I also introduced considerable variety into the stimuli, unrelated to the probe variable. The most important example of this was the inclusion of an additional, simultaneous mode of potential stream segregation for each example, that would contradict, rather than confirm, the consonant proportions of the target rhythm. The principle that allows for this simultaneity of segregations is a common, if subtle, principle in everyday musical experience. If a succession of events is dispersed between two interleaved loudness groups, the louder, or "accented" notes, taken by themselves, might (as in the example of Figure 2(c)) gain some independent rhythmic identity. The same succession – involving identical time intervals between adjacent events – might at the same time exhibit two interleaved "pitch" streams, which segregate the notes at a different pace and contour from that of the loudness distinction. The pitch-stream arrangement will thus generate a new pair of rhythms that differ from those distinguished by accent. This aspect of our experiment resembles models of investigation pioneered by Gregory (1994) and Tekman (1995, 1997), which investigate the relative strengths of competing dimensional distinctions that selectively articulate scalar patterns. The present model differs in that cues potentially affect a persistent (whole-stimulus) stream segregation, rather than a time-specific grouping boundary.

Figure 3 diagrams this possibility in the abstract. Two competing percepts (pitch streaming and intensity streaming) coexist in a single, multidimensional contour (Figure 3(a)). Pitch information and intensity

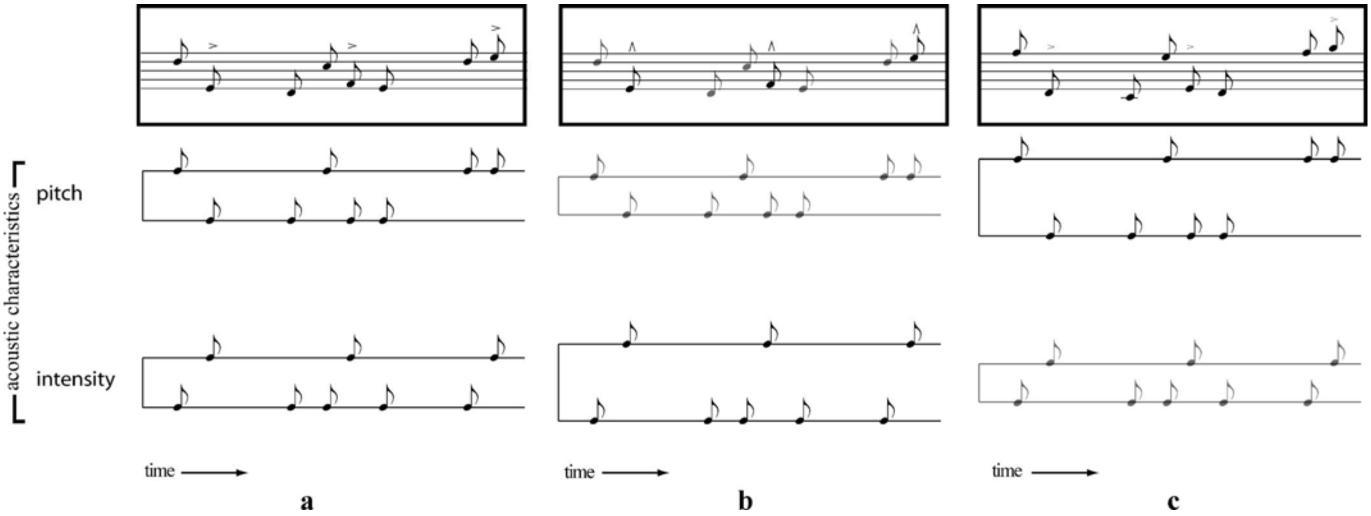


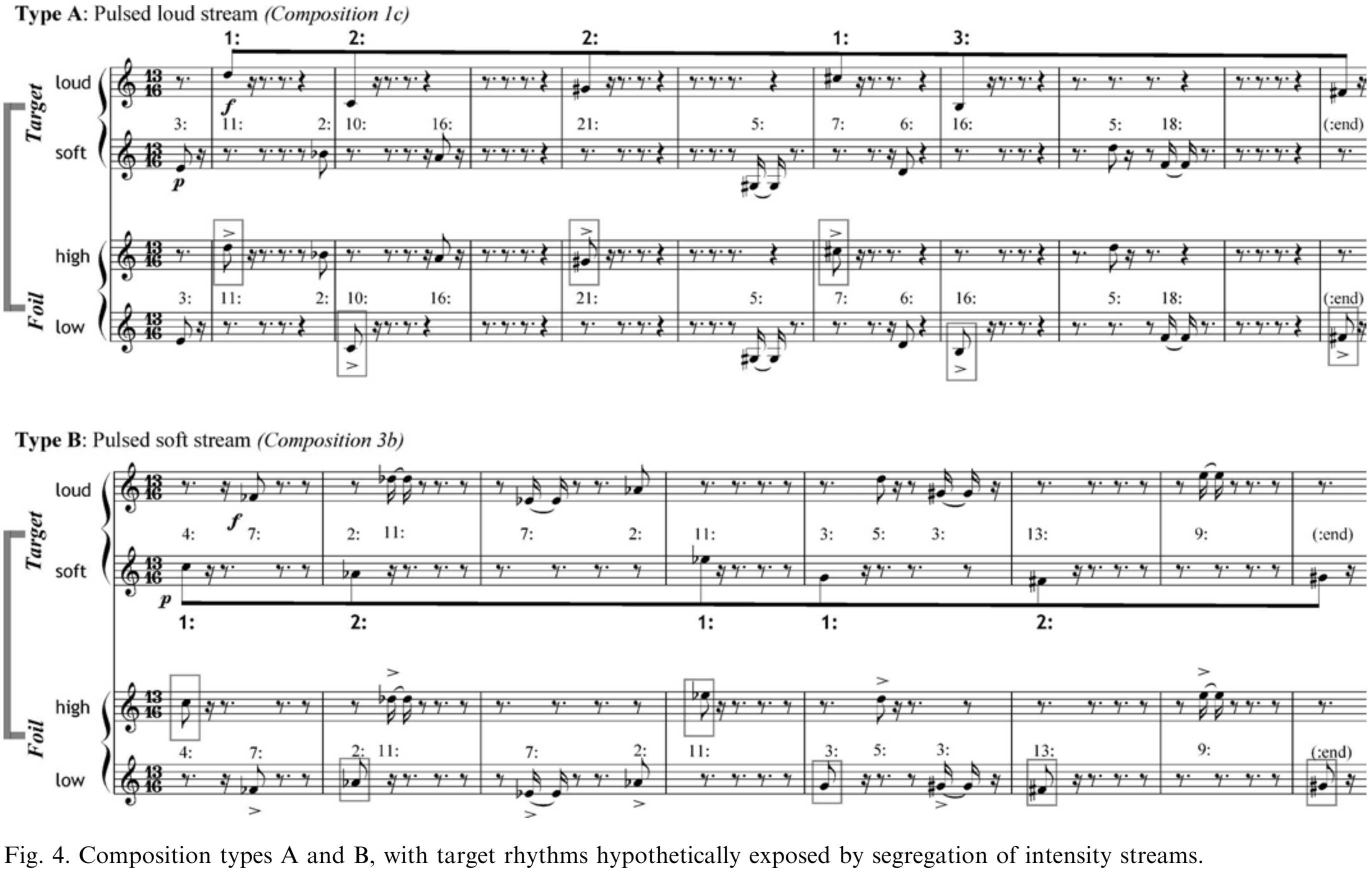
Fig. 3. Competing percepts in a melody with contrasted potentials for streaming.

information alternately prevail in the examples of Figures 3(b) and (c), affecting the melody's group organization to produce contrasting rhythmic results. Thus, a potential for two separate rhythmic percepts is available, and the strength of one percept should act as an inhibitor to the potential of the other. This additional level of complexity guards against an important confound: that the very phenomenon of stream-segregation, regardless of what rhythm it exposes, might have some independent effect on listeners' imaginary or real inferences of pulse. By tracking two conflicting types of segregation, we were able to insure that pulse confidence correlated positively with one and negatively with the other.

Figure 4 displays examples of melody types A and B, in which loudness segregation exposes the target rhythm. In each grand-staff displayed here, two staff pairs (“loud/soft” and “high/low”) represent the potentially conflicting stream organizations of the composition type, following the model of the pairs of horizontal lines in Figure 3, marked “intensity” and “pitch”. The two staff-pairs thus illustrate competing percepts, with one suggesting a pulsed percept – the target – and the other suggesting a percept in which pulse perception is inhibited – the foil. (It is crucial here to bear in mind that the aggregate of each 4-staff system represents *one* monodic event-succession, with each event represented twice – once in each staff-pair.) A long beam connects representations of the events of the target rhythm, which, if salient as a group, should increase listeners' confidence in a structure of regular pulse. In the pair of staves representing the foil percept, the events of the target rhythm are identified in faded, disjoined, rectangular boxes.

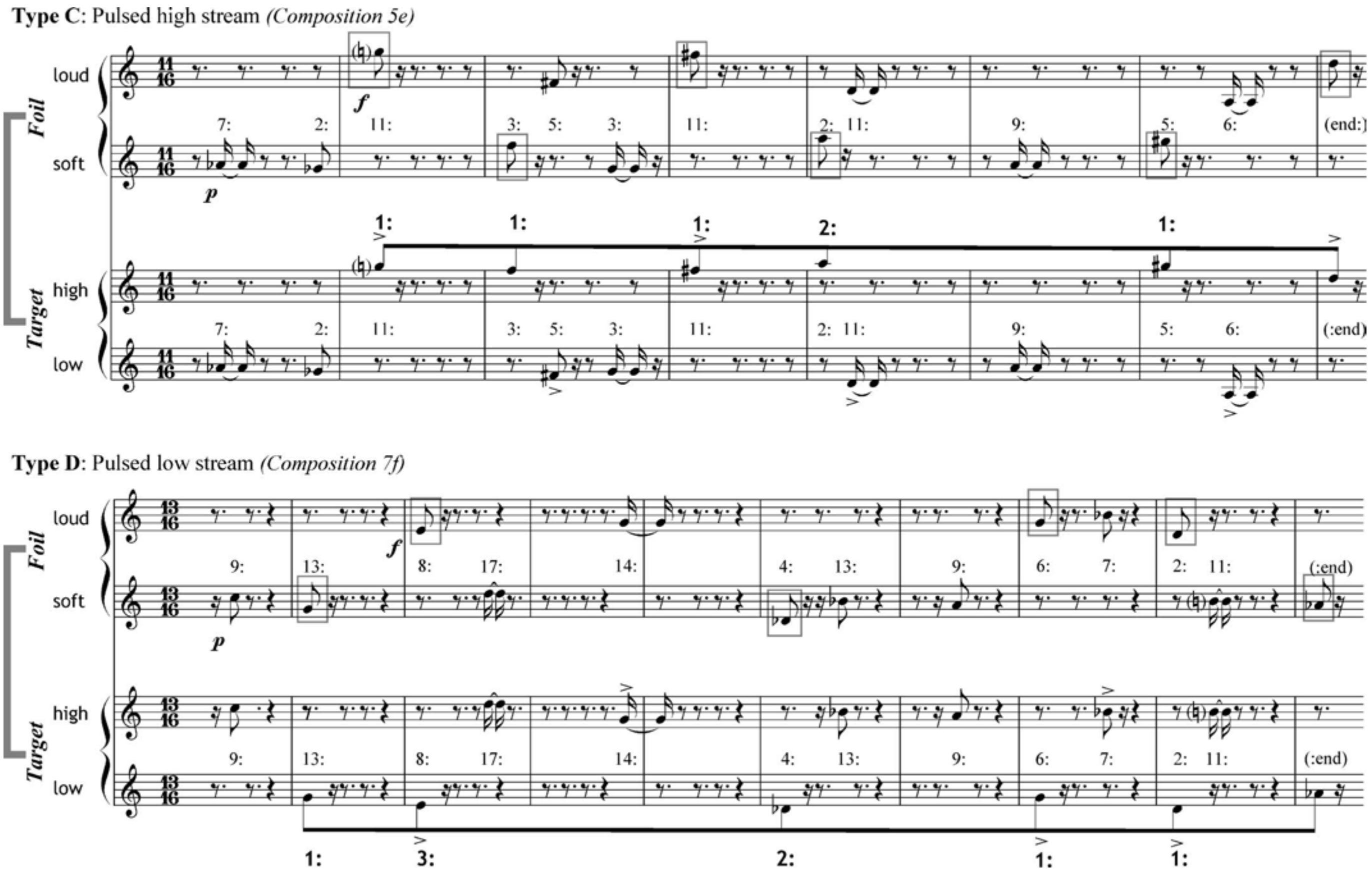
It should be noted that “A” and “B” in Figure 4 are not likely to elicit similar responses from listeners. Presumably, event-groups distinguished by intensity or loudness are *nested* rather than merely contrasting: minimum thresholds of intensity define group organization, while maximum thresholds do not, i.e. events may be too quiet for inclusion in a given group, but never too loud. Hereafter I will refer to this special condition, affecting intensity-based streaming, as the “hierarchical opposition” hypothesis: that the role of loudness plays a role in group distinction that is hierarchically opposed to that of softness. If this is true, then it seems unlikely that compositions of type B will be heard as “pulsed” or “simple”. On the other hand, special rhythmic segregation for loud notes is well represented in musical practice; the target rhythms in type A compositions might attract attention as simple rhythms.

Pitch-distinguished target rhythms (in compositions type C and D) are shown in Figure 5. These examples resemble the traditional concept of *compound melody* – melodic action that, in reductive harmonic analysis, would account for more than one polyphonic “voice”, even in a single stream of non-overlapping notes. The distinction of


Type A: Pulsed loud stream *(Composition 1c)*

Type B: Pulsed soft stream *(Composition 3b)*

Fig. 4. Composition types A and B, with target rhythms hypothetically exposed by segregation of intensity streams.


Type C: Pulsed high stream *(Composition 5e)*

Type D: Pulsed low stream *(Composition 7f)*

Foil
Target
loud
soft
high
low
f
p
7: 2: 11: 3: 5: 3: 11: 2: 11: 9: 5: 6: (end:)
1: 1: 1: 2: 1:
7: 2: 11: 3: 5: 3: 11: 2: 11: 9: 5: 6: (:end)
9: 13: 8: 17: 14: 4: 13: 9: 6: 7: 2: 11: (:end)
1: 3: 2: 1: 1:

Fig. 5. Composition types C and D, with target rhythms hypothetically exposed by segregation of pitch streams.

either low-stream or high-stream target rhythms will therefore resemble the common musical distinction of contrapuntal voices. Assuming relatively narrow-ranging pitch contours, and no voice crossings, the distinction between streams can be measured by a mean of total distances (in semitones) between voices taken at each of the composition's time-points.

Figure 6 presents a type-C composition from another view, to reinforce the complimentary and inverse roles stream-segregation and stream-confluence that coexist in the stimuli. This will also serve to clarify more precise aspects of the "target percept" we are attempting to locate. A single rhythm is displayed twice, in the staves initiated with a black circle. At top, it is further organized as loud and soft events; at bottom, as high and low. In the first organization, timespans among the loud events alone are just as ametric as those in the foreground. In the second organization, a simple pattern of consonant timespans emerges.

Lower-case letters in figure 6 mark events at which distinctive phases of *expectancy* in the accumulating rhythm can be imagined. Any letter "a" (including a', a", etc.) in this example marks times when a listener might assess an immediately preceding timespan *pair* as consonant, assuming the target stream (the treble clef) possesses salient identity as a group. Following generalized principles of beat-induction from Desain and Honig (1999, 29), a "bottom-up" process – in which a sense of underlying unit is induced – is theoretically possible in these moments. At all events /a′/, the prior pair is in the ratio 2:1, and at /a′/, the ratio is 1:1. At the event marked b, a larger pattern (the 4-unit sequence 1:2:1) has been heard twice in a row, facilitating a larger-scale "top-down" framework in which not only a beat, but metric order, might be inferred. But at /a′/ an unprecedented prior ratio of 1:3 can be heard, and the 1:2:1 pattern must again compete with a wide variety of other possible organizations. Since no special emphasis in the piece favours 1:2:1, and since divergent unpatterned organizations (including those suggested by low unpulsed notes) are also abundant, no singular metric reinforcement ever emerges completely. The target rhythm is thus an example of "consonant timespan ratios without metric reinforcement": at each "high-note event" a variety of proportional consonances is available, and a treble-dominant hearing of the passage should be a consonance-abundant experience.

3.1 Manipulation of the independent variables

Some readers may wonder why I have not tested these four streaming conditions (loudness interval above and

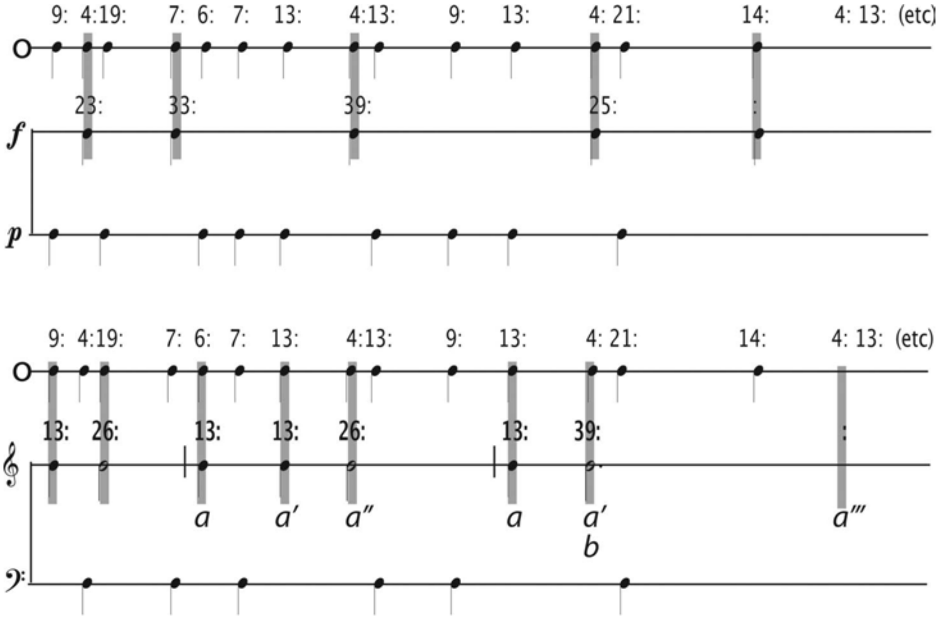

9: 4:19: 7: 6: 7: 13: 4:13: 9: 13: 4: 21: 14: 4: 13: (etc)
O
23: 33: 39: 25: :
f
p
9: 4:19: 7: 6: 7: 13: 4:13: 9: 13: 4: 21: 14: 4: 13: (etc)
O
13: 26: 13: 13: 26: 13: 39: :
a a′ a″ a a′ a‴
b

Fig. 6. Graphic representation of two different rhythmic percepts in an excerpt of Composition 5 (type C).

below, mean pitch interval above and below) in separate experiments. In fact, this combination of differing stimuli offers a distinct advantage, by limiting the repetition of familiar "solutions" within the narrative of stimuli, and thus reducing the participants' susceptibility to learning effects.

The seven versions of each composition, in alphabetical sequence according to labels a–g, are intended to mimic a progression between the generalized states represented in Figure 7, with version *a* favouring the loudness-streamed rhythms (the target in types A and B, or the foils in types C and D) and version *g* favouring those distinguished by pitch (the foils in A and B, or the target in C and D).

To further avoid learning effects, I distinguished stimuli within each of the four melody types in several other ways. First, each type is manifest in two contrasted rhythmic identities, for a total of eight basic compositions. (Four of these – compositions 1, 3, 5, and 7 – were shown in Figures 4 and 5, while composition 6 is represented in Figure 7.) Each composition's six-note "target rhythm" was unique, and none of the target rhythms, nor the aggregate, contained internally repeating patterns of rhythm or pitch. For purposes of consistency, the target rhythm included, in every case, the final note of the aggregate stimulus.

Our choice of specific streaming distances for the target and foil rhythms varied according to what I believe are conceptual differences among the four streaming variables. Figure 8 plots the treatments for all type A and type B stimuli. Two versions (exactly one version each of two compositions in the same type) occupy each data point, so that no data point would be bound to only one musical example. Thus, no composition treatment was without a counterpart that was significantly contrasted in numerous features (melodic contour, tempo, and MIDI instrument, as discussed below), but identical with respect to the core independent variables. For example, in Figure 8, both composition 1c and 2c (members of type A, but musically different) occupy the "+"-point, marked "c", with a mean inter-stream pitch interval of 10.1 semitones, and an inter-stream loudness interval of 4.2 phons.

Relative to type A examples, the role of the pitch-difference foil in type B was small – no example was inhibited by mean inter-stream differences greater than 9.9 semitones – because the "soft stream" target rhythms, if they are to be heard at all, hypothetically require the most favourable of circumstances. It also seemed prudent to concentrate the investigation of soft-stream target rhythms slightly more in the realm of smaller loudness differences, in case large loudness differences amplify the problem of the hierarchical opposition hypothesis. However, as shown in Figure 8, both type A and type B monodies were manifest by a controlled progression of pitch differences, allowing the statistical isolation of at least four increments of increasing inhibition to the loudness streams.

Characteristics of types C and D were distributed according to the plan shown in Figure 9. The pitch distances explored in these stimuli were of an overall higher range, for two reasons: first, initial informal probing of the variables suggested a low sensitivity to small pitch distances for interleaved melodies. Second, continued widening of the pitch-gap between the streams is not likely to strengthen the streaming effect

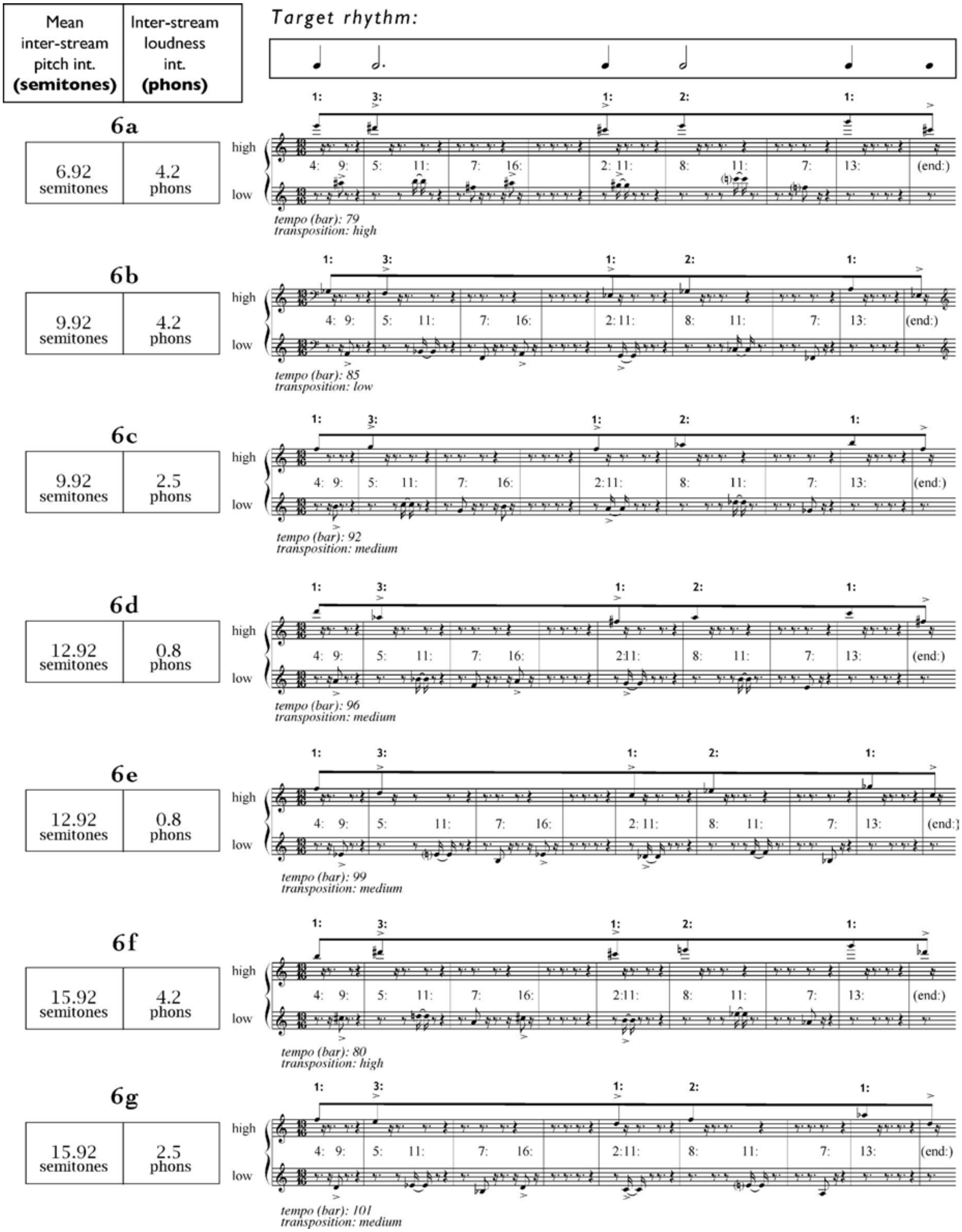


Fig. 7. Seven versions of composition 6, manifesting incremental progression between loudness-based streaming (6a) and pitch-based streaming (6g). Aspects of the composition change in register, contour, and tempo, independently of the progression of changes in streaming intervals.

indefinitely. (I predicted that the impact of increases in pitch-difference will diminish when both the initial difference and the increased difference are large.)

The treatments of composition types C and D were symmetrical, so that we could test both similarities and differences between the effects of high- and low-pitch

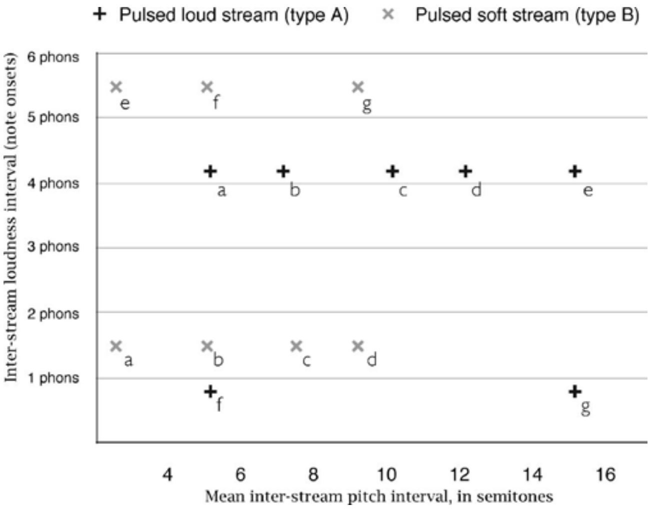


Fig. 8. Distribution of treatments for melodies type A and B.

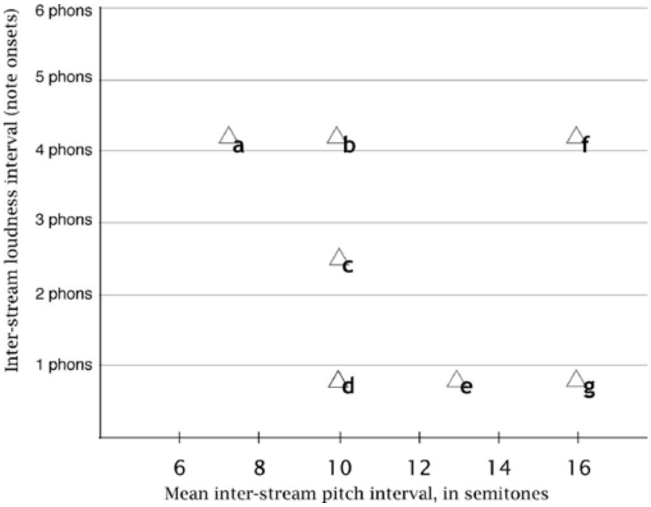


Fig. 9. Distribution of treatments for melodies type C and D.

streaming. The arrangement of the primary independent variables for melodies in both types enables several inferential operations: a three-sample test of pitch streaming effects at each of two loudness-streaming foils (4.2 phons and 0.8 phons), a three-sample test of loudness-streaming inhibition effects (when pitches stream at a mean of 9.9 semitones), and a two-factor analysis of variance combining elements of both three-sample tests (comparing across stream treatments at 0.8 and 4.2 phons, 9.9 and 16 semitones).

4. Procedure

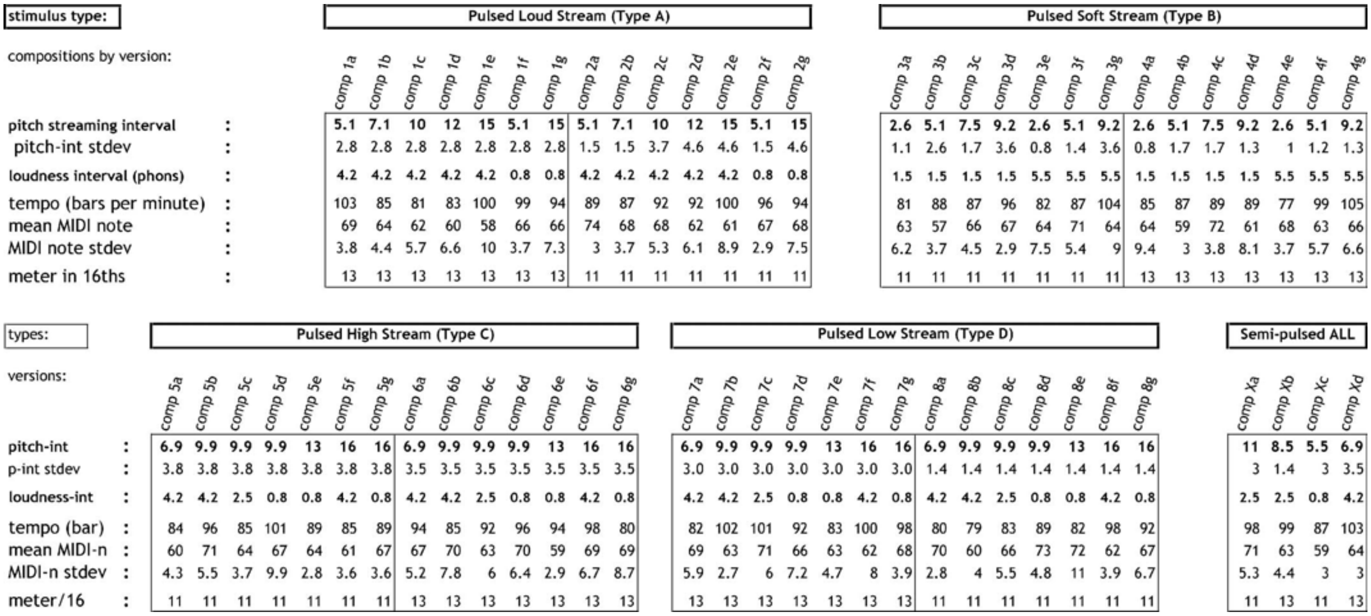
We tested 20 students in graduate seminars at the Santa Cruz campus of the University of California, ranging in age from 20 to 60, including both musically skilled and unskilled participants. All were fluent speakers of English and acknowledged comprehension of the test instructions.

The participants listened to a total of 60 stimuli in three separate sessions; the sessions consisted of groups of 20 stimuli, with two-to-three days' interval between sessions. I exposed each participant to the same three sets, in one of six orders. The stimuli orders within the three sets were carefully designed. No composition (1–8) occurred more than three times in any session, or more than twice in any period of seven stimuli. No version (a–g) of a given composition type (A–D) was repeated in any session.

The 20-stimuli sessions contained examples from all eight compositions, with no repeated versions. Because these compositions are stylistically homogenous, sharing general characteristics like atonality and (surface) ametricity, it seemed unlikely that participants would confidently acquire any salient thematic memory based on rhythmic similarities among the versions. Nevertheless, in producing the 56 stimuli, I introduced considerable variation that should discourage associative learning or unconscious categorization for one or more basic compositions. The differences shown in Figure 10 include tempo (measured in ''bars per minute'', since bars in our notation are the speed of beats in more conventional practice), melodic contour, overall pitch height (mean MIDI note), standard deviation in inter-stream pitch interval, and pitch-height standard deviation (mean MIDI note standard deviation). The wide variety of combined influences posed by these differing features made it unlikely that listeners would experience any priming effect, or recognize common features among the monodies as exact repetitions.

Variations in tempo were a particularly sensitive element of the experimental design. I found, initially, that listener confidence in assessing the regularity of these target rhythms was more likely when tempos were at markedly faster paces. (See also Figure 1; this effect was confirmed in the preliminary experiments on isolated and looped timespan pairs.) However, rigorously controlling that variable – by testing all stimuli with identical tempi – could produce other problems. In that scenario, the persistence of an inflexible underlying 16th-note unit might offer listeners an acquired metronomic sense, which could gradually wear away at the ambiguity of complex ratios. Over numerous repetitions, timespan ratios like 5:13, or 4:11, each being associated with specific and unchanging clock-time proportions, will also acquire a motivic familiarity. Such familiarity could easily be confused for pulse, and in addition to that problem, salient pulsed target rhythms that obtain motivic value early in a session would likely have an advantage in the confidence ratings given later in a session.

I addressed these problems in two ways. First, I varied the speeds of the compositions carefully across only a narrow range of tempos. This insured that listeners



stimulus type:		Pulsed Loud Stream (Type A)													
compositions by version:		comp 1a	comp 1b	comp 1c	comp 1d	comp 1e	comp 1f	comp 1g	comp 2a	comp 2b	comp 2c	comp 2d	comp 2e	comp 2f	comp 2g
pitch streaming interval	:	**5.1**	**7.1**	**10**	**12**	**15**	**5.1**	**15**	**5.1**	**7.1**	**10**	**12**	**15**	**5.1**	**15**
pitch-int stdev	:	2.8	2.8	2.8	2.8	2.8	2.8	2.8	1.5	1.5	3.7	4.6	4.6	1.5	4.6
loudness interval (phons)	:	**4.2**	**4.2**	**4.2**	**4.2**	**4.2**	**0.8**	**0.8**	**4.2**	**4.2**	**4.2**	**4.2**	**4.2**	**0.8**	**0.8**
tempo (bars per minute)	:	103	85	81	83	100	99	94	89	87	92	92	100	96	94
mean MIDI note	:	69	64	62	60	58	66	66	74	68	68	62	61	67	68
MIDI note stdev	:	3.8	4.4	5.7	6.6	10	3.7	7.3	3	3.7	5.3	6.1	8.9	2.9	7.5
meter in 16ths	:	13	13	13	13	13	13	13	11	11	11	11	11	11	11

stimulus type:		Pulsed Soft Stream (Type B)													
compositions by version:		comp 3a	comp 3b	comp 3c	comp 3d	comp 3e	comp 3f	comp 3g	comp 4a	comp 4b	comp 4c	comp 4d	comp 4e	comp 4f	comp 4g
pitch streaming interval	:	**2.6**	**5.1**	**7.5**	**9.2**	**2.6**	**5.1**	**9.2**	**2.6**	**5.1**	**7.5**	**9.2**	**2.6**	**5.1**	**9.2**
pitch-int stdev	:	1.1	2.6	1.7	3.6	0.8	1.4	3.6	0.8	1.7	1.7	1.3	1	1.2	1.3
loudness interval (phons)	:	**1.5**	**1.5**	**1.5**	**1.5**	**5.5**	**5.5**	**5.5**	**1.5**	**1.5**	**1.5**	**1.5**	**5.5**	**5.5**	**5.5**
tempo (bars per minute)	:	81	88	87	96	82	87	104	85	87	89	89	77	99	105
mean MIDI note	:	63	57	66	67	64	71	64	64	59	72	61	68	63	66
MIDI note stdev	:	6.2	3.7	4.5	2.9	7.5	5.4	9	9.4	3	3.8	8.1	3.7	5.7	6.6
meter in 16ths	:	11	11	11	11	11	11	11	13	13	13	13	13	13	13

types:		Pulsed High Stream (Type C)													
versions:		comp 5a	comp 5b	comp 5c	comp 5d	comp 5e	comp 5f	comp 5g	comp 6a	comp 6b	comp 6c	comp 6d	comp 6e	comp 6f	comp 6g
pitch-int	:	**6.9**	**9.9**	**9.9**	**9.9**	**13**	**16**	**16**	**6.9**	**9.9**	**9.9**	**9.9**	**13**	**16**	**16**
p-int stdev	:	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.5	3.5	3.5	3.5	3.5	3.5	3.5
loudness-int	:	**4.2**	**4.2**	**2.5**	**0.8**	**0.8**	**4.2**	**0.8**	**4.2**	**4.2**	**2.5**	**0.8**	**0.8**	**4.2**	**0.8**
tempo (bar)	:	84	96	85	101	89	85	89	94	85	92	96	94	98	80
mean MIDI-n	:	60	71	64	67	64	61	67	67	70	63	70	59	69	69
MIDI-n stdev	:	4.3	5.5	3.7	9.9	2.8	3.6	3.6	5.2	7.8	6	6.4	2.9	6.7	8.7
meter/16	:	11	11	11	11	11	11	11	13	13	13	13	13	13	13

types:		Pulsed Low Stream (Type D)													
versions:		comp 7a	comp 7b	comp 7c	comp 7d	comp 7e	comp 7f	comp 7g	comp 8a	comp 8b	comp 8c	comp 8d	comp 8e	comp 8f	comp 8g
pitch-int	:	**6.9**	**9.9**	**9.9**	**9.9**	**13**	**16**	**16**	**6.9**	**9.9**	**9.9**	**9.9**	**13**	**16**	**16**
p-int stdev	:	3.0	3.0	3.0	3.0	3.0	3.0	3.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4
loudness-int	:	**4.2**	**4.2**	**2.5**	**0.8**	**0.8**	**4.2**	**0.8**	**4.2**	**4.2**	**2.5**	**0.8**	**0.8**	**4.2**	**0.8**
tempo (bar)	:	82	102	101	92	83	100	98	80	79	83	89	82	98	92
mean MIDI-n	:	69	63	71	66	63	62	68	70	60	66	73	72	62	67
MIDI-n stdev	:	5.9	2.7	6	7.2	4.7	8	3.9	2.8	4	5.5	4.8	11	3.9	6.7
meter/16	:	13	13	13	13	13	13	13	11	11	11	11	11	11	11

types:		Semi-pulsed ALL			
versions:		comp Xa	comp Xb	comp Xc	comp Xd
pitch-int	:	**11**	**8.5**	**5.5**	**6.9**
p-int stdev	:	3	1.4	3	3.5
loudness-int	:	**2.5**	**2.5**	**0.8**	**4.2**
tempo (bar)	:	98	99	87	103
mean MIDI-n	:	71	63	59	64
MIDI-n stdev	:	5.3	4.4	3	3
meter/16	:	11	13	11	13

Fig. 10. A sample distribution of characteristics across three 20-stimulus sessions. Boldface rows indicate the primary independent variables, which did not change from session to session. Tempos ranged from 77 to 105 (bars per minute) in a normal distribution peaking at 91. Mean MIDI note ranged from 57 (A3) to 74 (D5), in a normal distribution peaking at 65.5. The assignment of features is pseudo-random, with manipulations to prevent contrast effects and correlations with the independent variables.

would never hear two hypothetically similar composition *versions* (in terms of the independent variables) at a similar pace. This strategy may further inhibit the potential for rhythmic/motivic value in an already chaotic landscape of musical examples. More important, it eliminates the potential that listeners might become acquainted with the underlying sixteenth-note texture of the unpulsed composite rhythm. We also pay a potential cost with this decision, in that listeners could more confidently rate the pulsedness of stimuli with higher tempos. However, the tempo variations were small enough to render that impact negligible, and if the impact is significant, we will measure the significance and account for it in our interpretations of the data.

Our second control against a familiarity affect was the insertion of four additional rhythmic variants into the mixture of stimuli, named Xa, Xb, Xc, and Xd (Figure 11). Composition "type X" is a heterogeneous category, but all four of its members differ from types A, B, C, and D in that simple/pulsed target rhythms are found in more than one of their potential streams, *as well as* in their "composite" forms; stream segregation should therefore have no meaningful impact on pulse confidence. The introduction of type X diversifies the materials presented in each session; it also serves as a control example in which high pulse confidence should be demonstrable. Even more importantly, because each version of Composition X resembles another composition in the stimuli set, the rhythmic variety that it introduces should discourage participants from the assumption that memorable features signal a rhythm identical to one heard previously in a session.

In order to be sensitive to disruption affects related to these two solutions, I also insured that change in tempo between any adjacent stimuli in the presentation order would be neither lesser than two nor greater than 15 bars per minute, except where adjacent examples moved to or from version "X". I distributed the tempos of the whole test on a bell curve, with infrequent values around 77 and 105, and frequent values around 91.

4.1 Stimuli, equipment, and environment

We realized these compositions using Nightingale notation and MIDI software, using the Macintosh-native synthesized piano instrument patch. I converted MIDI output files to AIFF format, and confirmed the rhythmic accuracy of the eight compositions by examining waveforms in the Audacity (freeware) sound editor. The examples averaged 5–6 s in length, and were played with 18 s intervals of silence between them. I used an Apple PowerBook G4 laptop computer, running iTunes music playback software, connected to DENON PM-915R "Precision Audio Component/Integrated Stereo Amplifier" with P-reference stereo loudspeakers, separated by 17′ in a $25 \times 25'$ soundproofed classroom environment at the University of California, Santa Cruz Music Center.

We took decibel readings from a SPER Scientific 840018 Sound Level Meter (OSHA compliant; ANSI S1.4 type 2). MIDI velocity indications were independently correlated with decibel differences at four pitch ranges (Figure 12). Velocity differences in the example

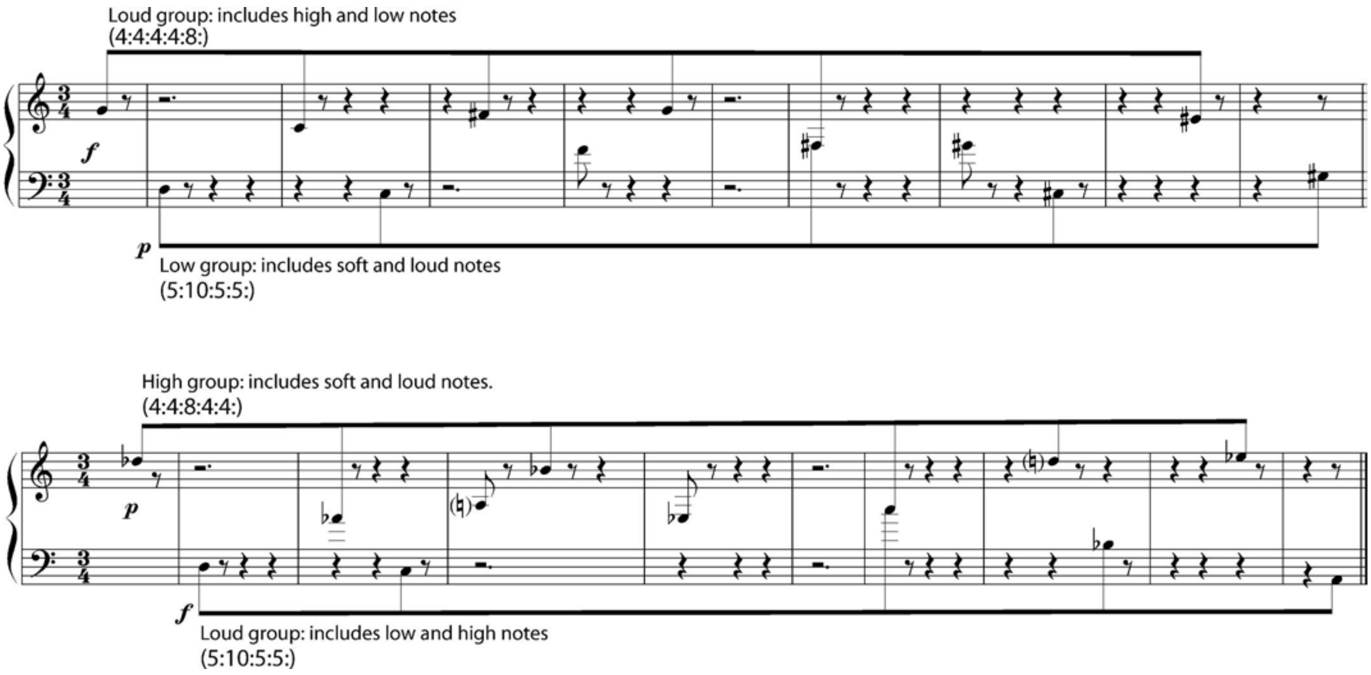
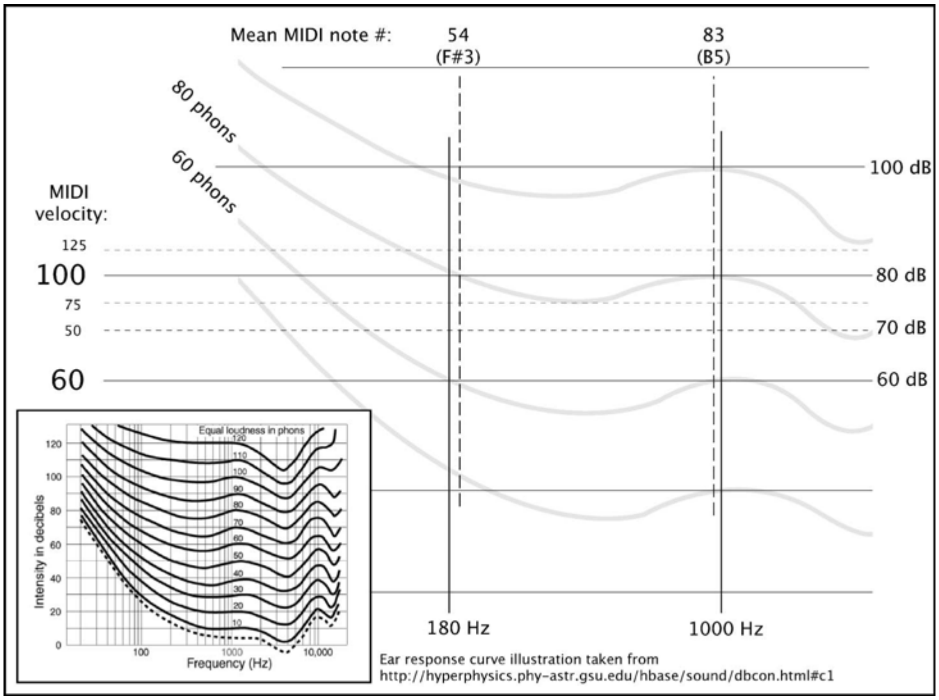

Loud group: includes high and low notes
(4:4:4:4:8:)
f
p
Low group: includes soft and loud notes
(5:10:5:5:)
High group: includes soft and loud notes.
(4:4:8:4:4:)
p
f
Loud group: includes low and high notes
(5:10:5:5:)

Fig. 11. "Type X" control stimuli a and b. Regardless of streaming effects (loud-to-soft versus low-to-high), compositions of type X should produce a rhythm similar to the target rhythms of types A, B, C, and D.



Ear response curve illustration taken from http://hyperphysics.phy-astr.gsu.edu/hbase/sound/dbcon.html#c1

Fig. 12. Source curves for the derivation of decibel optimizations at particular pitch ranges. We controlled upper and lower loudness streams of each example via MIDI velocity settings, the consistent trans-registral loudness of which six subjects confirmed. At mean MIDI notes 54 (185 Hz) and 83 (988 Hz), a range of 45 MIDI velocity points corresponded to a sound level difference measurement of 4.2 dB. At mean MIDI note 68 (415 Hz) and 105 (3520 Hz, not used in the experiment), the same range corresponded to 3.5 dB. Phon measurements were calibrated according to 4.2 phons = ±45 MIDI velocity points.

scores were consistent within an error of 2.0 dB across the full range of the composition versions. Normal ambient noise in the classroom during the tests was registered at 5 dB above reference, and peaked at no higher than +11 dB, with no more than one such disturbance (approximately 0.5 s in length) per test.

4.2 Execution

In advance of each session, we offered listeners a variety of practice stimuli that typified the possible presence or absence of pulse in the test examples. Some practice examples were similar in their reliance on obvious stream segregation for the salience of a pulse; others resembled type X, in that they involved multiple streams that could suggest a pulse more readily than any of the test examples, regardless of streaming effects. Still more priming examples were designed with a complete lack of metric regularity. The strong contrasts between extremes in these listening examples helped orient listeners to the subtlety of their assessment task, since none of the test examples resembles the simplicity and obvious "pulsedness" of ordinary musical experience.

We asked listeners to assess "how confidently [they could] find a pulse, or a beat, in each of these examples". Subjects were invited to consider rating 5 as equivalent to "*I am certain that I detected some regularity, even if only for a brief time*" and a 0 ranking as equivalent to "*I detected no regularity whatsoever, at any time*" Additional language suggested how the participants should interpret the individual ratings of 1, 2, 3, and 4, all of which suggested brief moments of suspected (but not confident) perception of regularity. Five seconds in advance of each example, participants were given a short spoken cue, consisting of the example number.

5. Methods of analysis for confidence ratings

In order to study confidence ratings in a careful way, I needed a reliable method of distinguishing between two types of variance among the participants. This task is made complicated by the subjective (and, strictly speaking, non-intervallic) nature of a confidence rating. Any difference between two participant's scores for a given treatment of the variable could reflect one or both of two entangled kinds of perceptual difference. First, the participants could be experiencing differences in their recognition of a pulse, based on their different ways of attending to the stimulus in the moment, their different ability to parse the streams or the time intervals, or their different overall approaches to rhythmic experience. Obviously, I must *preserve* these differences in the presentation of the data, because they are part of what I am studying, and they are intractable as listeners' expressions about the auditory stimulus. Second, participants will tend to have different understandings – even with abundant guidance and practice – of what should constitute "2", "3", "5", etc., on the scale. A pair of listeners, for example, might express highly similar *relative* rankings of confidence in relation to one of the variables, reflecting similar experiences of the stimuli, but represent those same rank-orders in very different distributions of numeric choices. In this type of difference, two listeners respond to similar percepts, in differently calibrated ways. In our experiment, knowledge of the success of the hypothesis depends on an ability to control *against* them.

We therefore produced a secondary set of scores for the whole test sample by converting each score into a *z score*, relative to the overall distribution of scores from its participant sample; this identifies a raw score in terms of its relationship to other scores *by the same participant*. My "z-conversions" do not assume, as would normally be the case, the "strict coherence" of ratings as interval scales; but they measure the way an individual listener has situated a particular stimulus in her whole collection of reports. The z-conversion illustrates important relationships by reducing one kind of intra-sample variance – derived from participant calibrations – that would otherwise be an obstacle to understanding basic relationships between subjects in the test.[4]

The effects of this technique on inferential statistical computation will be strong, as demonstrated by a preliminary exercise: I constructed an index for the hypothesis of pulse confidence for composition types C and D, which yielded, in general, the strongest pulse confidence correlations with the independent variables. Where P is pitch streaming distance, and I is loudness streaming distance, the function $f(X) = P/3 - (1 + \log_{10} I)$ yields results that increase in proportion to the exposure of the target rhythm ($P/3$), with a small attenuation for loudness streaming measured in the logarithmic decibel scale. Although this is a crude hypothesis, I found it strongly related to z-conversions of confidence ratings for collected samples in types C and D ($t(1,27) = +13.59$,

[4] Strict adherents to Stevens' (1951) typology proscriptions will not permit a z-score transformation of results from a ratings system or of subjective assessments. However, concern has grown steadily over the past few decades in the field of statistics, that Stevens' restrictions on treatments of ordinal data have unnecessarily "limited the ability to detect anomalies" in strictly ordinal psychometrics, and that in many cases will "restrict consideration of [experimental] errors to random perturbations" (Velleman, 1993; see also Tukey, 1957, 1977). Most prominently, Tukey (1977) has pointed out that Stevens' typologies will reject many conventional interval scales as technically ordinal, because errors in them are potentially systematic; nevertheless, their consequential "demotion" from all interval-based transformations would render entire categories of inquiry impossible.In their use, z-scores reflect something *real* contained completely within the collected information from any individual participant – namely, how one participant claimed to understand a musical stimulus *in ordinal, quasi-interval, comparison* to the rest of her understandings of musical stimuli. (We say "quasi-interval" here because subjects were instructed to treat zero as a true zero, reflecting no detection of pulse, and to consider the ratings as reflections of equidistant degrees of confidence.) Any errors – momentary or pervasive – that a subject commits in the *rough* assessment of the difference /0-1/ as equivalent to the difference /1–2/, are not ruled out, but preserved, in the transformation.

$p < 0.0001$), while the relationship of those ratings to the raw scores was poor ($t(1,27) = +0.33$, $p > 0.05$).

In the discussion below, I will use the term "z-scores" as shorthand for "pulse confidence ratings represented by their z-scores in relation to the participant's overall score distribution". However, the significance of my hypotheses for individual compositions does not depend on the z-score conversion method, and the results of raw confidence ratings will also be displayed.

5.1 Potential confounds in the results

We eliminated the results of two participants on grounds of non-completion (one case), and a pattern of responses that seemed to be determined solely by its resulting appearance on the response sheet (one case).

Composition "version pairs" within each type (1a/2a, 1b/2b, etc.) often produced mean confidence ratings that were significantly different, as should be expected given the number of difference parameters (tempo, register, melodic features, and overall rhythmic identity). Nevertheless, ANOVA results for the response collections across versions for six of the eight combined composition pairs – 1 and 2 (type A), 5 and 6 (type C), 7 and 8 (type D) – showed significant overall treatment effects. This indicated that the widening of the streaming variables directly affecting the target rhythm played a consistent role (among other factors) in pulse confidence for those composition types.

We found one significant correlation of pulse confidence to tempo, for composition types C and D at 4.2 phons, with 9.9 and 12.5 semitone streaming distances for the target rhythm ($r(11) = 0.58$, $p < 0.05$), $t(1,11) = +3.1$, $p < 0.05$). However, the relationship of the same scores to pitch streaming in two-factor ANOVA was much stronger ($t(1,11) = +9.59$, $p < 0.0001$), even accounting for variance in samples containing compositions at radically different tempos. We accounted for other possible tempo effects by comparison of variance: for example, when the streaming distance was 9.9 semitones, an apparent relationship of confidence to tempo was nonsignificant ($t(1,11) = 3.5$, $p > 0.05$), while the inverse relationship between z-scores and loudness streaming in the same score sample was significant $t(1,11) = +19.34$, $p < 0.0001$.

We observed no significant correlation of z-scores to variables such as the standard deviations of pitch streaming-distance, the overall pitch height of the stimulus, or slight differences in the abundance of non-target notes.

6. Results and discussion

The aggregate of 18 listeners' pulse-confidence reports was significantly correlated with the target rhythm streaming distance in compositions of types A, C, and D. These general results show that unmetered simple timespan ratios are appreciable, and occur to listeners as "perceptually simple", when they are sufficiently distinguished by either loudness accents, or by structures of contrapuntal voice distinguished by pitch.

For the "accented" target rhythms in type A compositions, pulse confidence decreased as the pitch-foil distance increased. Figure 13 illustrates the difference between mean confidence ratings, according to the treatments of the pitch foil, for compositions of types A and B. Compositions 1 and 2 (type A, "loud-stream target-rhythm" compositions) show significant negative correlation to the pitch streaming of the foil (Composition 1: $F(4,17) = 6.48$, $p < 0.001$; Composition 2: $F(4,17) = 12.49$, $p < 0.0001$). Mean confidence ranking across both compositions, showed an even stronger negative relationship to the foil ($F(4,34) = 14.50$, $p < 0.001$). In addition, the mean confidence at each pitch-streaming distance correlated with the distance of loudness streams ($F(4,35) = 14.50$, $p < 0.0001$). Thus, ranking of both the individual compositions, and the type A compositions together, show a positive correlation of confidence with the loudness-distance of the target rhythm, and a negative correlation of confidence with the pitch-distance of the foil.

Rankings for Composition 1 (represented by down-pointing triangles) remained steady in relation to the three smallest foil distances, but were radically lower when the foil distances were at their largest.[5] Conversely, rankings for Composition 2 were affected radically by differences among the small foil distances, and less so by the large ones.[6] These differences potentially reflect distinctive musical features of the compositions, which are discussed in more detail below.

Multi-factor analysis of variance among the collected reports of corresponding versions of Compositions 1 and 2 (type A) at 4.2 phons reveals a significant pitch-streaming effect among the streaming distances 5.1, 7.1,

[5] Tukey post-ANOVA testing for the relationships between foil distances in Composition 1 at 4.2 phons showed significance for relationships between streaming distance 15.1 and distances 5.1 ($p < 0.01$), 7.1 ($p < 0.01$), and 12.1 ($p < 0.05$; Tukey HSD $[0.05] = 0.67$, HSD $[0.01] = 0.08$), and in no other relationships among the distances. Thus, the streaming distance of this particular "pitch foil" only resulted in a pronounced loss of pulse confidence when it reached an interval of 15 semitones; in other instances, any detection and assessment of "relatively loud" event groups was not significantly hindered by large swerves in pitch contour.

[6] By contrast, only the smaller foil distances (5.1 and 7.1) in Composition 2 (still at 4.2 phons) were significant, with z-scores at 5.1 semitones significantly (and negatively) correlated with differences from z-scores at 10.1 ($p < 0.01$), 12.1 ($p < 0.01$), and 15.1 ($p < 0.01$; HSD $[0.01] = 0.88$) semitones. Z-scores at 7.1 semitones and 12.1 semitones were also significantly distinct ($p < 0.05$; HSD $[0.05] = 0.73$).

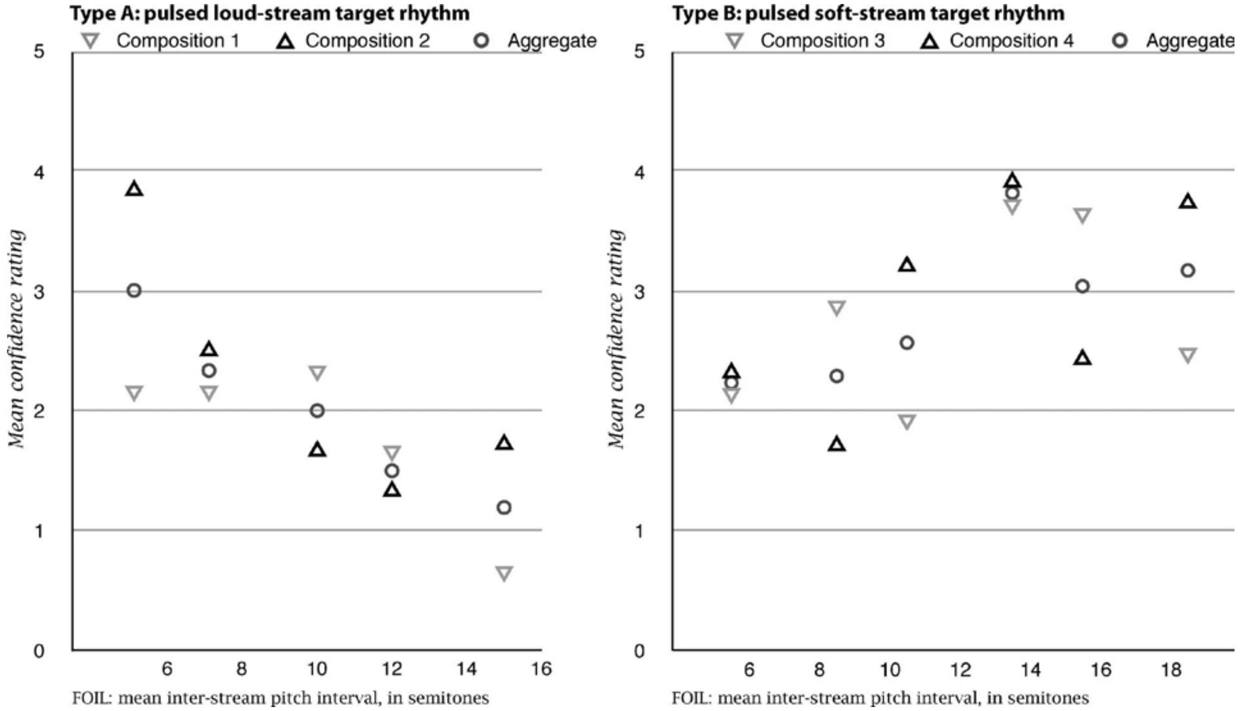


Fig. 13. Confidence ratings for target rhythms distinguished as loud streams (left, type A), and soft streams (right, type B).

and 10.1, as well as in the relationships between the two lowest distances and all other streaming distances (Tukey HSD [0.05] = 0.51, HSD [0.01] = 0.61, $p < 0.05$). Relationships among the three highest streaming distances (10.1, 12.1, and 15.1) were nonsignificant. For type A compositions, mean z-scores at loudness-streaming distances of 4.2 phons correlated negatively with the pitch-distance foil ($r(4) = -0.69$, $p < 0.01$).

As we expected, composition type B (Compositions 3 and 4) – where the target rhythm was only distinguished by its quietness – produced no significant results, even when pitch-streaming foil was small. This supports our common-sense hypothesis that loudness distinguishes notes in a hierarchically oppositional manner, with loud notes forming a distinct identity, but soft-notes being an indistinct part of the whole surface.

When target rhythms were distinguished by pitch, in either direction, the pitch-distance of the stream had a positive impact on pulse confidence, in most cases. Mean confidence ratings for compositions of types C and D are shown in Figure 14. The effect of our three pitch-streaming treatments for the "high-stream" target rhythms in Compositions 5 and 6 (type C) was pronounced and consistent, both in individual compositions, and throughout the aggregated type. The same effect is apparent in type D, where the pulsed rhythm was in a low voice (although a distortion is apparent in the 14-semitone treatment of Composition 7, possibly resulting from octave- and fifth-effects at that treatment, reinforcing a patterned relationship between the voices).

The "symmetrical" combination of all the type C and D compositions also showed a clear effect in some conditions. As will be seen below, the nature of those effects emerges and recedes from view according to some reliable factors, and we will speculate on how those factors affected our general hypothesis about pitch-streaming distance.

6.1 Detailed discussion of results for types A and B

Figure 15 examines trajectories in the confidence ratings of type A compositions for individual listeners whose reports differed least from the sample mean. Correlations were strong for all scores in Composition 2 ($r(3) = -0.97$ [DM], -0.96 [RJ], -0.96 [ET], -0.91 [BL], $p < 0.05$); and for the aggregate scores ($r(8) = -0.71$ [DM], -0.81 [RJ], -0.76 [ET], -0.84 [BL], $p < 0.05$). Correlations among individual participants' responses to Composition 1 were nonsignificant.

Whereas Figure 13 showed the negative correlation of the pitch-foil in the perception of target rhythms distinguished by loudness (type A), Figure 16 offers more detail about type A, showing effects of both pitch-streaming distance and loudness-streaming distance. The more complex graph hierarchizes the treatments of both target-rhythm and foil-rhythm variables, distinguishing the greater target-rhythm streaming distance of 4.2 phons (in dark bars) from the lesser distance of 0.8 phons (in faded bars). Difference in confidence between the loudness distance treatments is shown in each of

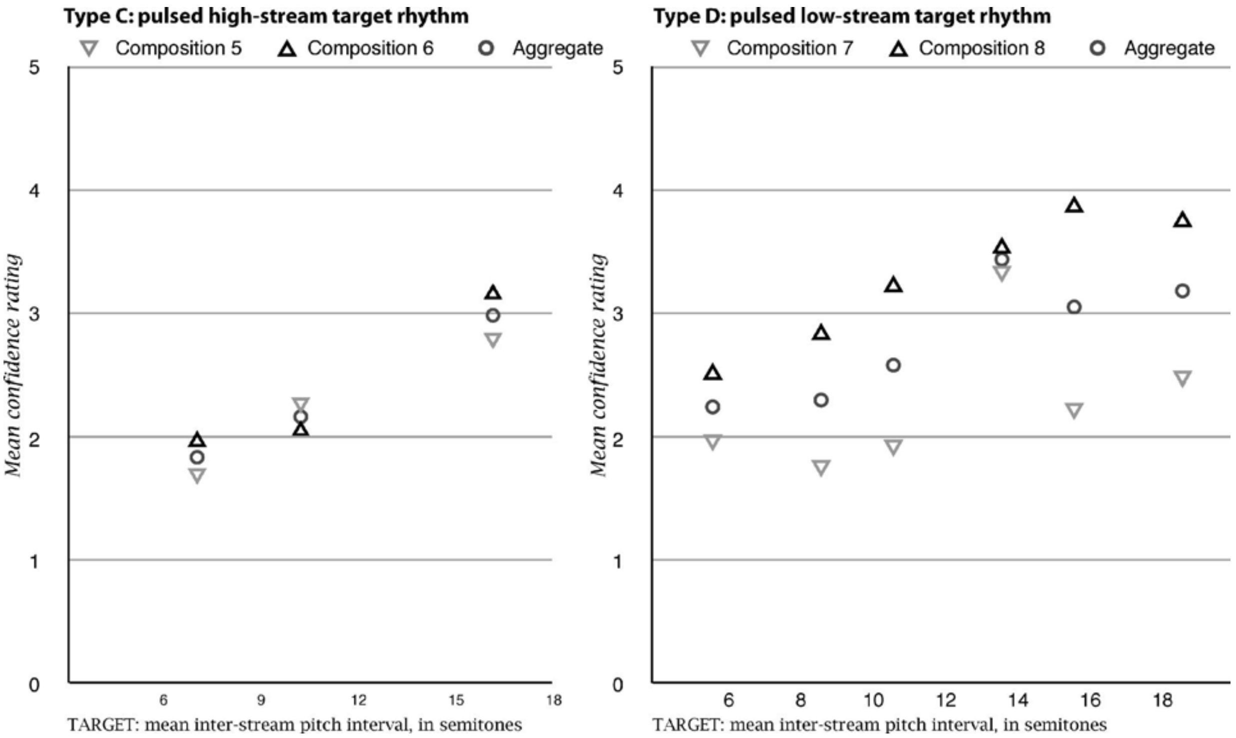


Fig. 14. Confidence ratings for target rhythms distinguished as high streams (left, type C, and low streams (right, type D).

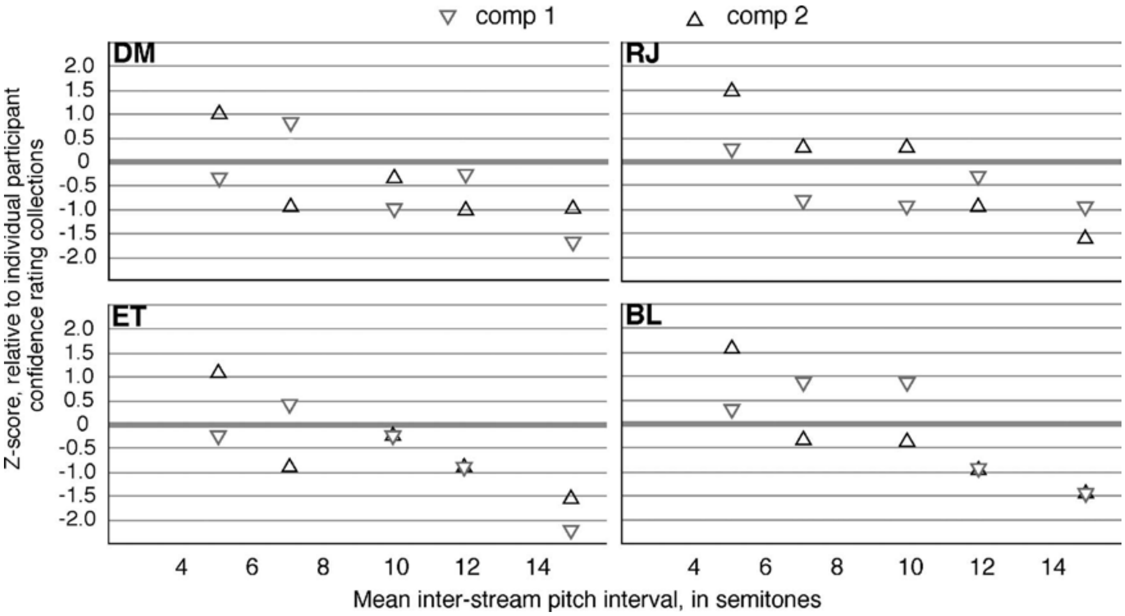


Fig. 15. Pulsed loud stream. Z-score conversions of type A pulse-confidence ratings for four representative participants. The target rhythm occurred at a 4.2-phon difference from its remainder.

four columns (left-to-right: small foil distance for Compositions 1 and 2, then large foil distance for Compositions 1 and 2). In each pair of z-score means (adjacent bar-pairs, distinguishing 4.2-phon from 0.8-phon treatments), pulse confidence is higher when the target rhythm differs by 4.2 phons (versions a and e), than when it differs by only 0.8 phons (versions f and g).

In the same figure, another strong correlation is shown in the progression between 5.1-semitone (left half of the graph) and 14.9-semitone (right half) mean pitch-streaming effects (a versus e with the 4.2-phon foil, and f versus g with the 0.8-phon foil). Variance across both factors was significant for Composition 1 ($F(3,17) = 6.49$, $p < 0.05$), and Composition 2 ($F(3,17) = 17.23$, $p < 0.001$).

By comparison, versions of type B (pulsed soft-stream target rhythm), displayed in the same way (Figure 17), demonstrate that neither the foil nor the target streaming treatments produced a significant effect across versions within the type. When the pitch-foil treatment was minimal (2.6 semitones), I did observe a significant negative correlation of confidence to target streaming distance (version – a versus version – e; $F(5,17) = 19.27$, $p < 0.0005$). However, this possible "target-rhythm quietness" effect was nonsignificant at larger pitch-streaming distances.

Comparison of specific sample pairs in type A compositions has yielded interesting results. In particular, it appears that the breaking point for the loud target rhythm – the moment at which the pitch-foil first begins to inhibit the target rhythm's perception – was much higher for Composition 1 than for Composition 2. (In other words, increased pitch contour in Composition 1 could not as easily disrupt the salient rhythm of accents, as it could in Composition 2.) Looking at the details of these particular compositions (Figure 18), as they relate to the outcome represented in Figure 13, we can make a few speculations about this distinction.

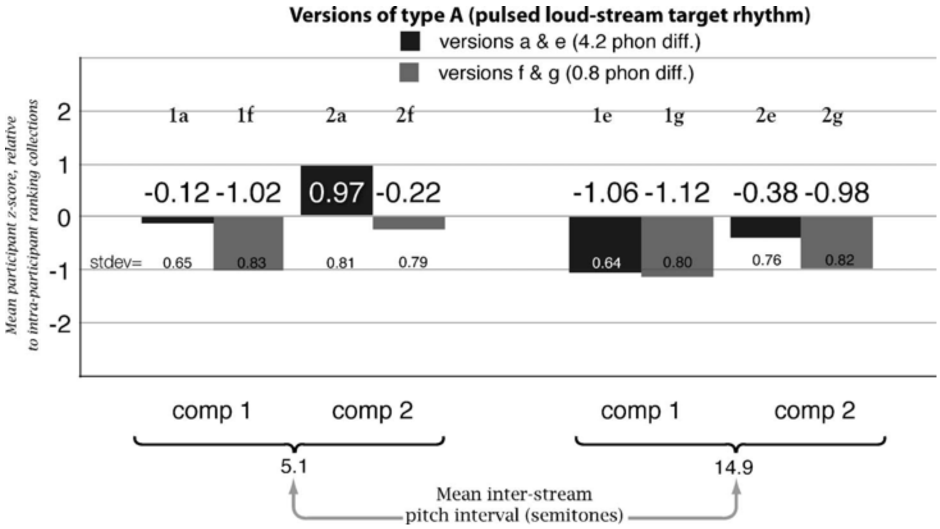


Fig. 16. Pulsed loud stream. Significant two-factor comparisons in four type A versions.

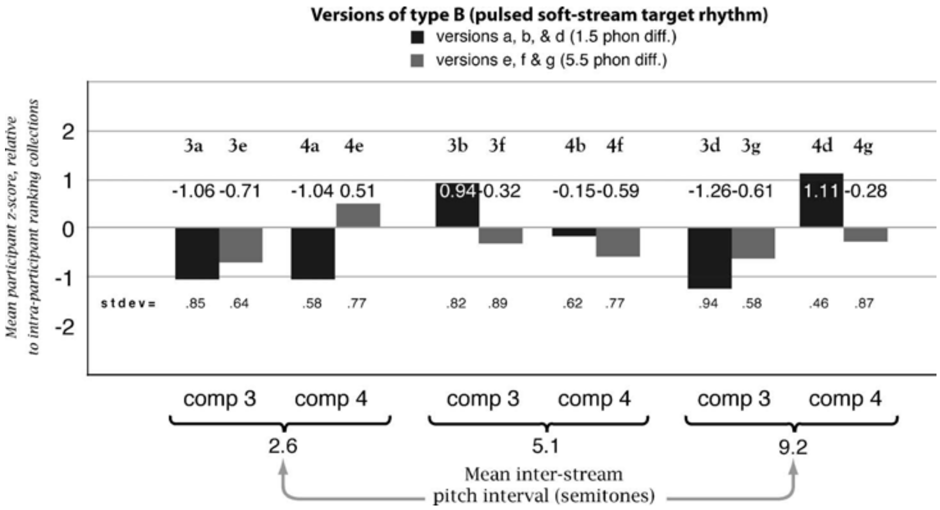


Fig. 17. Pulsed soft stream, two-factor comparison. Type B compositions (3 and 4) show no obvious overall effect of pitch-stream differences for any composition or version.

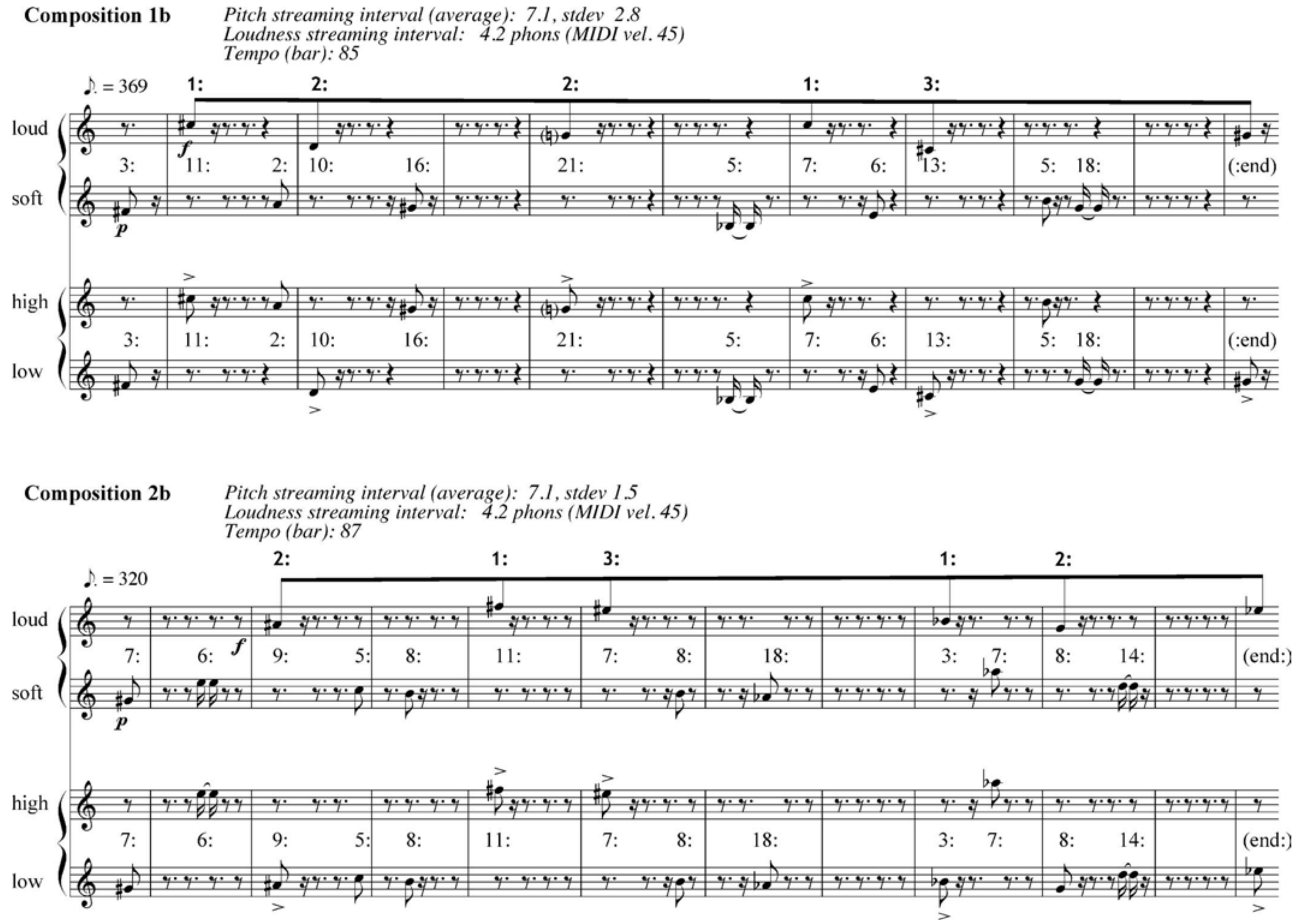

Composition 1b
Pitch streaming interval (average): 7.1, stdev 2.8
Loudness streaming interval: 4.2 phons (MIDI vel. 45)
Tempo (bar): 85
♪. = 369
loud
soft
high
low
1: 2: 2: 1: 3:
3: 11: 2: 10: 16: 21: 5: 7: 6: 13: 5: 18: (:end)
f
p
Composition 2b
Pitch streaming interval (average): 7.1, stdev 1.5
Loudness streaming interval: 4.2 phons (MIDI vel. 45)
Tempo (bar): 87
♪. = 320
2: 1: 3: 1: 2:
7: 6: 9: 5: 8: 11: 7: 8: 18: 3: 7: 8: 14: (end:)

Fig. 18. Comparison of Compositions 1 and 2 for distinctive streaming confounds.

In version b of both compositions, the ranges represented by the individual pitch streams are significantly contrasted: Composition 1 spans 6 semitones in the treble, and 8 in the bass, while Composition 2 streams span 4 and 5 semitones, respectively. A narrower range of notes in the foil streams may have facilitated the pitch-streaming percept in Composition 2, especially since its 3-note opening treble gesture persists within a small 2-semitone range for such a large portion of the example. In addition, Composition 2b establishes both voices in the "counterpoint" immediately in the first two notes, and before the onset of the loud stream; it may have been difficult for listeners to avoid "locking in" to those relationships, at the expense of the accent-based target rhythm. (Careful examination of the other compositions in the test will reveal numerous similar confounds. In allowing these, I gamble that the results will occasionally be less significant, but likewise, when relationships are shown, I believe them to be of broader potential importance, as observations about general musical experience.)

The "hierarchical opposition" hypothesis supported by our data also deserves more attention. In addition to the expected effect, in which loudness affected grouping differently than softness, we noticed a second, and somewhat unexpected, expression of the hypothesis in the results for type B compositions: counter-intuitively, when the loudness interval was small (so that the target rhythm's softness was *less* distinct in relation to the whole) confidence in pulse perception increased.

This suggests a different kind of asymmetry in event-group identity, than what was expected: even if listeners were theoretically able to hear a quiet group unto itself, it could be that their grasp of the target rhythm was nevertheless inhibited by steep differences between soft and non-soft notes. This represents a different take on the hierarchical opposition hypothesis, allowing for the theoretical possibility of soft groups. In this view, when we reduce the disparity of attention between the two group identities, and further reinforce attention to the soft group by some other factor (perhaps, in this case, the distinct reinforcement of consonant timespans), salient identity for quiet events is feasible. This question deserves a programme of separate tests that would be to isolate and examine those specific circumstances carefully.

6.2 Detailed discussion of results for types C and D

Since Compositions 5, 6, 7 and 8 were manifested identically in terms of measurements for the streaming variables, additional explorations of data are possible. By doubling sample sizes for some inquiries, we can observe stronger statistical relationships, while at the same time slightly broadening the range of musical conditions under which these relationships appear.

Figures 19 and 20 illustrate mean participant responses to all type C & D compositions in which the foil rhythms were distinguished in 4.2-phon streams. At pitch-streaming distances of 6.9, 9.9, and 15.9 semitones, the relationship among versions of Composition 5 was especially strong ($r(1) = +1.00$, $p < 0.01$; $F(2,34) = 8.89$, $p < 0.001$). The between-pairs comparison was significant for all relationships involving the highest streaming distance in Composition 5 (6.9 semitones versus 15.9 semitones: Tukey HSD = 0.95, $p < 01$; 9.9 semitones versus 15.9 semitones: HSD = 0.95, $p < 0.05$) and Composition 6 (6.9 s versus 15.9 s: HSD = 0.95, $p < 0.01$; 9.9 s versus 15.9 s: HSD = 0.95, $p < 0.05$), but not for the relationship between the smaller treatments. In the aggregate of samples from Compositions 5 and 6, analysis of variance for the pitch-stream variable yielded similar between-samples results as did the compositions on their own, but at much greater power ($F(2,70) = 22.13$, $p < 0.0001$).

Type D compositions with the same treatments shared some properties with type C, showing the same effect of 15.9 semitones in Composition 7 ($F(2,34) = 14.67$, $p < 0.0001$; HSD[0.01] = 0.78, $p < 0.01$), and to a lesser extent, in Composition 8 ($F(2,34) = 3.85$, $p < 0.05$,) with nonsignificant distinctions involving the lower streaming variables.

When the target rhythm streamed at distances of 9.9, 12.5, and 15.9 semitones with a 0.8-phon foil, the correlations of means were sometimes weak or non-existent. In Compositions 5 and 6, inter-sample variance was both nonsignificant and non-correlating, so I did not study the aggregate in these conditions. In Composition 7, participant reports for 9.9-semitone and 15.9 semitone variants were significantly distinct ($r(1) = +1.00$, $p < 0.05$; $F(2,34) = 5.21$, $p < 0.05$; Tukey HSD = 0.99, $p < 0.01$). The correlations in Composition 8 were not independently conclusive, but in the aggregate of

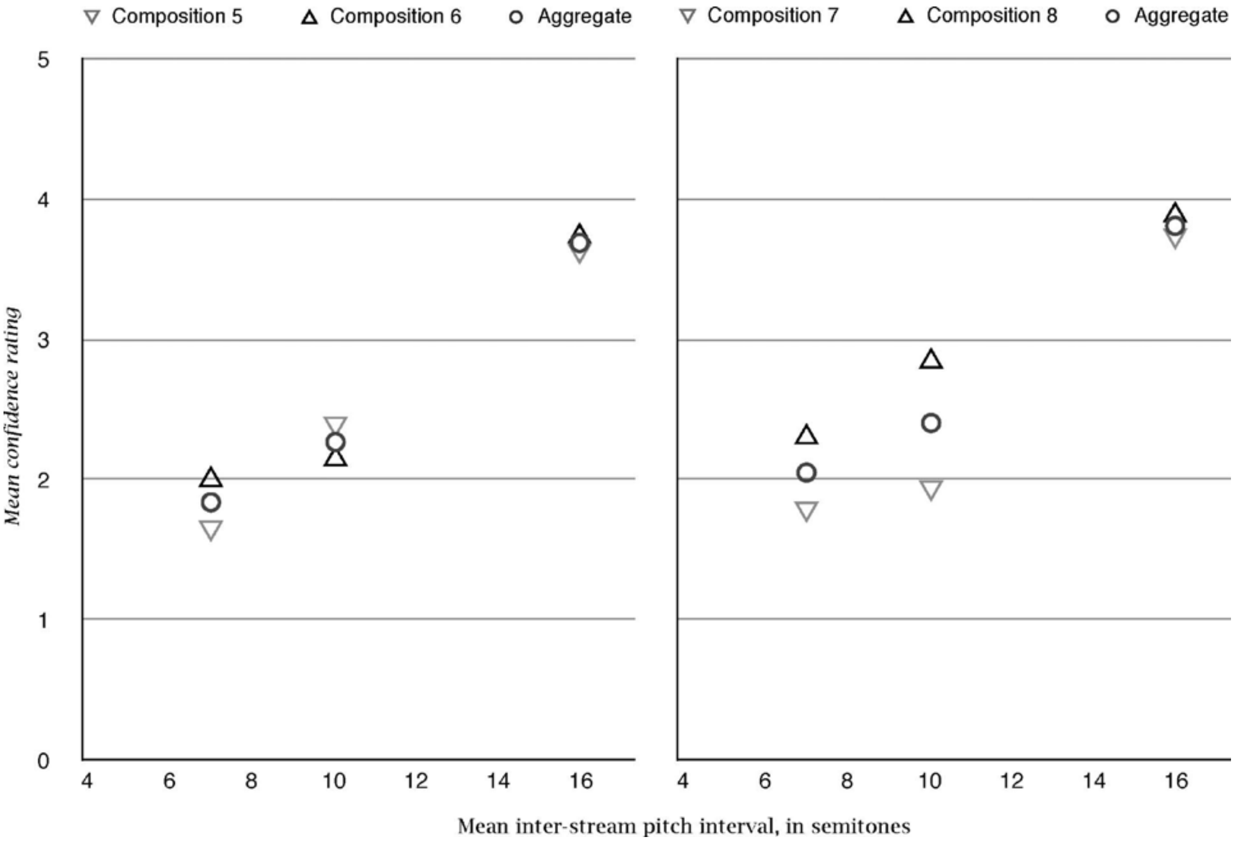


Fig. 19. Confidence ratings for target rhythms distinguished by pitch, with a loudness foil of 4.2-phons. Pulsed high streams (Compositions 5 amd 6) and pulsed low streams (Compositions 7 and 8).

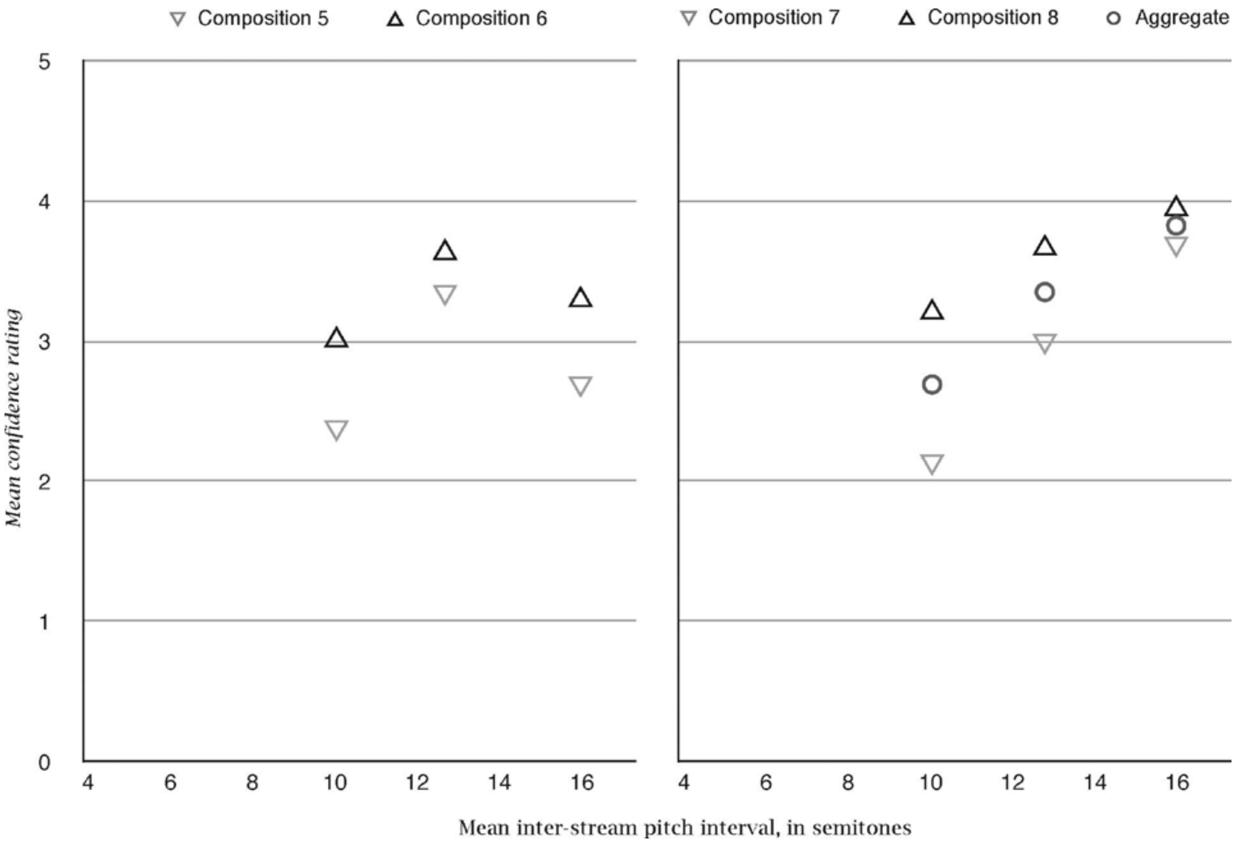


Fig. 20. Confidence ratings for target rhythm distinguished by pitch, with a loudness foil of 0.8-phons. Mean confidence ratings for pulsed high streams (Compositions 5 and 6) and pulsed low streams (Compositions 7 and 8).

Composition 7 and 8 (all of type D), the correlation was strong ($r(1) = +1.00$, $p < 0.05$) the overall significance was (like that of Composition 7 alone) limited to relationships between the smallest and largest streaming conditions ($F(2,70) = 5.98$, $p < 0.005$; HSD = 0.63, $p < 0.01$).

According to the expectations expressed in the design, I also performed ANOVA on relationships between two conditions in each of the streaming variables for composition types C and D. Figures 21 and 22 show another hierarchy of relationships between the results of those independent variables when the target rhythm was in a high or low stream, respectively.

Figure 21 shows that the effect of the foil was minimal, or inconclusive, whenever a regular pulse was in the high voice (type C aggregate ANOVA: $F(2,70) = 1.45$, $p = 0.24$; Tukey HSD for 5b and 6b versus 5b and 6d was nonsignificant). (We can speculate here that increased connections among accented notes across pitch-distinctions did not significantly inhibit the rhythmic independence of the contrapuntal "voices".) The figure also illustrates two general relationships contrary to the hypothesis, which are apparent in a comparison of the left halves of Figures 19 and 20. First, when the high target rhythm's pitch distance was more pronounced (15.9 semitones), the effect of the loudness-streaming foil was the opposite of the main hypothesis; confidence for stimuli 5d and 5g – the 4.2-phon foil – were both rated with significantly greater confidence than 5g and 6g – the 0.8-phon foil (type C aggregate ANOVA: $F(2,70) = 21.68$, $p < .0001$). Second, when the pitch-distance of the high-streaming target rhythms was 9.9 semitones, the foil had no apparent effect on composition 5.[7] Finally, in the one-way analyses of 9.9-semitone targets in all type C compositions, I checked for effects of change in the foil streaming between 4.2, 2.5, and 0.8 phons, individually and collectively. This test reiterated the nonsignificance of the foil.

Overall results in versions b, f, d, and g of all type D compositions (Figure 22) were much stronger, showing significant (and consistent) correlating effects from both variables. Composition 7, whose relationships were stronger than other compositions under the 4.2-phon streaming condition (see Figure 19), were also stronger here ($F(2,34) = 25.69$, $p < 0.0001$); Composition 8 also showed significant effects ($F(2,34) = 6.71$,

[7] In fact, the apparent effect in Composition 6 – correlated negatively with the foil streaming.

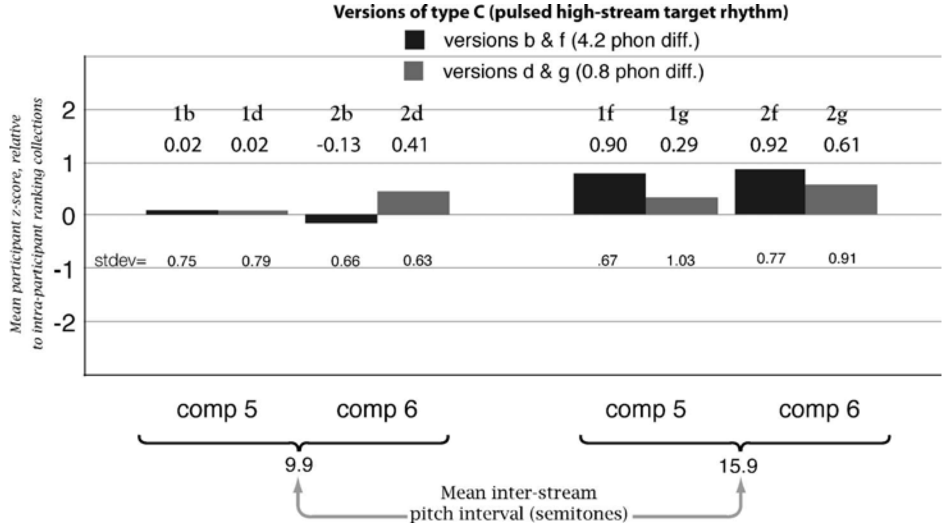


Fig. 21. Two-factor comparisons in four high-stream target rhythms.

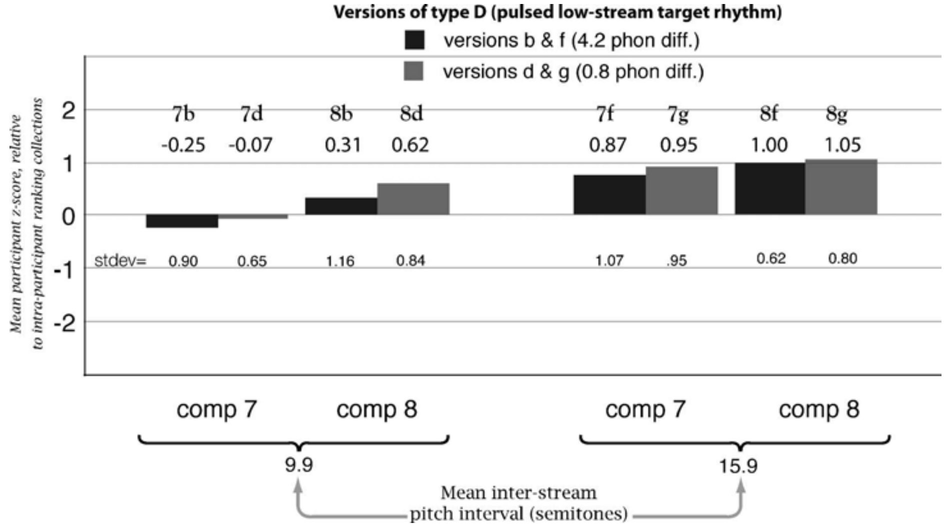
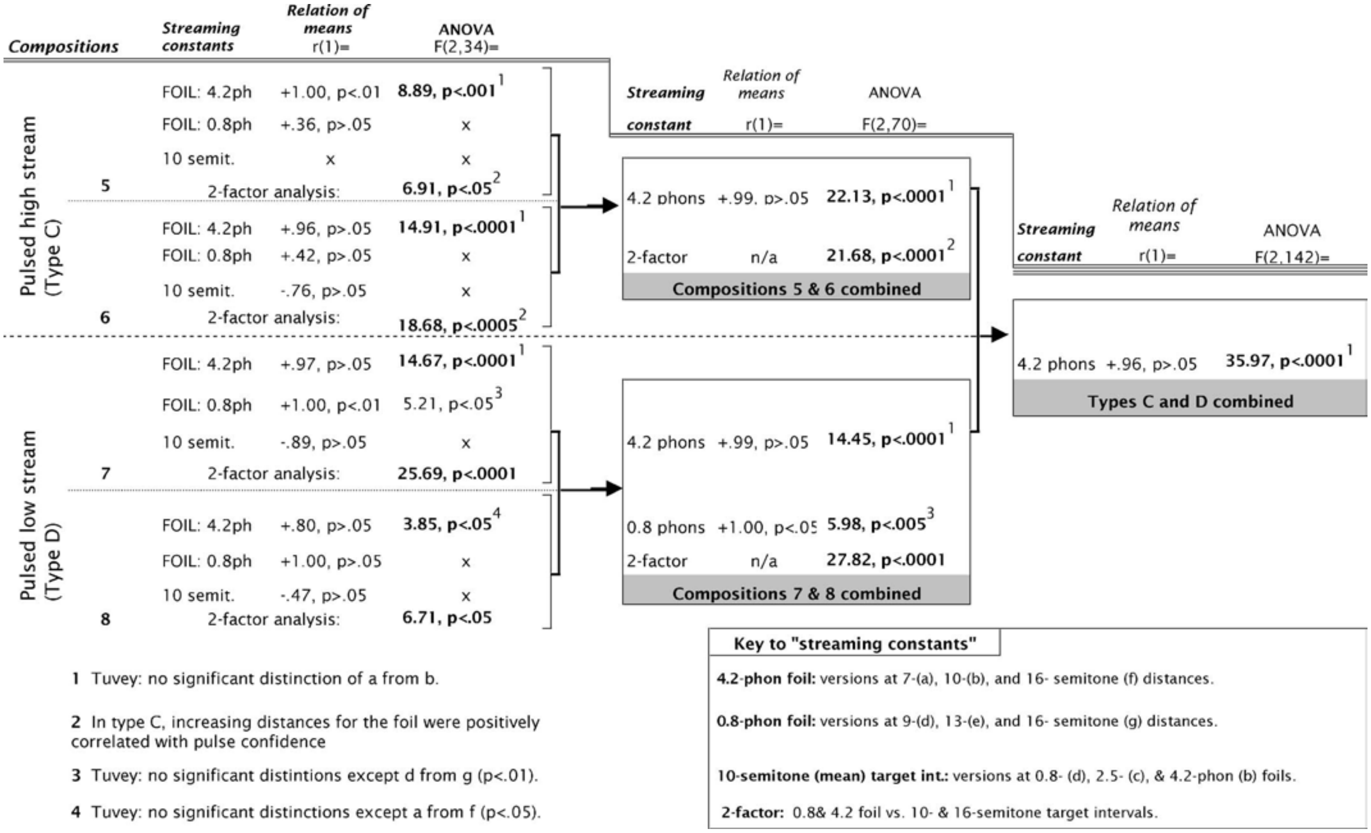


Fig. 22. Two-factor comparisons in four low-stream target rhythms.

$p < 0.05$), as did the aggregate sample ($F(2,34) = 27.82$, $p < 0.0001$).

In Figure 23, I display analyses of variance and correlation for all four compositions with pitch-distinguished target rhythms. Not all analyses showed conclusive relationships. First, the three distinct levels of effect for the loudness foil did not generate any significant results. (In this test, I constrained the loudness foil to distances that are common in musicians' interpretations of melodic material. In future trials, it may be worthwhile, instead, to examine a more sparse distribution.) Relationships in Figure 23 also suggest that listeners were much less consistent in distinguishing levels of separation for the target rhythm when the foil was smaller; however, recall that these pitch-streaming distances (at 0.8 phons) were exclusively high. This confirms, again, the intuitive idea that small distances *between* high or large streaming distances will have a lesser impact on the salience of the target rhythm than small distances between small or moderate streaming distances. Finally, I have noted that even in the most successful condition (the 4.2-phon foil) there were consistently weak or nonsignificant distinctions between versions a and b.

Nevertheless, whenever I paired weak relationships with strong ones, across compositions within the same type, or between the collected results of the types themselves, a larger analysis was justifiable, and that has allowed us to make a few powerful generalizations. Although results for the 0.8-phon foil conditions were nonconclusive with respect to Composition 8, and of limited power and extent in Composition 7



	Compositions	Streaming constants	Relation of means r(1)=	ANOVA F(2,34)=
Pulsed high stream (Type C)	5	FOIL: 4.2ph	+1.00, p<.01	**8.89, p<.001**[1]
		FOIL: 0.8ph	+.36, p>.05	x
		10 semit.	x	x
		2-factor analysis:		**6.91, p<.05**[2]
	6	FOIL: 4.2ph	+.96, p>.05	**14.91, p<.0001**[1]
		FOIL: 0.8ph	+.42, p>.05	x
		10 semit.	-.76, p>.05	x
		2-factor analysis:		**18.68, p<.0005**[2]
Pulsed low stream (Type D)	7	FOIL: 4.2ph	+.97, p>.05	**14.67, p<.0001**[1]
		FOIL: 0.8ph	+1.00, p<.01	5.21, p<.05[3]
		10 semit.	-.89, p>.05	x
		2-factor analysis:		**25.69, p<.0001**
	8	FOIL: 4.2ph	+.80, p>.05	**3.85, p<.05**[4]
		FOIL: 0.8ph	+1.00, p>.05	x
		10 semit.	-.47, p>.05	x
		2-factor analysis:		**6.71, p<.05**

Compositions 5 & 6 combined

Streaming constant	Relation of means r(1)=	ANOVA F(2,70)=
4.2 phons	+.99, p>.05	**22.13, p<.0001**[1]
2-factor	n/a	**21.68, p<.0001**[2]

Compositions 7 & 8 combined

Streaming constant	Relation of means r(1)=	ANOVA F(2,70)=
4.2 phons	+.99, p>.05	**14.45, p<.0001**[1]
0.8 phons	+1.00, p<.05	**5.98, p<.005**[3]
2-factor	n/a	**27.82, p<.0001**

Types C and D combined

Streaming constant	Relation of means r(1)=	ANOVA F(2,142)=
4.2 phons	+.96, p>.05	**35.97, p<.0001**[1]

[1] Tuvey: no significant distinction of a from b.

[2] In type C, increasing distances for the foil were positively correlated with pulse confidence

[3] Tuvey: no significant distintions except d from g (p<.01).

[4] Tuvey: no significant distinctions except a from f (p<.05).

Key to "streaming constants"

4.2-phon foil: versions at 7-(a), 10-(b), and 16- semitone (f) distances.

0.8-phon foil: versions at 9-(d), 13-(e), and 16- semitone (g) distances.

10-semitone (mean) target int.: versions at 0.8- (d), 2.5- (c), & 4.2-phon (b) foils.

2-factor: 0.8& 4.2 foil vs. 10- & 16-semitone target intervals.

Fig. 23. Overview of inferential statistics for target rhythms distinguished by pitch. In all combinations, the effect of increased pitch streaming was significant between b and f (4.2-phon foil) and – except in Composition 8 – between versions a and f.

($F(2,34) = 5.21$, $p < 0.05$), I found a significant correlation of means in the *aggregate* of type D, between the 9.9- and 15.9-phon foils. The aggregates of types C and D, ANOVA in both the trio of "4.2-phon foil" streaming conditions and on the 2-factor comparison of target distances and foil distances, we also achieved significant results. Finally, the significances found in all four sets of 4.2-phon foil treatments (12 samples in all, with related conditions), allowed us to offer a general analysis of the pitch-streamed targets regardless of their streaming positions as high or low. The results here ($F(2,142) = 35.97$, $p < 0.0001$) show a powerful overall distinction for the 15.9 streaming condition from its smaller associates.

7. Conclusion

In this study, I have asked, essentially, whether the property of "proportional consonance" is a viable and distinct percept for ametric sequences of timespans. I constructed event sequences that, in the foreground, were non-repetitive and non-consonant, but which in the middleground exposed or concealed a distinct set of consonant timespans, through controllable linear transformations. This style of control for the stimuli was risky, as it tested not only the main hypothesis, but the underlying premise that the exposure of a target-rhythm, by stream distance alone, would strengthen its identity *as* a rhythm. Because whole (combined- streams) groups of events, and their "unpulsed" overall successions of timespans, were the exclusive surface aspect in all treatments of the independent variable, I was uncertain as to whether group exposure would equate to a clear and distinct rhythmic percept. I have affirmed both the hypothesis and the basic experimental design concept in many of its instantiations.

The results of the experiment have shown, in general, what I could not demonstrate with brief ametric stimuli in the preliminary investigation (Figure 1). The relationships between stream exposure and pulse perception were often conclusive and unambiguous, and I found those relationships only in cases where the exposed stream was a pulsed target rhythm; no known factors in the test, other than the target rhythm, account for the correlation of segregation with confident perception of pulse. The participants heard timespans among event-subsets, essentially, as free-floating abstractions, rather than as subordinate parts of a linear metric whole; showing sensitivity to the difference between "consonant" and "dissonant" timespan proportions without benefit of metric reinforcement.

I was also able to show that effects for widening inter-stream pitch intervals were significant at small and moderate pitch distances, regardless of whether the target rhythm was in the upper or lower "voice". (At greater pitch distances, the effects for the foil variable were inconclusive.) I found that the "upper voice" target rhythms (type C: the pulsed high stream) affected pulse confidence, in a manner similar to those in the lower voice (type D: the pulsed low stream). The results here also show that more widely spaced loudness-foil variables may be necessary when target rhythms are distinguished by pitch; the parameter interactions in that condition were nonsignificant. In a wide variety of conditions, however – including the distinction of events by softness, and the more detailed difference between low-pitched and high-pitched streaming effects, this experiment leaves many questions available for further research.

Moreover, our main findings suggest unfamiliar and new questions about larger issues of temporal perception. Our data suggest the availability of a synchronic mode of temporal perception that coexists with a more commonly observed diachronic *expectancy*. While it is well known that listeners organize sound by expecting patterns within a linear continuum, and then either refuting or confirming them with actual auditory phenomena, this study confirms the simultaneous availability of a more collapsible temporal perception, in which the structures to be perceived cannot have been meaningfully assessed by their listeners until the moment of their completion. The data connected to timespan ratio consonance in this study confirm that listeners can gather to mind such freely collapsible structures, in at least some parts of their listening experience.

The existence of that kind of listening would hint, in a preliminary way, that listeners might foster other out-of-time structural differentiations and associations of a kind unfamiliar to systematic approaches in music theory. The structured temporal difference that we usually observe, when we begin with questions meter and expectancy, are *mediated* temporal differences. Deleuze (1994, p. 29) explains that "differences" assessed as intervals, between objects or markers, on a linear continuum, are not neutral assessments, but rather attempts to "rescue difference [for example, differences between consonance and dissonance in timespan ratios] from its maledictory state". The "rescue" in this view, is successful when differences "in themselves" are conceptually suppressed, and transformed into coherences, by expectancy. The views of Meyer (1959) and Kurth (1991, trans. of 1917), exemplify that process, as the tether-like aspect of an inevitable and progressive musical time produces a distinctive musical present; all events are subsumed in relation to a hierarchy of fixed reference points. Indeed, the ratio objects studied here are dependent (for their emergence as percepts) on a continuum of fixed reference points, but inasmuch as they determine a quality of "pulsedness" or "unpulsedness" out of time, the rhythm objects can be heard as differences that are plainly unencumbered. From behind a striated and unilateral regime of unified time, they emerge as individuated parts of a whole memory of temporal experience, producing what Deleuze might have called a more "dissemblant" and "oceanic" field of undifferentiation (Deleuze, 1994 [1968], 262).

References

Ashley, R. (2002). Do[n't] change a hair for me: the art of jazz rubato. *Music Perception*, *19*, 311–322.

Bregman, A.S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge: MIT Press.

Deleuze, G. (1994). *Difference and Repetition*, trans. by P. Patton. London: Athlone Press. [From *Difference et Repetition*. Paris: Presses Universitaires de France, 1968.]

Desain, P. (1992). A (de-)composable theory of rhythm perception. *Music Perception*, *9*, 439–454.

Desain, P. & Honig, H. (1999). Computation models of beat-induction: the rule-based approach. *Journal of Music Research*, *28*(1), 29–42.

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics*, *14*(1), 37–40.

Eck, D. (2001). A positive-evidence model for rhythmical beat induction. *Music Perception, 30*(2), 187–200.

Essens, P. (1995). Structuring temporal sequences: comparison of models and factors of complexity. *Perception & Psychophysics, 57*(4), 519–532.

Gregory, A. (1994). Timbre and auditory streaming. *Music Perception, 12*(2), 161–174.

Handel, S. (1993). The effect of tempo and tone duration on rhythmic discrimination. *Perception & Psychophysics*, *54*, 370–382.

Hasty, C. (1997). *Meter as rhythm*. New York: Oxford University Press.

Idson, W. & Massaro, D. (1976). Cross-octave masking of single tones and musical sequences: the effects of structure on auditory recognition. *Perception & Psychophysics, 19*, 155–175.

Jones & Boltz (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Large, E.W. & Jones, M.R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*, 119–159.

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Keller, P. & Burnham, D. (2005). Musical meter in attention to multipart rhythm. *Music Perception, 22*(4), 629–661.

Kurth, E. (1991). Foundations of Linear Counterpoint [1917] In L. Rothfarb (Ed. Translator), *Ernst Kurth: Selected Writings*. Cambridge: Cambridge University Press.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Narmour, E. (2000). Music expectation by cognitive rule-mapping. *Music Perception, 17*(3), 329–398.

Pfordresher, P. (2003). The role of melodic and rhythmic accents in musical structure. *Music Perception, 20*(4), 431–464.

Povel, D.-J., and Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–40.

Repp, B.H., London, J., and Keller, P.E. (2005). Production and synchronization of uneven rhythms at fast tempi. *Music Perception, 23*(1), 61–78.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysics In S.S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley.

Tekman, H.G. (1995). Cue trading in the perception of rhythmic structure. *Music Perception, 13*, 17–38.

Tekman, H.G. (1997). Interactions of perceived intensity, duration, and pitch in pure tone sequences. *Music Perception, 14*, 281–294.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Timmers, R., Ashley, R., Desain, P. & Heijink, H. (2000). The influence of musical context on tempo rubato. *Journal of New Music Research, 29*, 131–158.

Tukey, J.W. (1957). On the comparative anatomy of transformations. *Annals of Mathematical Statistics, 28*, 602–632.

Tukey, J.W. (1977). *Exploratory data analysis*. Boston: Addison-Wesley.

Velleman, P. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician, 47*(1), 65–72.