

Perception of Timbre Intervals

by

Christopher Wood



Music Technology Area, Department of Music Research
Schulich School of Music
McGill University, Montreal
August 2015

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Arts

© Copyright 2015
by
Christopher Wood

Abstract

An experiment tested participants' ability to recognize and reproduce timbre intervals. Listeners performed an analogy completion task in which they were presented with a change in either attack time or spectral centroid (A:B interval) and were instructed to adjust tone D in a second interval (C:D) until the two intervals were perceptually equivalent in terms of both magnitude and direction. The two intervals were related by transposition, where point A was transposed to C and D was a point selected by participants using a slider. There were **two** possible interval sizes (small and large) and **three** possible sizes of transpositions (small, medium, and large). Both intervals and transpositions were either upward or downward. Configurations of these directions were categorized as either congruent (both in the same direction) or incongruent (opposite directions). Responses in the spectral-centroid block showed significantly higher accuracy and consensus among participants than did the responses in the attack-time block. Listeners were likely to correctly reproduce interval directions in their responses, an effect that increased in large interval conditions and incongruent conditions. Participants responded with smaller than expected C:D intervals in congruent conditions and larger than expected intervals in incongruent conditions, suggesting a tendency to be biased in the direction of point B. This effect was seen in both blocks, but to a greater extent in the attack-time block. The effect increased with transposition size (i.e., distance from B to D), though less reliably for conditions with downward interval directions. Similarities found between trends and interactions in both blocks suggest that the same cognitive functions are used in the completion of each task, regardless of dimension.

Résumé

Nous avons testé la capacité des participants à reconnaître et à reproduire des intervalles de timbre. Les auditeurs effectuent une tâche d'achèvement d'analogie dans laquelle on leur présente un changement du temps d'attaque ou du centre de gravité spectral formant un intervalle A:B; ils doivent ajuster le deuxième son d'un autre intervalle C:D jusqu'à ce que les deux intervalles soient équivalents en grandeur et en direction du changement. Le deuxième intervalle est censé être une transposition du premier: le point A est transposé à C et l'auditeur doit trouver le point D approprié en ajustant un curseur. Il y a deux grandeurs d'intervalle (petite et grande) et trois grandeurs de transposition (petite, moyenne, grande). Les intervalles et les transpositions pouvaient être ascendants ou descendants. Les configurations de ces directions sont catégorisées comme congruentes (les deux dans la même direction) ou incongrues (directions opposées). Les réponses pour le centre de gravité spectral présentent une précision et consensus significativement plus élevés que celles pour le temps d'attaque. Les auditeurs reproduisent souvent la bonne direction de l'intervalle, un effet qui augmente dans les conditions avec des grandes intervalles et dans les conditions incongrues. Les participants répondent avec des intervalles C:D plus petits que prévus dans les conditions congruentes et plus grands que prévus dans les conditions incongrues, suggérant ainsi un biais vers le point B dans les réponses. Cet effet se produit pour les deux dimensions, mais est plus grand pour le temps d'attaque. Cet effet augmente également avec la grandeur de transposition (c'est-à-dire la distance entre le point B et le point D idéal), bien que de façon moins fiable pour les intervalles descendants. Des ressemblances trouvées entre les tendances et les interactions pour les deux dimensions du timbre suggèrent que les mêmes fonctions cognitives sont en jeu dans la réalisation de la tâche, indépendamment de la dimension testée.

Acknowledgments

First and foremost I would like to thank my supervisor Stephen McAdams for his tireless patience and support throughout the completion of this paper. Secondly, I wish to thank Bennett Smith for programming the experimental interface, as well as assisting in the installation of PsiExp. Thanks go out to Daryl Cameron as well for all of his help in the wake of hard drive failures and other technical issues. I would like to thank the members of the Music Perception and Cognition lab that provided help and feedback during the loudness matching and piloting phases of the experiment: Cecilia Taher, Eddy Kazazis, Rachel Kim, Kai Siedenburg, Yinan Cao, Meghan Goodchild and especially Chelsea Douglas for her additional help with LaTeX formatting.

Author contributions

As the author of this thesis, I participated in designing the experiment and preparing the stimuli. I was also responsible for running the subjects, performing statistical analyses of the data, interpreting the results, and writing and editing the thesis. As the thesis supervisor, Stephen McAdams provided laboratory equipment, oversaw the experimental design, and guided data analysis and interpretation of the results.

Contents

1	Introduction	1
2	Methods	5
3	Results	10
3.1	Directional/contour accuracy	10
3.2	Response analysis	15
3.3	Effects of point placement and congruency	17
3.4	Interactions	21
4	Discussion	22
4.1	Directional/contour accuracy	22
4.2	Response analysis	23
4.3	Effects of point placement and congruency	23
4.4	Interactions	25
5	Conclusion	47
	References	49
	Appendix	51

List of Figures

2.1	Point placement categories	8
2.2	Experimental interface	9
3.1	Percentages of attack-time block responses with correct interval directions, contrasted by interval size	11
3.2	Percentages of attack-time block responses with correct interval directions, contrasted by congruency	12
3.3	Percentages of spectral-centroid block responses with correct interval directions, contrasted by interval size	13
3.4	Percentages of spectral-centroid block responses with correct interval directions, contrasted by congruency	14
3.5	Comparison of the histograms of all responses in the attack-time and spectral-centroid blocks	16
3.6	Congruent underestimation and incongruent overestimation	18
4.1	Transposition Direction \times Transposition Size interaction in both attack-time and spectral-centroid blocks	27
4.2	Interval Direction \times Interval Size interaction in both attack-time and spectral-centroid blocks	29
4.3	Transposition Size \times Interval Size interaction in the attack-time block.	30
4.4	Interval Direction \times Transposition Size interaction in the spectral-centroid block. . .	31
4.5	Interval Direction \times Transposition Size \times Interval Size in the attack-time block. . . .	33
4.6	Interval Direction \times Transposition Size \times Interval Size in the spectral-centroid block.	34
4.7	Conditions 13 and 19 in the attack-time block	35
4.8	Conditions 13 and 19 in the spectral-centroid block	36
4.9	Conditions 15 and 21 in the attack-time block	37
4.10	Conditions 15 and 21 in the spectral-centroid block	38
4.11	Interval Direction \times Transposition Direction \times Transposition Size in the attack-time block.	40
4.12	Transposition Direction \times Transposition Size \times Interval Size in the attack-time block.	41
4.13	Condition 17 in both attack-time and spectral-centroid blocks	42

4.14	Transposition Direction×Transposition Size×Interval Size in the spectral-centroid block.	43
4.15	Interval Direction×Transposition Direction×Transposition Size in the spectral-centroid block	45
4.16	Conditions 19 and 20 in the spectral-centroid block	46
5.1	Loudness matching results	52

List of Tables

2.1	Conversions between 1–10 scale and attack time (ms) and spectral centroid (harmonic rank)	7
3.1	Summary of the results in section 3.3	20
5.1	List of conditions and their properties	51
5.2	Table of participants' primary instruments	53

Chapter 1

Introduction

When we think of timbre as a form-bearing dimension (McAdams, 1989), one issue that invariably arises is the notion of intervals, particularly, the extent to which we can draw an analogy between intervals of pitch and timbre. The first step is to define exactly what constitutes a timbre interval. McAdams, Winsberg, Donnadieu, De Soete & Krimphoff (1995) conducted a study in which participants were asked to rate the dissimilarities between simulations of common orchestral instruments (played at the same pitch and loudness) using a numerical scale. Using these data, they performed a multidimensional scaling (MDS) analysis to assign a set of coordinates to each timbre, such that the participants' dissimilarity ratings were represented as distances between these sets of coordinates. From this information, they were able to diagram the relationships between instrumental timbres by plotting them as points in a 3D “timbre space”. Timbre intervals can thus be conceived as the relations (distances) between timbres (points) in this perceptual space.

Caclin, McAdams, Smith & Winsberg (2005) investigated this space as well, instead using synthesized tones to confirm the findings of McAdams et al. (1995). They found that spectral centroid, attack time, and jaggedness of the spectral envelope were the three most contributing factors that determine where a particular timbre lies in the space.

Attack time describes the initial portion of a stimulus—it is the duration required for a stimulus to reach its peak loudness starting from 0 dB. Spectral jaggedness, also referred to as even harmonic attenuation, describes the relative amplitudes of the even harmonics compared to the odd ones. Finally, spectral centroid refers to the weighted average, or center of gravity, of the spectrum of the stimulus, which is calculated based on the relative amplitudes of all harmonics present. For example, a tone with relatively intense high-frequency harmonics will have a high spectral centroid, whereas a tone with mostly low harmonics will have a low spectral centroid. A precise description of the calculation is given by Caclin et al. (2005). Spectral centroid is often used as a metric for characterizing “bright” vs. “dull” timbres (Iverson 1995; Russo & Thompson 2005a). In the present study, we express spectral centroid in harmonic rank, which is a multiple of the fundamental frequency f_0 of a tone (i.e., a spectral centroid of 3.5 means that the weighted center of the spectrum occurs at $3.5f_0$, or halfway between the third and fourth harmonics).

Other factors, such as spectral flux and specificities, have been shown to influence perception of timbre relationships as well, especially among orchestral instruments (McAdams et al. 1995, Krumhansl 1989). Spectral flux refers to the degree to which the amplitudes of harmonics evolve over time, whereas specificities are unique, distinguishing qualities of certain timbres that can contribute to perceptual dissimilarity ratings. In order to remove specificities as a compounding factor in the present experiment, we decided to use only synthesized tones with carefully controlled harmonics and amplitudes. We also chose to focus only on the dimensions of spectral centroid and attack time, as these have been shown to be particularly influential dimensions of timbre (McAdams et al. 1995, Iverson 1995, Krumhansl 1989). Another important finding of Caclin et al. (2005) was that the attack time dimension of the timbre space provided the strongest correlation when it incremented logarithmically. In other words, when comparing stimuli with longer attack times, a greater difference is required in order to produce a perceptually equivalent distance.

By contrast, spectral centroid (expressed as harmonic rank) was found to be perceptually linear. However, it is important to keep in mind that the range of centroids used in their stimuli is significantly smaller than the one used for this experiment (3.0–4.5 compared to 1.5–15.0). For this reason, we hypothesize that spectral centroid is in fact perceptually logarithmic as well, because it shares the same sensory continuum used to encode frequencies and detect pitch height. Pitch and spectral centroid have been shown to interact in interesting ways. Russo & Thompson (2005a) found that melodic intervals between pitches were perceptually larger when accompanied by a congruent change in brightness (i.e., an ascending interval from dull to bright timbre, or descending interval from bright to dull). Musicians are able to faster and more accurately recognize changes in pitch across two tones if they contain a change in timbre as well (Krumhansl & Iverson 1992, Melara & Marks 1990a), although non-musicians are more easily distracted by timbral differences (Pitt 1994). These findings suggest that the brightness of a tone has a non-negligible effect on the perception of its pitch. It would be interesting to see the extent to which the similarities between these dimensions carry over into the realm of timbre interval perception. By using a logarithmic scale in our experiment we can indirectly shed some light on whether spectral centroid, like pitch, is perceived logarithmically.

The 3D timbre spaces investigated by McAdams et al. (1995) and Caclin et al. (2005) have many other applications as well. They serve as models of the perceptual representation of timbre and can be used to predict how timbre differences affect auditory stream segregation. Iverson (1995) showed that participants could more easily identify the contour of a target melody interleaved with a distractor melody when the two had sufficiently different timbres. In a similar study, Bey & McAdams (2003) found that timbre differences and large pitch differences allowed subjects to more accurately recognize the interleaved target melody when a comparison melody was subsequently played back without the distractor sequence. Timbre differences also improve listeners' ability to implicitly learn the probabilities of sequential relations among timbres in Klangfarbenmelodien (Tillman & McAdams, 2004). Thus, it is clear that timbre plays an important role in our ability to both attend selectively to melodies and commit them to memory.

Composers in the 20th century began exploring the ways in which melodies and timbre could interact as well. In his *Five Pieces for Orchestra* (Op. 16), Schoenberg used subtle timbral changes to evoke changes in pitch when in fact there are none (from an orchestrator's perspective), a technique that he later referred to as *Klangfarbenmelodien*. Such seamless blending was not always the goal, however—other composers used stark and abrupt timbral changes to create a more pointillistic texture (e.g., Webern's *Symphony Op. 21*). The difference between these two styles is relatively simple to understand in terms of auditory scene analysis. It is a matter of whether the timbres used are sufficiently different to induce stream segregation, or whether they are similar enough to be perceived as one melody. Understanding the various timbral relationships within a timbre space will empower composers and help them create the specific textures that they desire. In this regard, research on timbre intervals has particularly exciting implications as well—imagine if one were able to transpose a melody over timbre instead of pitch!

However, the issue of transposing timbre intervals within this space has yet to be fully explored. If we define a timbre interval as a vector representing the distance between two points in the space, then transposition of this interval is defined as translation of the vector, reproducing the same amount of change along each dimension of the timbre space. The extent to which people can recognize or reproduce these transpositions is not yet clear. Two studies have investigated this issue by framing the transposition as an analogy (interval A:B is transposed to C:D). Put in these terms, participants showed a slight preference for transpositions that maintained magnitude and direction, when asked to select the D timbre that they felt best completed the analogy A is to B as C is to D (Ehresman & Wessel, 1978; McAdams & Cunibile, 1992).

An interesting point to consider here is that because these intervals consist of both a magnitude and a direction, it may be the case that participants are more adept at reproducing one than the other. Contour has been shown to play a significant role in sequence perception—participants are able to recognize familiar melodies whose contours are transcribed into loudness or brightness (McDermott, Lehr & Oxenham 2008).

Melara & Marks (1990a) conducted an experiment in which participants were asked to perform a speeded classification of changes in pitch, timbre, or loudness. The task was to assess whether or not a particular dimension changed across two sequentially presented tones, while ignoring any changes in a distractor dimension. Performance increased when the two dimensions shared the same contour (e.g., low pitch, dull timbre to high pitch, bright timbre), a phenomenon the authors refer to as *redundancy gain*. This effect was also seen for timbre and loudness (low pitch, low loudness to high pitch, high loudness). In a series of similar follow-up experiments, they explored not only auditory dimensions, but spatial and linguistic ones as well. Redundancy gain was found for pitch in conjunction with the visual words HI and LO, vertical position with visual words, spoken words with vertical position, and finally pitch with spoken word (Melara & Marks 1990b). The findings suggest that some type of cognitive contour schema is responsible for recognizing and processing analogous contours across all dimensions.

However, it remains unclear whether contour or magnitude plays a larger role in determining subjects' preference for a particular D timbre in an analogy completion task, or whether they are equally important. In this study, we aim to delve further into this issue by eliminating as many confounding factors as possible and simply working within a 1D timbre space of either spectral centroid or attack time. The experiment consists of two parts—one for each parameter—and uses a similar A:B::C:D analogy format. However, participants are allowed to choose the D timbre using a continuous slider. This provides a more detailed look at how people perceive timbre intervals, because they are not restricted by a multiple-choice task. In addition, we hope to gain insight into the mechanisms at play in the performance of this task, namely whether participants rely more heavily on precise value judgments or simple contour assessments. In the case of the former, we would expect to see a high level of concurrence among the responses, regardless of whether or not they happen to fall on our predicted point D. In the case of the latter, however, we would see lower levels of agreement among participants, but a consistent pattern of interval direction preservation (e.g., if A:B is an increasing interval, so is C:D).

Chapter 2

Methods

Participants

There were 22 participants (5 males and 17 females), aged 19–32 years with an average age of 24 years. All participants were McGill University music students who had completed at least their first-year (200-level) courses. No participants reported any hearing problems.

Stimuli

The stimuli used in this experiment were generated using additive synthesis in PureData software (Pd-extended version 0.43.4, <http://puredata.info>) at a sampling rate of 44,100 Hz. Each stimulus contained 40 harmonics with a fundamental frequency of 311.13 Hz (E \flat 4).

The amplitudes of each successive harmonic decreased linearly in dB, and the spectral centroid was adjusted by altering the slope of this line. In the spectral-centroid portion of the experiment, each tone had a 15-ms linear rise time, followed by a 400-ms plateau and 100-ms exponential decay to –60 dB relative to the maximum of the amplitude envelope. Because attack time remained unvaried in this block, all stimuli shared this 515-ms structure. The spectral centroids of the stimuli presented directly to the participants (tones A, B, and C) ranged from 2.38 to 9.46 (expressed in harmonic rank). However, the slider that participants used to lock in their responses permitted them to choose any centroid from 1.5 to 15 for tone D. Spectral centroid varied logarithmically as a function of the slider position.

The onsets of all stimuli were simple linear ramps. In order to maintain the perceptual duration of the stimuli when changing the attack time, the plateau of each tone was adjusted according to the formula $t_2 = 412 - 0.8t_1$ (where t_1 is the attack time and t_2 is the sustain duration, in ms). Caclin et al. (2005) derived this equation for a similar purpose in their experiment—they found that the perceived duration of their stimuli remained unaffected so long as the attack and sustain times were consistent with the formula. The decay portion in our stimuli was always a 100-ms exponential decrease to –60 dB, regardless of any changes in attack or sustain time.

In the attack-time block of the experiment, each stimulus had a fixed spectral centroid of 3.96. Stimuli directly presented to the participants had attack times ranging from 20 ms to 152 ms. The slider for adjusting the attack time of tone D ranged from 10 ms to 300 ms and increased logarithmically. Taking into account the sustain time adjustments, this means that the effective durations of the stimuli in this block ranged from 514 to 572 ms.

All stimuli were matched for loudness prior to the experiment. To accomplish this, a small experiment was created in which six equidistant points on the slider (−2, 1, 4, 7, 10, and 13 in Table 2.1) were chosen to be compared to a reference tone with a spectral centroid of 3.96 and an attack time of 15 ms. We chose a centroid of 3.96 because it is close to the center of our scale while also falling well within the range of 3.0–4.5 used by Caclin et al. (2005). Therefore, it is reasonable to assume that this spectral centroid should not produce any unwanted effects, neither here nor in the attack-time block of the main experiment.

Seven members of the Music Perception and Cognition lab matched each of the six points individually to the reference using an interface similar to that in Fig. 2.2. It contained a slider to adjust the loudness of the comparison stimulus and two buttons to replay either the comparison or reference as desired. Responses had an average standard deviation of 1.45 for attack time and 1.98 for spectral centroid. A more detailed summary of the results can be found in the Appendix (Fig. 5.1).

Whereas the attack-time results were not significant enough to warrant loudness correction, we derived the following linear regression for use in the spectral-centroid block:

$$\Delta dB = -6.35 \ln(C) + 7.76$$

where C is the spectral centroid of the tone at that particular point on the slider and ΔdB represents the amplitude change to be applied. Note that it is necessary to take the logarithm of C , because spectral centroid increased exponentially as a function of slider position. Once this formula was applied, every point on the slider had equivalent perceived loudness.

Procedure

The task in this experiment was to complete an analogy—A is to B as C is to D. In each condition, participants were presented with the tones A, B, and C, which varied in either attack time or spectral centroid, depending on the parameter being tested. Participants were instructed to adjust the on-screen slider “until the change from C to D sounds analogous to the change from A to B, both in terms of size and direction. That is, if A to B sounds like a decrease in brightness, you should adjust D so that C to D sounds like an equivalent decrease in brightness.” At the end of each condition, whatever attack time or spectral centroid that the participant had assigned to tone D was recorded.

In the experimental design, we used two different A:B interval sizes (small and large) and three different transposition sizes (small, medium, and large distance between A and C). These sizes were defined in terms of logarithmic differences in attack time and spectral centroid. For

the sake of simplicity and for our own analyses, we created a linear 1–10 scale to represent the placement of the points and the intervals between them. Table 2.1 displays the actual spectral centroids and attack times of these points in harmonic rank and ms, respectively. Points 1–10 represent stimuli presented directly to participants, whereas the **three** points above and below this represent the extra room they were given for their responses. The formulas for converting between the linear scale and the actual values are given by:

$$C = 1.5 \left(10^{(n+2)/15} \right) \quad \text{and} \quad A = 10 \left(30^{(n+2)/15} \right)$$

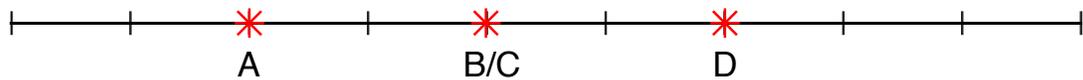
where C is the spectral centroid in harmonic rank, A is the attack time in ms, and n is the number of the point on the table. Small, medium, and large intervals were defined as **two, three, and four** steps on the linear 1–10 scale, respectively. This applied to both interval sizes as well as transposition sizes, although interval sizes could only be small or large.

Point	Attack time (ms)	Spectral centroid
-2	10	1.5
-1	13	1.749
0	16	2.039
1	20	2.377
2	25	2.771
3	31	3.232
4	39	3.768
5	49	4.393
6	61	5.122
7	77	5.972
8	97	6.962
9	121	8.118
10	152	9.464
11	191	11.035
12	239	12.865
13	300	15

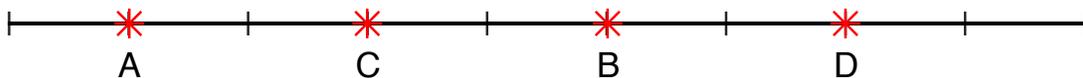
Table 2.1 The attack time and spectral centroid values for each point on the 1–10 scale, in ms and harmonic rank, respectively.

Both intervals and transpositions could be either upwards or downwards. All possible combinations of these parameters were considered, yielding a total of 24 interval conditions (2 interval directions \times 2 transposition directions \times 2 interval sizes \times 3 transposition sizes). If the interval direction is the same as the transposition direction, the condition is said to be congruent. If the two directions are opposite, the condition is incongruent.

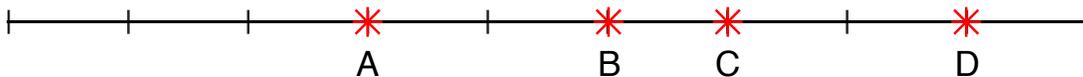
The interval and transposition sizes that we used naturally gave rise to three different subcategories (concatenated, interleaved, and separated) based on the relative positions of the points. Concatenated conditions were those in which the A:B and C:D intervals were placed directly next to one another with no space in between ($B = C$ in the congruent case, $A = D$ in the incongruent case). In interleaved conditions, C fell between A and B if the condition was congruent, whereas D fell between A and B if it was incongruent. In separated conditions, a gap existed between the the A:B interval and the C:D interval, regardless of congruency. Fig. 2.1 provides visual examples of these three categories.



(a) A concatenated condition. This occurs when the transposition size equals the interval size.



(b) An interleaved condition. This occurs when the transposition size is smaller than the interval size.



(c) A separated condition. This occurs when the transposition size is larger than the interval size.

Fig. 2.1 Examples of the three types of point placement: a) concatenated, b) interleaved, and c) separated.

The interface consisted of four circles representing each of the four tones, A, B, C and D (see Fig. 2.2). Below this was a slider that could be used to adjust either the attack time or the spectral centroid of tone D, whichever parameter was being tested.

To avoid biasing the participants, the slider was left blank in each condition until it was clicked on, at which point the diamond-shaped cursor would appear. Participants could replay the A:B pair or the C:D pair as many times as desired by clicking the A or C buttons. They could also play the entire A:B::C:D sequence using the ‘play all’ button. This button had to be clicked at least once before the experiment would allow the participant to register their rating and proceed to the next condition. This procedure ensured that for each condition, the participant heard the full sequence at least once. In all sequences, there was a 600-ms silence between successive tones.

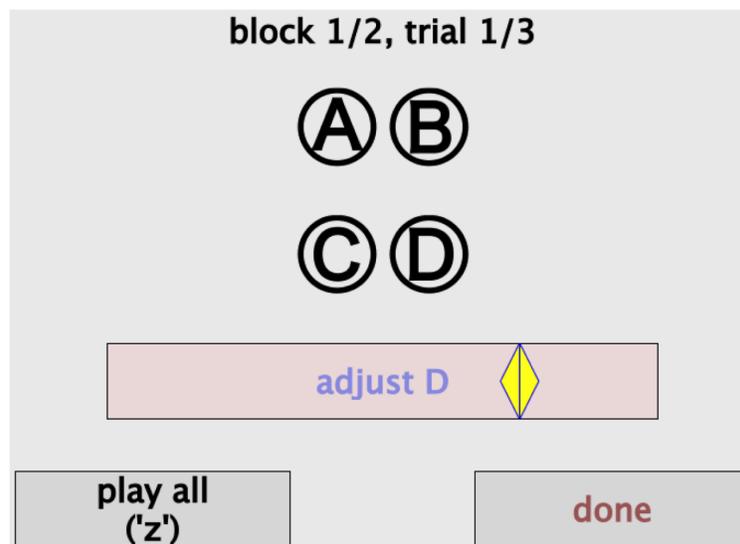


Fig. 2.2 The experimental interface. The ‘adjust D’ region was left blank at the start of each condition. Clicking in this region caused the cursor to appear where the participant had clicked.

The experiment began with three practice conditions for the given block (attack time or spectral centroid), followed by 24 experimental conditions. The same structure was then presented for the other block (spectral centroid or attack time). All conditions were presented in random order within their respective block. Odd-numbered participants began with the attack time task, and even-numbered participants began with spectral centroid. After both blocks were completed, participants were debriefed and given a questionnaire that assessed their hearing, level of musical training, and music listening habits. A summary of the participants’ primary instruments is given in Table 5.2 in the Appendix.

Apparatus

The experiment was carried out in an IAC model 120act-3 double-walled soundproof booth (IAC Acoustics, Bronx, NY). Participants used a Mac Pro 5 computer running OS 10.6.8 (Apple Computer, Inc. Cupertino, CA). All audio was synthesized with PureData software (Pd-extended version 0.43.4) and amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA). It was transmitted through Sennheiser HD280 Pro earphones (Sennheiser Electronic GmbH, Wedemark, Germany). Sound levels of the stimuli ranged from 65.5 dB to 69.6 dB, as measured with a B&K type 2250 sound level meter and a B&K type 4153 Artificial Ear with a flat-plate coupler. The experimental interface was created using the PsiExp computer environment (Smith, 1995). Participants interacted with this interface using a standard mouse.

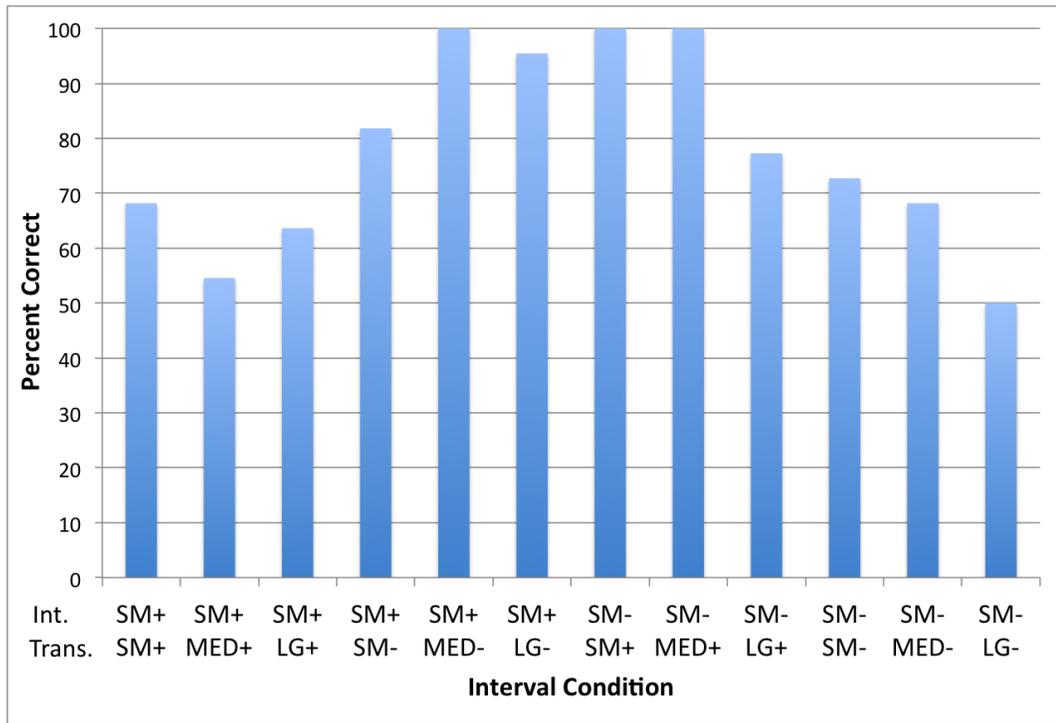
Chapter 3

Results

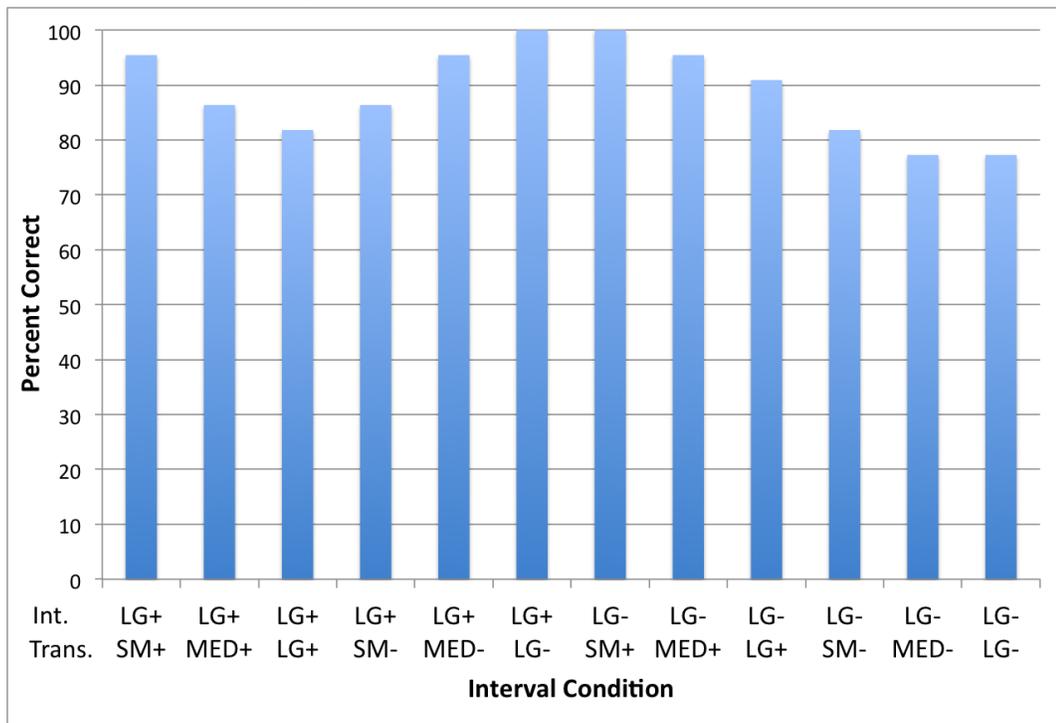
3.1 Directional/contour accuracy

Responses of the participants were analyzed to determine whether or not they had preserved interval direction, i.e., whether the direction of the C:D interval matched that of A:B. We were interested to see the role that contour plays in the performance of the task. Given that the intervals we used consist of both a magnitude and a direction, we wanted to know if participants were at least able to identify the correct direction in their response. Averaged across all conditions, participants responded with the correct direction 83.3% of the time in the attack-time block and 96.4% of the time in the spectral-centroid block.

Three main factors were found to contribute to these results. The first is simply the dimension—participants scored higher on average in the spectral-centroid block. Participants also preserved direction more often in conditions that used large intervals (89.0% in the attack-time block, 99.6% in spectral centroid) compared with small interval conditions (77.7% in attack, 93.2% in centroid). The third factor is whether the condition was congruent (interval direction matches transposition direction) or incongruent (interval and transposition directions are opposite). In congruent conditions, participants scored lower on average (73.1% in attack, 94.3% in centroid) compared to incongruent conditions (93.6% in attack, 98.5% in centroid). These results are shown in Figures 3.1–3.4.

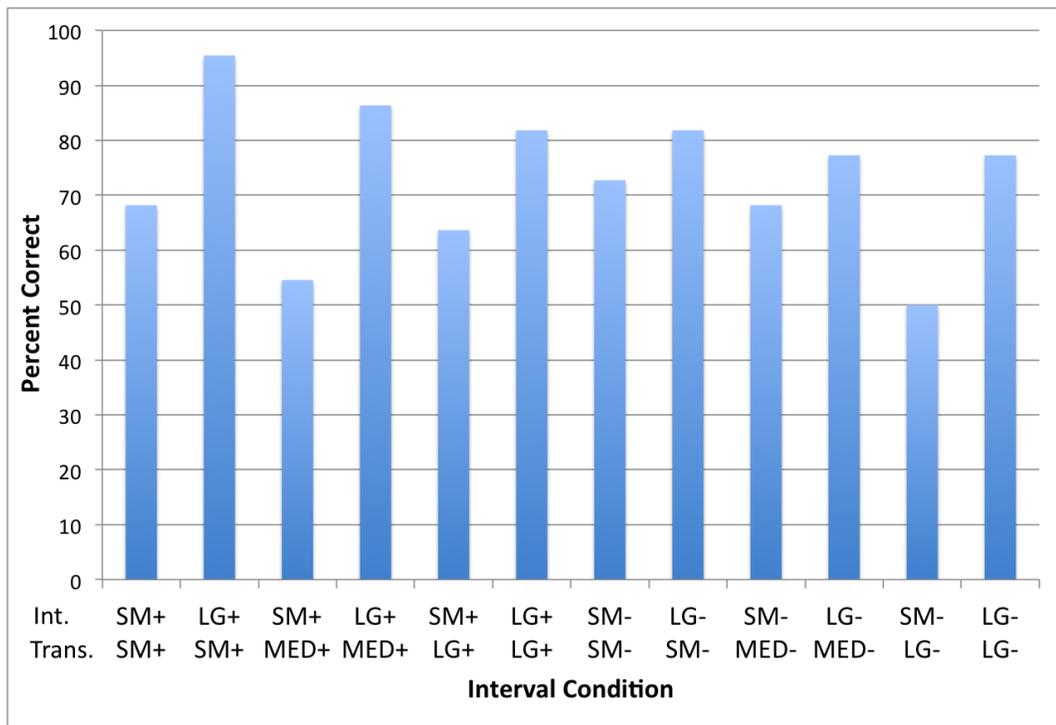


(a) Attack-time conditions with small intervals.

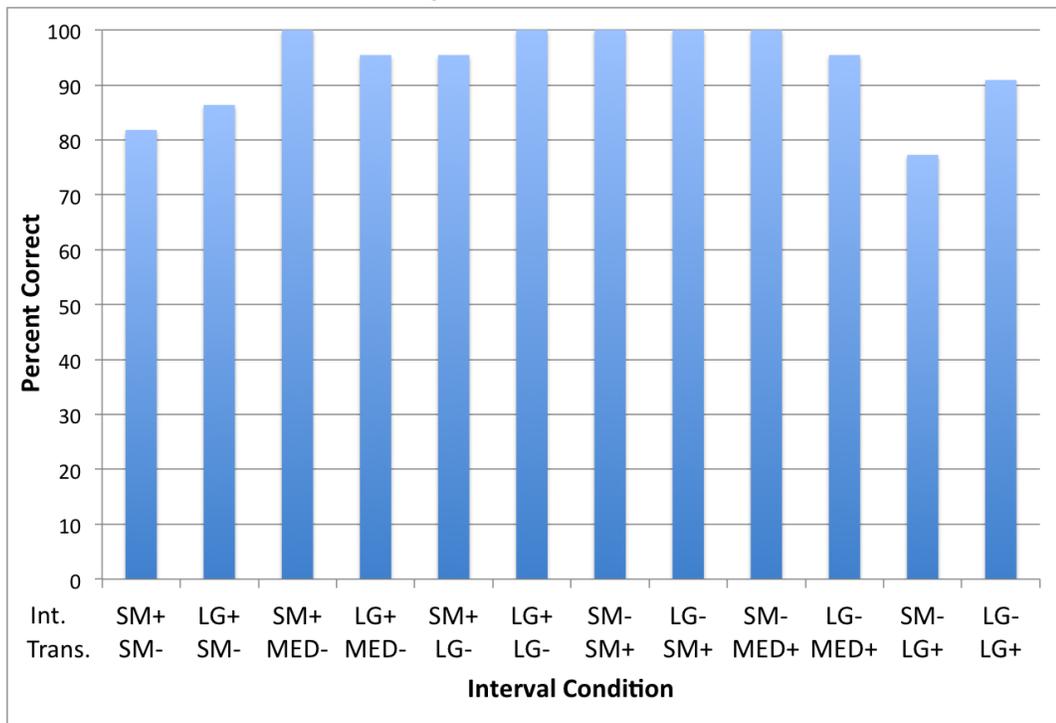


(b) Attack-time conditions with large intervals.

Fig. 3.1 Percentages of responses with correct interval directions, contrasted by interval size. Sizes are denoted by SM, MED, and LG. + signifies an upward direction whereas - is downward. Interval sizes are in the upper row, whereas transposition sizes are in the lower one.

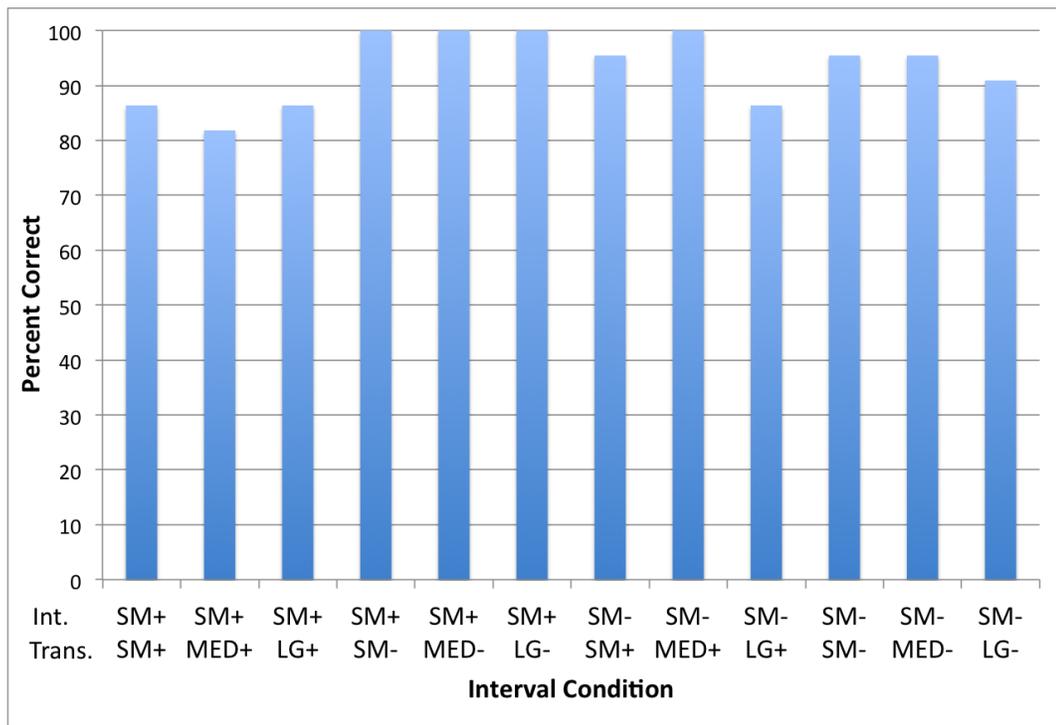


(a) Congruent attack-time conditions.

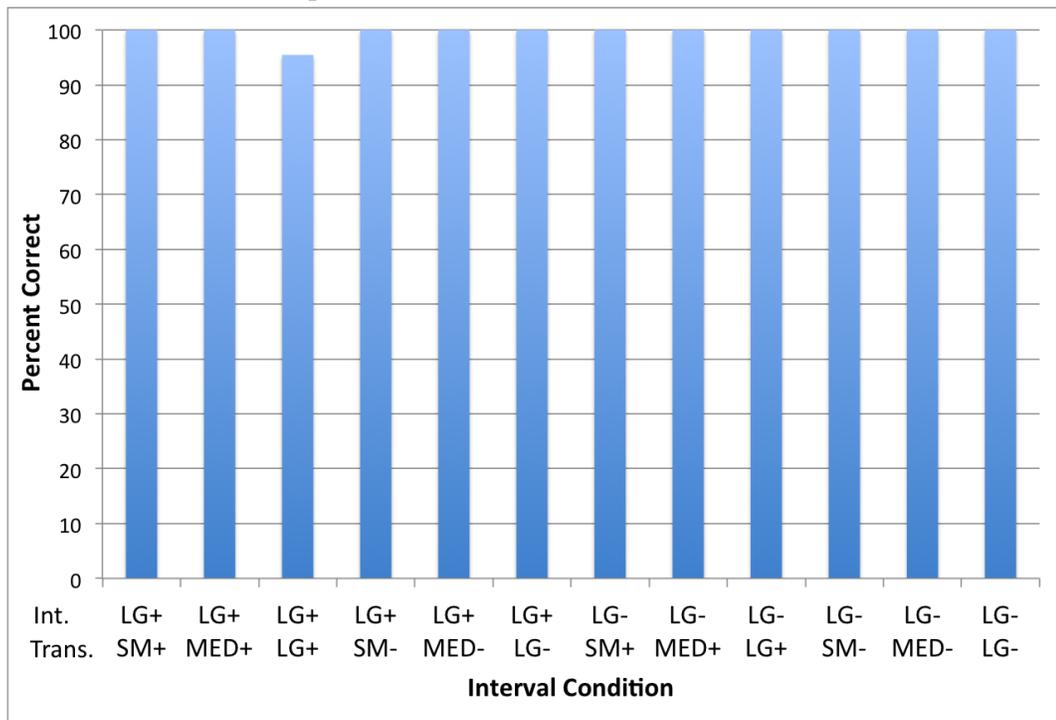


(b) Incongruent attack-time conditions.

Fig. 3.2 Percentages of responses with correct interval directions, contrasted by congruency. Sizes are denoted by SM, MED, and LG. + signifies an upward direction whereas - is downward. Interval sizes are in the upper row, whereas transposition sizes are in the lower one.

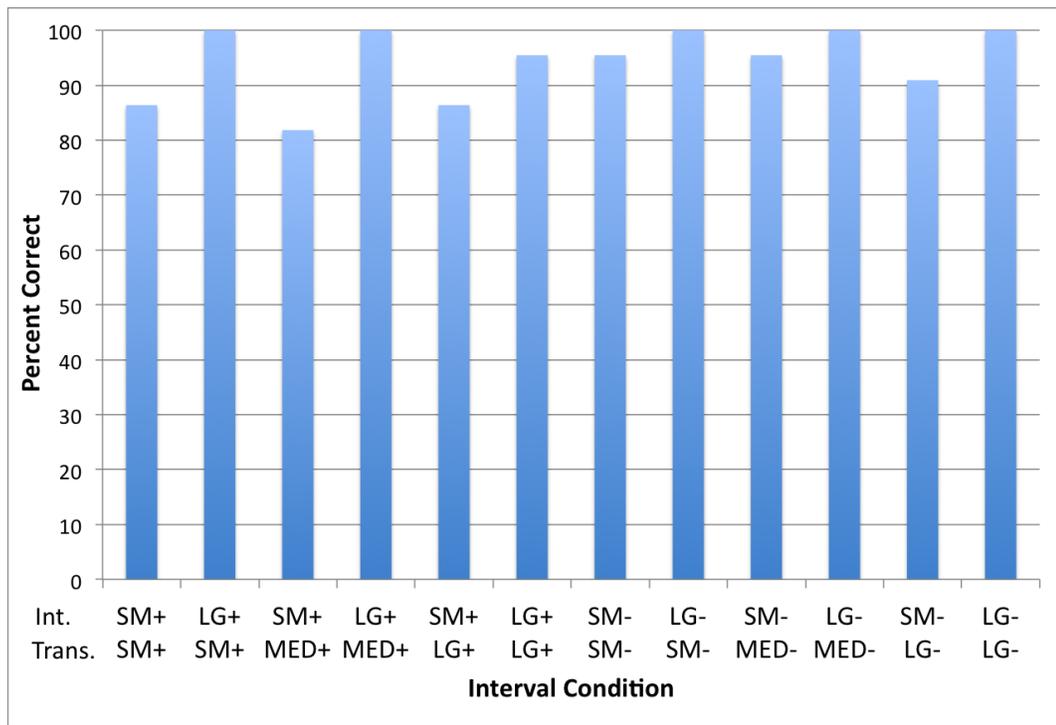


(a) Spectral-centroid conditions with small intervals.

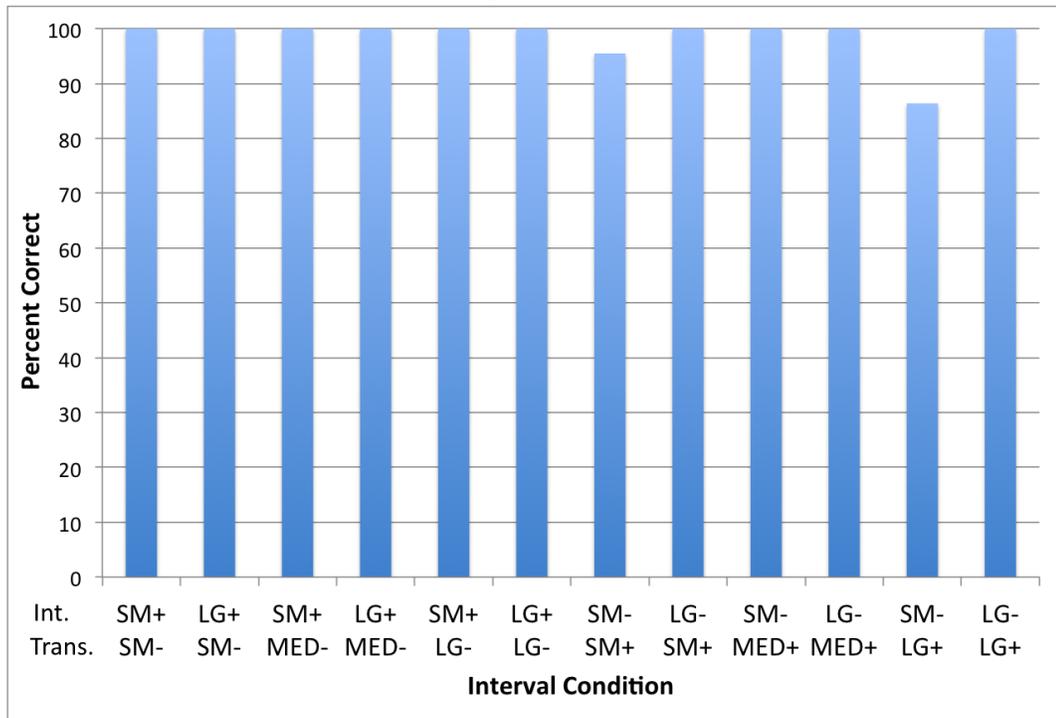


(b) Spectral-centroid conditions with large intervals.

Fig. 3.3 Percentages of responses with correct interval directions, contrasted by interval size. Sizes are denoted by SM, MED, and LG. + signifies an upward direction whereas - is downward. Interval sizes are in the upper row, whereas transposition sizes are in the lower one.



(a) Congruent spectral-centroid conditions.



(b) Incongruent spectral-centroid conditions.

Fig. 3.4 Percentages of responses with correct interval directions, contrasted by congruency. Sizes are denoted by SM, MED, and LG. + signifies an upward direction whereas - is downward. Interval sizes are in the upper row, whereas transposition sizes are in the lower one.

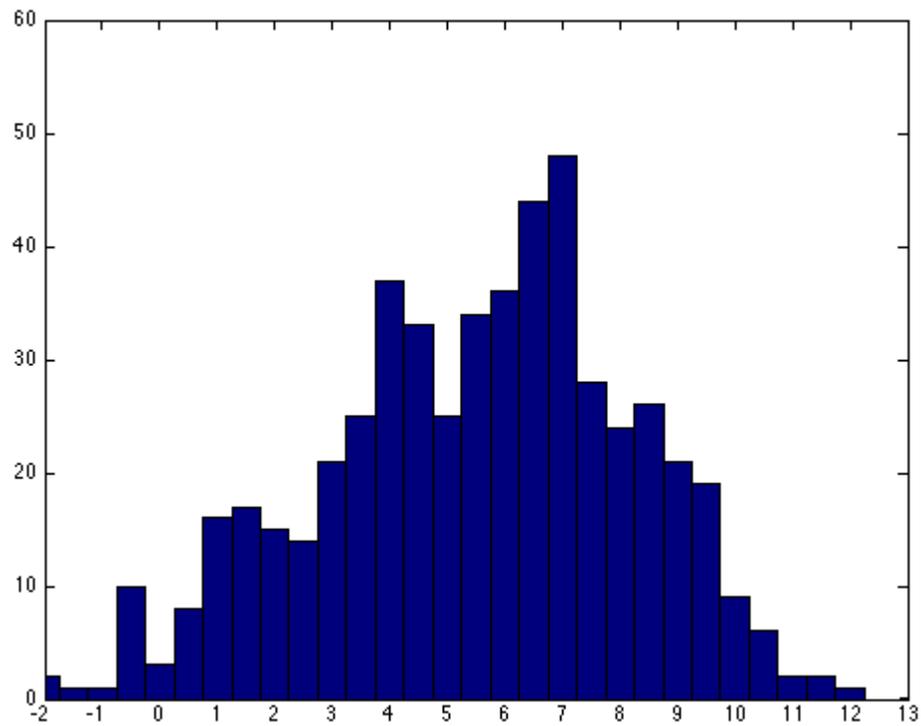
3.2 Response analysis

For both attack-time and spectral-centroid blocks, a histogram was created from all participants' responses across all 24 interval conditions (see Fig. 3.5). This was done to ensure that there was not any strange behavior in the data, i.e., that participants were satisfied with the length of the slider and that their responses were distributed evenly. In both cases, the histogram took on a bell shape, but the spectral-centroid bell was more condensed with fewer responses at or near the endpoints. The bell shape is telling because it matches the frequency with which the **ten** points on the slider were chosen as the expected point D. The **four** points in the middle (4–7) were the expected answer in **three** conditions each, whereas the **three** points on either side of this (1–3, 8–10) were expected twice. We designed the experiment this way in order to balance the expected outcomes as evenly as possible across the **ten** points.

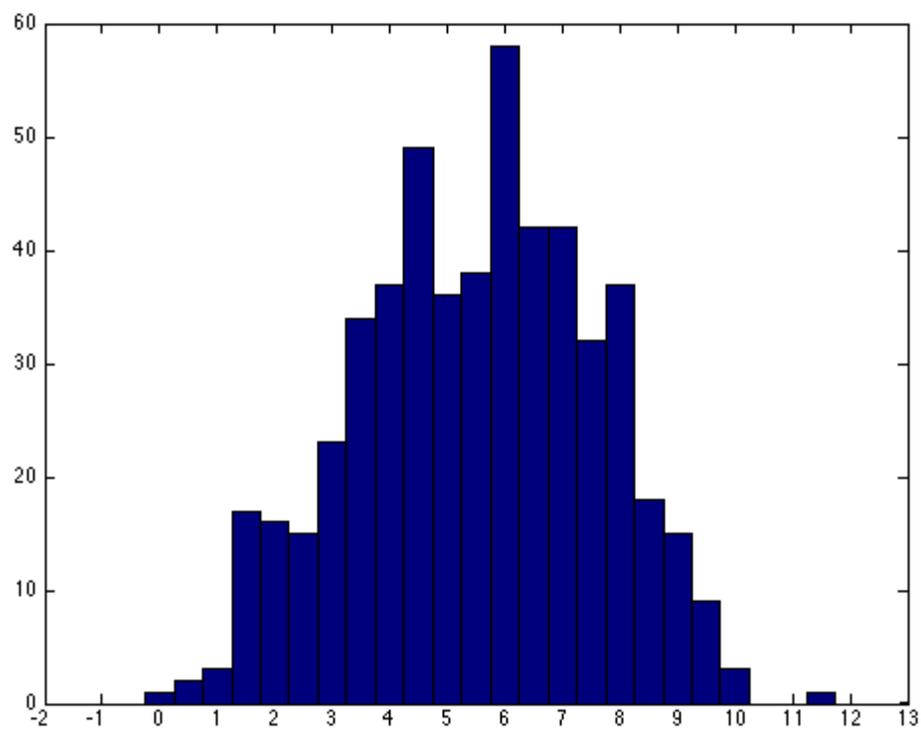
Histograms were also created for each of the 24 individual conditions in order to observe how the participants' responses were spread out around the mean and median (see Appendix). A few major differences between the two blocks stood out. For example, the spectral-centroid histograms were much more condensed overall. Using the 1–10 scale, the average range of responses across all conditions was 3.9 for centroid and 9.3 for attack time.

In addition to this, responses in the centroid block showed a greater degree of consensus between participants. The peak sizes of the histograms were considerably higher in this block—the average peak size was 3.8 participants out of 24 with a median of 4.0 for attack time, whereas the centroid block had an average peak size of 6.5 participants with a median of 6.0. For the sake of clarity, we halved the resolution of the attack-time histograms in the diagrams shown in this paper, because our original bin size of 0.5 produced charts that were too flat to be meaningful. This brought the average and median peak sizes up to 6.0.

We can see the greater degree of consensus in the spectral-centroid block reflected in the overall standard deviation across all conditions as well (1.0 for centroid, 2.4 for attack time). Furthermore, mean and median responses are typically closer to the expected point D in the centroid block. The overall root-mean-squared error around this predicted point was 1.4 for centroid and 2.9 for attack time, again expressed in the 1–10 scale.



(a) Attack time



(b) Spectral centroid

Fig. 3.5 Histograms created using responses across all attack-time conditions and all spectral-centroid conditions. X-axis displays the full range of the slider participants used to mark their response.

3.3 Effects of point placement and congruency

The most prominent, overarching trend in the results correlates with the congruency of the conditions. For congruent conditions, participants tend to undershoot the predicted D point, making the C:D interval too small. In incongruent conditions, by contrast, they tend to overshoot the predicted point and make the C:D interval too large. This trend appears in both the attack-time and spectral-centroid data, regardless of interval or transposition size. Fig. 3.6 shows two examples of this pattern using conditions from the spectral-centroid block. See Appendix for a complete list of histograms for all conditions.

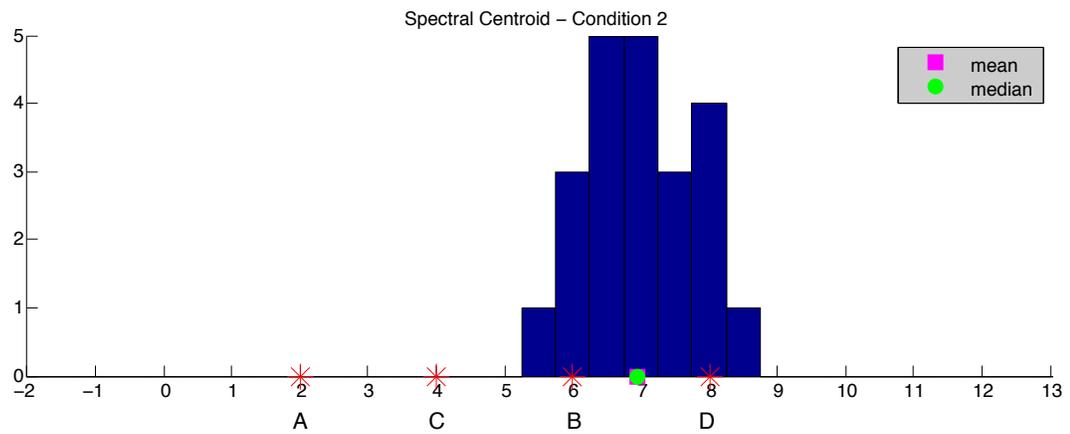
In order to study this over/underestimation effect in more detail, we performed a specialized analysis of the means and standard deviations of these overestimations and underestimations, i.e., the mean distance from the predicted D and the standard deviation from this mean. To make this analysis more meaningful, it was necessary to switch the sign of this distance in conditions with downwards interval directions, so that an underestimation was consistently defined as a negative value, whereas an overestimation was defined as positive. We then performed a paired-samples *t*-test to assess whether or not the differences in these means were significant.

Using this strategy, we found that the mean underestimation in congruent conditions was -1.2 for spectral centroid and -1.9 for attack time, with standard deviations of 0.9 and 2.6, respectively. Incongruent conditions saw a mean overestimation of 0.6 for centroid (standard deviation 1.2) and 0.9 for attack (standard deviation 2.4). In both blocks, the difference between the overestimation and underestimation means was significant ($t(263) = -18.7, p < 0.001$ in spectral centroid, $t(263) = -13.4, p < 0.001$ in attack time).

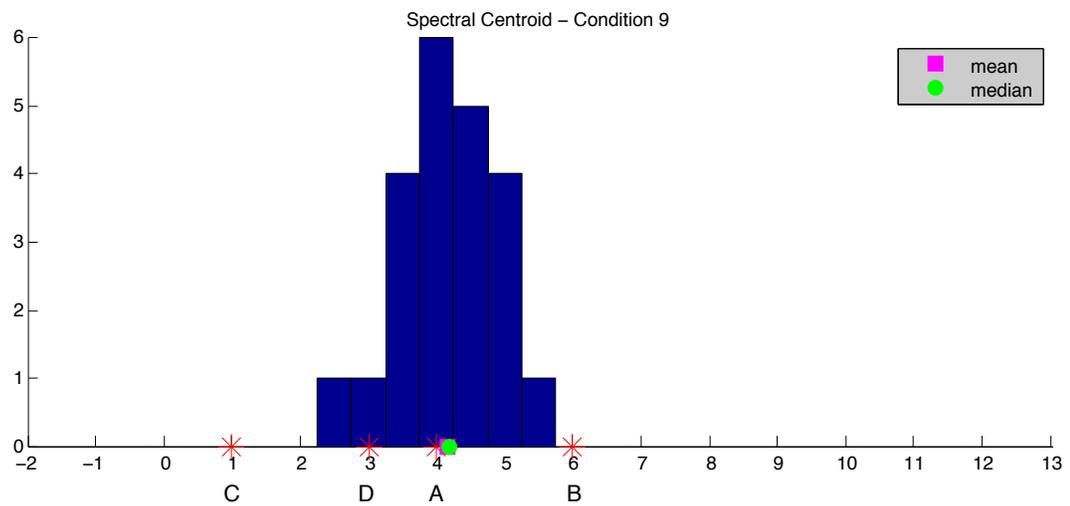
To study the effects of point placement in more detail, the conditions were divided into three categories: concatenated, interleaved, and separated. Concatenated conditions are those in which two points (including the projected value of D) share identical attack times or spectral centroids. In other words, these are conditions in which the intervals A:B and C:D are perfectly next to one another, regardless of which one comes first.

In concatenated congruent conditions (where $B=C$), the mean underestimation was -1.3 for spectral centroid and -2.0 for attack time, with standard deviations of 1.1 and 2.6, respectively. Incongruent conditions showed a mean overestimation of 0.7 for centroid (standard deviation 1.1) and 1.1 for attack (standard deviation 2.4). The differences in these means were significant as well, with $t(87) = -11.8, p < 0.001$ in spectral centroid and $t(87) = -8.0, p < 0.001$ in attack time.

Interleaved conditions were those in which the interval size was larger than the transposition. In other words, point B in these conditions was placed somewhere between C and D in the congruent case, whereas point A fell between C and D in the incongruent case. In congruent interleaved conditions, the mean underestimation was -1.2 for centroid with a standard deviation of 0.9. For attack time, mean underestimation was -1.8 with a standard deviation of 2.3. Incongruent conditions had a mean overestimation of 0.2 for centroid and 0.4 for attack, with respective standard deviations of 1.2 and 2.3. Mean differences were significant in both blocks with $t(87) = -8.9, p < 0.001$ in spectral centroid and $t(87) = -6.5, p < 0.001$ in attack time.



(a) Interleaved congruent condition with a large upward interval and a small upward transposition



(b) Separated incongruent condition with a small upward interval and a medium downward transposition

Fig. 3.6 Examples of congruent underestimation and incongruent overestimation in the spectral-centroid block.

Separated conditions are those in which there is an empty region between the two intervals. Congruent separated conditions had a mean underestimation of -1.1 for spectral centroid and -1.9 for attack time, with standard deviations of 0.8 and 2.8 , respectively. Incongruent conditions in this category showed mean overestimations of 0.8 for centroid (standard deviation 1.2) and 1.2 for attack (standard deviation 2.3). Differences in these means were again significant in both blocks with $t(87) = -12.0, p < 0.001$ in spectral centroid and $t(87) = -8.8, p < 0.001$ in attack time.

For each of these categories, we calculated the root-mean-squared error around our predicted point D to see how close, on average, our participants were to the expected values. We also looked at the standard deviation to get an idea of how much consensus there was across participants. In concatenated congruent conditions, the standard deviation was 0.9 for centroid and 2.6 for attack, with respective root-mean-squared errors of 1.7 and 3.3 . The standard deviations in concatenated incongruent conditions were 1.1 for centroid and 2.4 for attack, with root-mean-squared errors of 1.3 and 2.6 . Interleaved congruent conditions had standard deviations of 0.8 for centroid and 2.3 for attack, with root-mean-squared errors of 1.5 and 2.9 . Standard deviations in interleaved incongruent conditions were 1.2 for centroid and 2.3 for attack, with root-mean-squared errors of 1.2 and 2.3 . Finally, in the separated category, congruent conditions had standard deviations of 0.8 for centroid and 2.8 for attack, with 1.3 and 3.4 root-mean-squared errors. Incongruent conditions in this category had standard deviations of 1.1 for centroid and 2.1 for attack, with root-mean-squared errors of 1.4 and 2.6 . The results of this section are summarized in Table 3.1.

A few notable trends are revealed by these data. Conditions in the attack-time block showed higher magnitude values than their spectral-centroid counterparts in all analysis factors (congruent underestimations, incongruent overestimations, standard deviations around these over and underestimations, standard deviations around the condition mean, and root-mean-squared error). Incongruent overestimations were consistently lower in magnitude than congruent overestimations across all categories and for both blocks. Overestimations also varied more dramatically as a function of point placement, with interleaved conditions having lower overestimations compared to concatenated and separated conditions. Finally, incongruent conditions tended to have lower root-mean-squared errors than their congruent counterparts, with the lowest errors seen in incongruent interleaved conditions and the highest seen in concatenated and separated congruent conditions.

Point Placement	Congruency	Mean O/U	SD of Mean O/U	SD of Cond.	RMS Error
Concatenated	Congruent	-2.0	2.6	2.6	3.3
	Incongruent	1.1	2.4	2.4	2.6
Interleaved	Congruent	-1.8	2.3	2.3	2.9
	Incongruent	0.4	2.3	2.3	2.3
Separated	Congruent	-1.9	2.8	2.8	3.4
	Incongruent	1.2	2.3	2.1	2.6

(a) Attack-time block

Point Placement	Congruency	Mean O/U	SD of Mean O/U	SD of Cond.	RMS Error
Concatenated	Congruent	-1.3	1.1	0.9	1.7
	Incongruent	0.7	1.1	1.1	1.3
Interleaved	Congruent	-1.2	0.9	0.8	1.5
	Incongruent	0.2	1.2	1.2	1.2
Separated	Congruent	-1.1	0.8	0.8	1.3
	Incongruent	0.8	1.2	1.1	1.4

(b) Spectral-centroid block

Table 3.1 Summary of the results in section 3.3. Mean O/U refers to the mean over and underestimations in incongruent and congruent trials, respectively. The standard deviation of these means are displayed, as well as the standard deviation of the condition mean and the root-mean-squared error

3.4 Interactions

A repeated-measures ANOVA was performed using SPSS statistical software in order to study the interactions between the parameters of interval size, interval direction, transposition size, and transposition direction.

In the attack-time block, the only significant 2-way interactions were between transposition direction and transposition size ($F(2,42) = 5.38, p = 0.008, \eta_p^2 = 0.20$), interval direction and interval size ($F(1,21) = 8.02, p = .010, \eta_p^2 = 0.28$), and transposition size and interval size ($F(2,42) = 4.48, p = 0.017, \eta_p^2 = 0.18$). Significant 3-way interactions were found between interval direction, transposition direction, and transposition size ($F(2,42) = 6.33, p = 0.004, \eta_p^2 = 0.23$), interval direction, transposition size, and interval size ($F(2,42) = 3.95, p = 0.027, \eta_p^2 = 0.16$), and transposition direction, transposition size, and interval size ($F(2,42) = 4.81, p = 0.013, \eta_p^2 = 0.19$). The 4-way interaction of these parameters was not significant. These effects will be analyzed in more detail and interpreted in terms of the experimental task in the Discussion section.

In the spectral-centroid block, the significant 2-way interactions were those between interval direction and transposition size ($F(2,42) = 3.34, p = 0.045, \eta_p^2 = 0.147$), transposition direction and transposition size ($F(2,42) = 25.48, p < 0.001, \eta_p^2 = 0.55$), and interval direction and interval size ($F(1,21) = 13.74, p = .001, \eta_p^2 = 0.40$). Significant 3-way interactions were found between interval direction, transposition direction, and transposition size ($F(2,42) = 4.78, p = 0.014, \eta_p^2 = 0.19$), interval direction, transposition size, and interval size ($F(2,42) = 4.79, p = 0.013, \eta_p^2 = 0.19$), and transposition direction, transposition size, and interval size ($F(2,42) = 14.66, p < 0.001, \eta_p^2 = 0.41$). As with attack time, the 4-way interaction was not significant in this block.

Chapter 4

Discussion

4.1 Directional/contour accuracy

Interestingly, participants maintained interval direction with a surprisingly high success rate, considering their overall performance on the task. This indicates that they had a much easier time identifying directions of changes in attack time or spectral centroid rather than their exact magnitude. It would seem that contour or direction is a much more firmly ingrained listening schema than precise timbre intervals, especially in the case of attack time.

The other contributing factors, interval size and congruency, shed some light on this as well. It seems fitting that directional accuracy would increase with larger interval sizes, since a larger change in timbre makes the direction of the change more apparent. Incongruent conditions might also be effective at conveying directional information indirectly by presenting higher relative amounts of change. In these conditions, the direction of the A:B interval will always be the opposite of the subsequent change from B to C. If the A:B interval is upward, for example, then by the end of the initial condition presentation the participant will have heard an upward change (A:B) as well as a downward one (B followed by C). Moreover, this change from B to C is always at least the size of a large interval, because even in concatenated small interval conditions—where A, B, and C are as close together as possible—the distance from B to C is **four** steps on the scale, which equates to a large interval. Perhaps this motion, when contrasted with the opposite-direction A:B interval immediately prior, makes participants more likely to ascertain the direction of the A:B interval and preserve it in their response.

It is worth noting that because the percentages were so high overall for the spectral-centroid block, we do not see the scores change as drastically across other factors. For example, in the attack-time block, going from small interval conditions to large interval ones raises the average success rate by 11.3% (77.7% to 89.0%), whereas in the spectral-centroid block we only see an increase of 6.4% (93.2% to 99.6%), which represents a ceiling effect. This is most likely due to the fact that the task in the spectral-centroid block was simply much easier, as reported by numerous participants. Because the baseline for correctly predicting the direction in this block is already so high (average of 96.4% across all conditions), interval sizes or congruency do not have much of an opportunity to further influence the scores.

4.2 Response analysis

The histograms of all participants' responses across all conditions reveal informative results. The fact that the points are mostly clustered in the region where all A, B, C, and predicted D points lie tells us that the participants found the slider to cover an adequate range of timbres. Only rarely did they feel the need to respond with points at or near the endpoints of the slider. This is especially true in the spectral-centroid block, where the histogram is more condensed. It also indicates that participants had an easier time with this block, because their responses are more collapsed in the region in which we expected them to lie. This observation is consistent with comments made by several of the participants after the experiment as well.

This trend carries over into the histograms of individual conditions. On average, conditions in the attack-time block had a much larger range of responses than their spectral-centroid counterparts, as well as a higher standard deviation, higher root-mean-squared error around D, and a lower average peak size. This indicates that not only did subjects agree with one another more often in their responses to the spectral-centroid conditions, but they were also closer to our predicted D points. From this, we can deduce that our expected points on the logarithmic scale hold more predictive power in the spectral-centroid block than in that of attack time. However, it is important to note that the attack-time block shows roughly equivalent increases in both root-mean-squared error and standard deviation (1.4 and 1.0 for centroid up to 2.9 and 2.4 for attack, respectively). Whereas it is true that the logarithmic scale used in the attack-time block has less predictive power than that of spectral centroid, we can also see from the standard deviation that attack-time intervals are, in a word, unpredictable. High root-mean-squared errors in this block are not necessarily indicative of any inherent failures of the logarithmic scale used, but may have more to do with perceptual resolution along this dimension being poorer.

By contrast, it appears that the perceptual resolution of spectral centroid is larger than that of attack time. In other words, although participants in the attack-time block would seem to have relied primarily on contour to complete the task, they evidently had at least some notion of precise position along the spectral centroid dimension. The biggest piece of evidence for this claim is the significant decrease in standard deviation in this block, although the lower range of responses and higher average peak sizes lend some credit as well. One explanation for this phenomenon is that unlike attack time, spectral centroid perception relies on the same underlying sensory continuum used to encode frequencies and detect pitch height (McAdams & Bregman, 1979), and the ability of musicians to recognize transposed pitch intervals is well established (Attneave & Olson 1971; Bartlett & Dowling 1980; Cuddy & Cohen 1976).

4.3 Effects of point placement and congruency

The most interesting and most baffling finding in our data was the tendency for participants to underestimate our predicted point in congruent conditions and to overestimate it in incongruent conditions. To study this in further detail, we divided the conditions into three

categories—concatenated, interleaved, and separated—based on the positions of the A, B, C, and predicted D points. The most plausible explanation for the undershoot/overshoot effect is likely a result of participants becoming biased to answer closer to point B. In congruent conditions, biasing towards B will result in an underestimation. Because transposition direction matches interval direction, points B and C will always be on the same side relative to point D. Response biasing towards B makes the C:D interval smaller as well, resulting in underestimation. By contrast, B biasing in incongruent conditions results in an overestimation, since B and C are on opposite sides of the point D.

We see this reflected in a number of trends in the data. For example, separated incongruent conditions were the ones with the highest mean overestimation in both attack-time and spectral-centroid blocks (0.8 for centroid, 1.2 for attack time). These are also the conditions in which the B point is furthest away from the expected D value. Similarly, interleaved incongruent conditions showed the lowest mean overestimation in both blocks (0.2 for centroid, 0.4 for attack time), and these are the conditions for which points B and predicted D were closest to one another.

A similar trend can be found with congruent conditions in the attack-time block, though to a lesser extent. The conditions with the greatest average underestimation (in terms of magnitude) were concatenated and separated congruent conditions (−2.0 and −1.9, respectively). Naturally, this means that interleaved congruent conditions had the lowest average underestimation (−1.8). However, we see a small deviation from the trend here, as concatenated conditions have a larger mean underestimation than separated ones. The spectral-centroid block deviates even further from this trend, as concatenated congruent conditions showed the largest underestimation (−1.3), but separated congruent conditions showed the smallest (−1.1). It appears that the B biasing effect is not as strong in this case, which seems appropriate given that participants have a stronger notion of precise position in this block.

Interestingly, point placement had less of an effect in congruent conditions than in incongruent conditions, where overestimation spanned from 0.2 to 0.8 for spectral centroid and 0.4 to 1.2 for attack time. One possible explanation for this is that underestimation in congruent conditions is somewhat bounded by the tendency to preserve interval direction. Whereas there is no limit to how much a participant can overestimate in an incongruent condition while still maintaining the contour of the A:B interval, in congruent conditions a sufficiently large underestimation will obfuscate the direction of the C:D interval and weaken the perceptual analogy. In essence, this leads to a ceiling effect where increasing the distance from B to D does not produce as large of changes in underestimation as one might expect.

Additional evidence of the effects of B-to-D distance can be seen in the root-mean-squared errors of the conditions. In the attack-time block, separated congruent conditions had the highest root-mean-squared error (3.4) whereas interleaved incongruent conditions had the lowest (2.3). This is most likely due to large underestimations in separated congruent conditions and low overestimations in interleaved incongruent conditions, which again correlate with the distance from B to D. Interleaved incongruent conditions also had the lowest root-means-square error in

the centroid block (1.2). However, it was the concatenated congruent conditions—rather than the separated ones—that had the highest root-mean-squared error (1.7). Again, this is likely due to the fact that the B biasing phenomenon is not as influential a factor in the spectral-centroid task.

It is worth noting that overestimation in incongruent conditions is consistently smaller than underestimation in congruent conditions. It appears that if the B biasing effect were not present, the natural tendency would be for participants to slightly underestimate the length of the presented interval, regardless of whether the condition were congruent or incongruent. This is most likely due to biases associated with the method of adjustment (Gescheider 1976). The tendency in this method is for participants to underestimate their goal—when asked to match a given stimulus to a reference stimulus, for example, answers tend to fall above or below the reference point if the given stimulus is also above or below the point, respectively.

In the case of congruent conditions, this underestimation is exacerbated by the fact that points B and C are placed on the same side relative to the predicted D point. In incongruent conditions, with B and C on opposite sides of D, the B biasing effect counteracts the underestimation associated with adjustment, producing an overestimation. However, the resulting overestimation has a lower magnitude than the underestimations seen in congruent conditions.

The B biasing phenomenon seen in this study is likely an example of psychological anchoring, an unconscious cognitive bias toward previously presented information. The effects of anchoring are well established. In a study by Tversky & Kahneman (1974), participants were asked to calculate either $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ or $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ in only 5 seconds, forcing them to take a guess. Participants presented with the descending series guessed higher than their ascending series counterparts. Moreover, effects of anchoring can be seen even when the presented stimulus has no apparent relation to the question at hand. In another study, they presented participants with a roulette wheel that was predetermined (unbeknownst to them) to land on a specific number. They were then asked whether the number of African countries in the UN was greater or fewer than the number on the roulette wheel, and by how much. The median guess was 25 countries for participants who saw the wheel land on 10 (i.e., a guess of 15 countries over 10), whereas those whose wheel landed on 65 gave a median estimate of 45 (20 fewer than 65).

In the present study, the data suggest that participants are anchoring onto point B in their responses. Because the task was explained as an analogy—A is to B as C is to D—it seems natural to unconsciously associate B and D, despite being instructed to make A:B and C:D perceptually equidistant. As a result, responses were closer to B than expected, creating the overestimation and underestimation trend that we see in the data.

4.4 Interactions

Significant 2-way interactions in the attack-time block include Transposition Direction \times Transposition Size, Interval Direction \times Interval Size, and Transposition Size \times Interval Size. The estimated marginal means depicted in these interactions represent the mean distance

from the expected point D, averaged across all conditions that share the attributes specified by the graph dimensions. For example, in the Transposition Direction \times Transposition Size interaction (Fig. 4.1), the point at small transposition size on the upward transposition line represents the mean of all conditions with small, upward transpositions, regardless of interval size or direction.

The important fact to keep in mind when analyzing interaction data is that unlike our previous analysis, the estimated marginal means in our ANOVA do not distinguish between congruent underestimation and incongruent overestimation. They simply represent the mean distance above (positive) or below (negative) the expected point D. In other words, a congruent underestimation could be a negative or a positive value, depending on the transposition direction. This leads to a certain degree of “cancellation” in certain interactions. For example, the estimated marginal mean across conditions with upward interval directions pits congruent underestimates against incongruent overestimates, because both upward and downward transposition directions are included in this calculation.

In the Transposition Direction \times Transposition Size interaction (Fig. 4.1), we see positive estimated marginal means for conditions with downward transposition direction and negative means for conditions with upward transposition direction. This is expected, as upward transposition conditions contain congruent underestimations and incongruent overestimations, which are both negatively valued in this context. A similar argument applies to the downward transposition conditions. Moreover, the magnitude of these means increases slightly as transposition size increases, which accounts for the significance of the interaction. The reason behind this interaction is that as transposition size increases, so does the distance from point B to the expected point D, effectively augmenting the B biasing phenomenon responsible for over and underestimations. Interestingly, Transposition Direction \times Transposition Size is a significant interaction in the spectral-centroid block as well. The graph of this interaction bears a close resemblance to that of the attack-time block, indicating that similar mechanisms are likely at play (see Fig. 4.1).

Interval Direction \times Interval Size is the next significant 2-way interaction in the attack-time block. In Fig. 4.2, we see negative estimated marginal means in conditions with upward interval direction. As mentioned before, this calculation weighs congruent underestimations against incongruent overestimations, so the fact that it is negative demonstrates again how underestimations are greater on average than overestimations. Similarly, downward interval conditions have positive means. The interaction in this case results from the fact that conditions with a large interval size show a significantly greater disparity between the two positive and negative means. This is likely due to the fact that small intervals offer less room for underestimation. Given that contour likely plays an important role in this task, and that participants maintained interval direction a large majority of the time in their responses, it naturally follows that underestimations will be less severe. As a result, congruent underestimations are more equally counterbalanced by incongruent overestimations, yielding a mean closer to 0. Once again, the Interval Direction \times Interval Size diagram in the spectral-centroid block mirrors the trends of the attack-time interaction to a surprising degree,

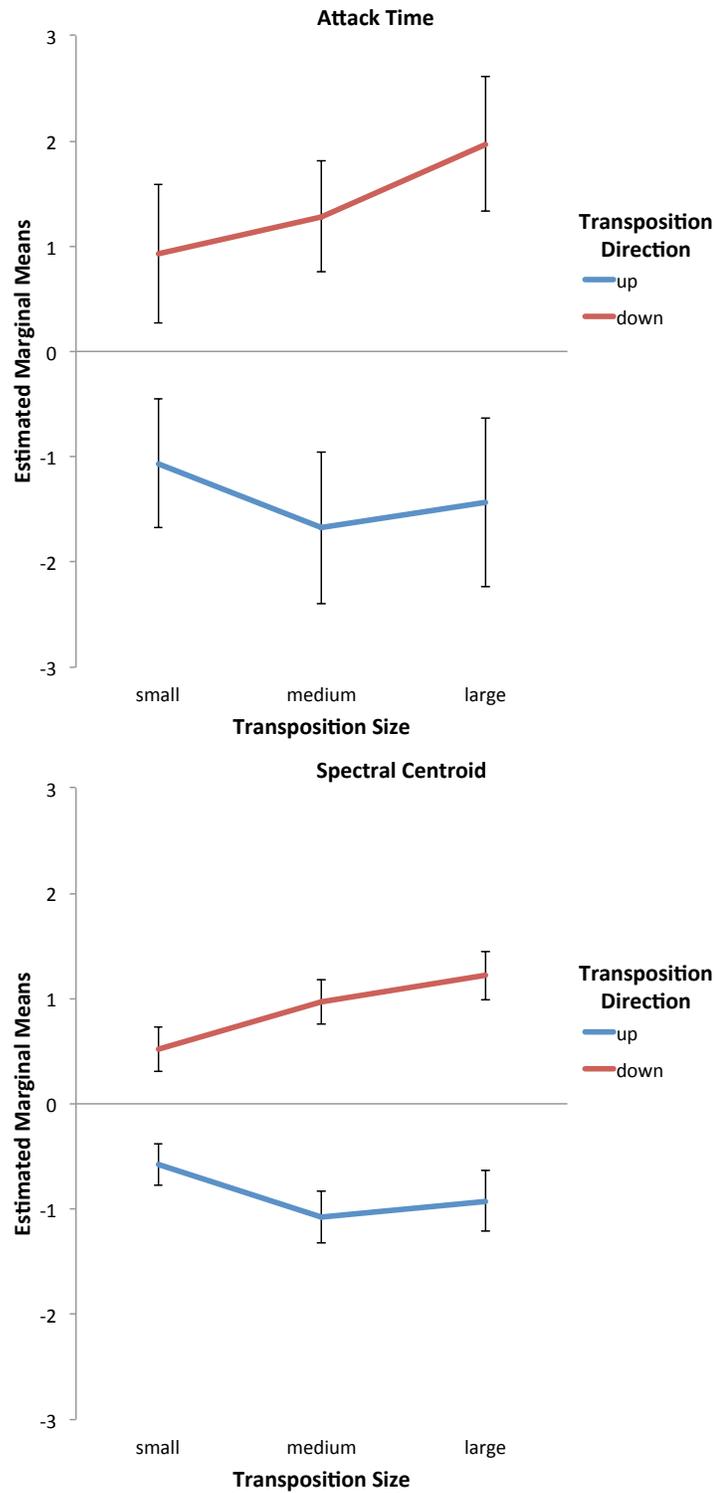


Fig. 4.1 Comparison of the Transposition Direction \times Transposition Size interaction for both attack time and spectral centroid.

indicating that similar arguments would explain the behavior of this graph as well (see Fig 4.2).

The final significant 2-way interaction in the attack-time block is Transposition Size \times Interval Size (Fig. 4.3). However, since the parameters in this interaction tell us nothing about the direction of the intervals or their transpositions, it follows that the estimated marginal mean calculations in this interaction weigh congruent underestimations and incongruent overestimations against their direct counterparts with opposite transposition direction. As a result, most of the differences in means in this graph are insignificant, with the sole exception being that small interval conditions have a significantly larger, positive mean only when the transposition size is large. Essentially what this means is that the positive values from congruent underestimation and incongruent overestimation in downward transposition conditions are great enough to outweigh their counterparts in upward transposition conditions, though only when the interval size is small and the transposition large. For small interval conditions with small or medium transposition size, the opposite is true. It is unclear exactly why this occurs, however, because this is the least significant interaction in the attack-time block and it is insignificant in the centroid block; it appears that B biasing and contour preservation are insufficient to explain this particular trend. It could be the case that this interaction arose purely from variability in the attack-time block, as this block has been shown to have significantly greater variance than that of spectral centroid.

One interaction that was significant in the centroid block (though not that of attack) is Interval Direction \times Transposition Size (shown in Fig. 4.4). On this chart, we again see negative estimated marginal means for conditions with upward interval directions, and positive means for downward interval conditions, which demonstrates how congruent underestimations outweigh incongruent overestimations. The interaction in this graph comes from the fact that downward interval conditions with large transposition sizes have a significantly higher mean than those with medium or small transposition sizes. One could argue that this is again due the B biasing effect, which increases as transposition size (and hence the distance from B to D) increases. However, this raises the question of why a similar trend is not seen in upward interval conditions with large transposition sizes. One possible explanation is that descending spectral-centroid intervals are perceptually larger than ascending ones, much like how descending melodic pitch intervals are perceptually larger than their ascending counterparts (Russo & Thompson 2005b). If this were the case, we would expect participants to underestimate to a greater degree in congruent downward interval conditions in order to compensate for perceptual largeness. As a result, the estimated marginal mean is raised. However, conditions with small or medium interval sizes are less affected because there is less room to underestimate while still maintaining correct contour. This could also explain why this interaction is not significant in the attack-time block, because attack does not share the same underlying sensory representation with pitch that spectral centroid does.

Significant 3-way interactions were the same in both attack-time and centroid blocks—Interval Direction \times Transposition Size \times Interval Size, Interval Direction \times Transposition Direction \times Transposition Size, and Transposition Direction \times Transposition Size \times Interval Size.

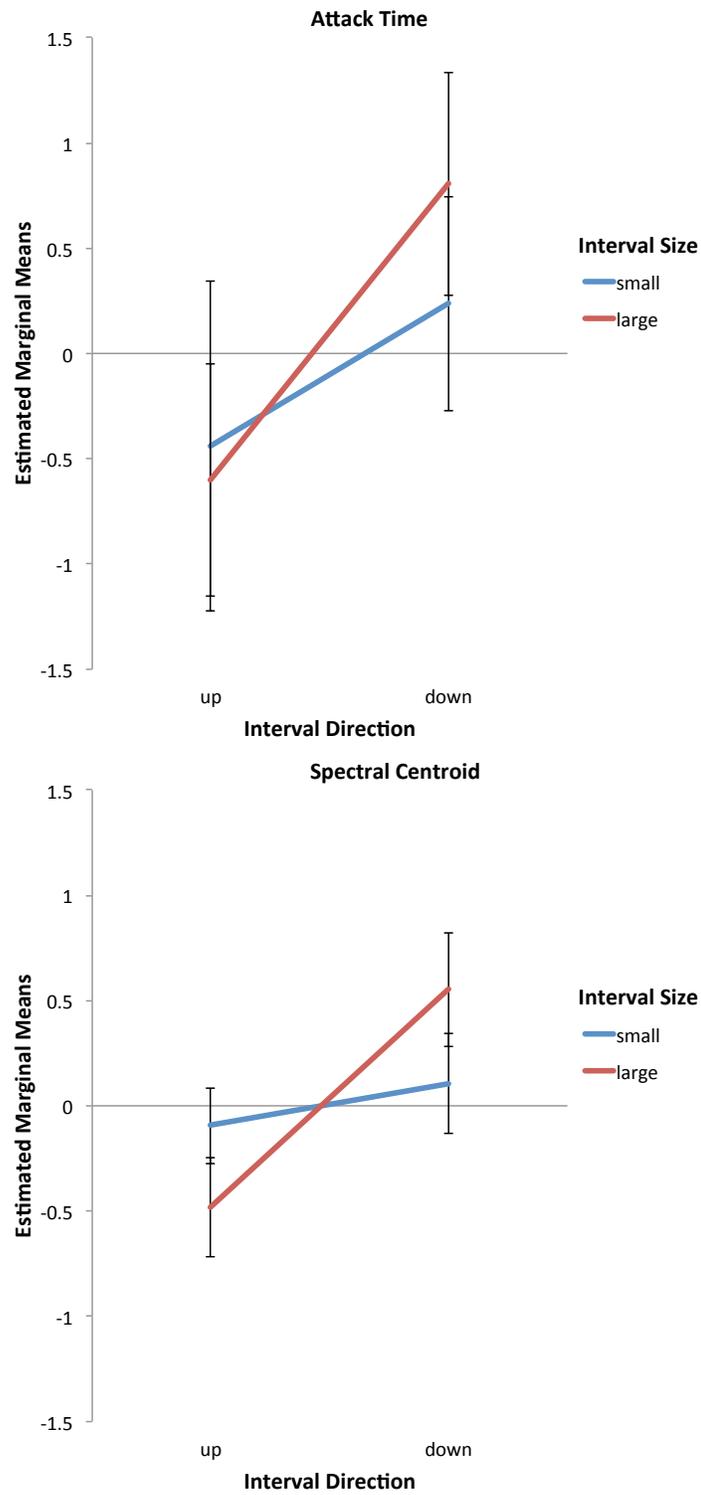


Fig. 4.2 Comparison of the Interval Direction \times Interval Size interaction for both attack time and spectral centroid.

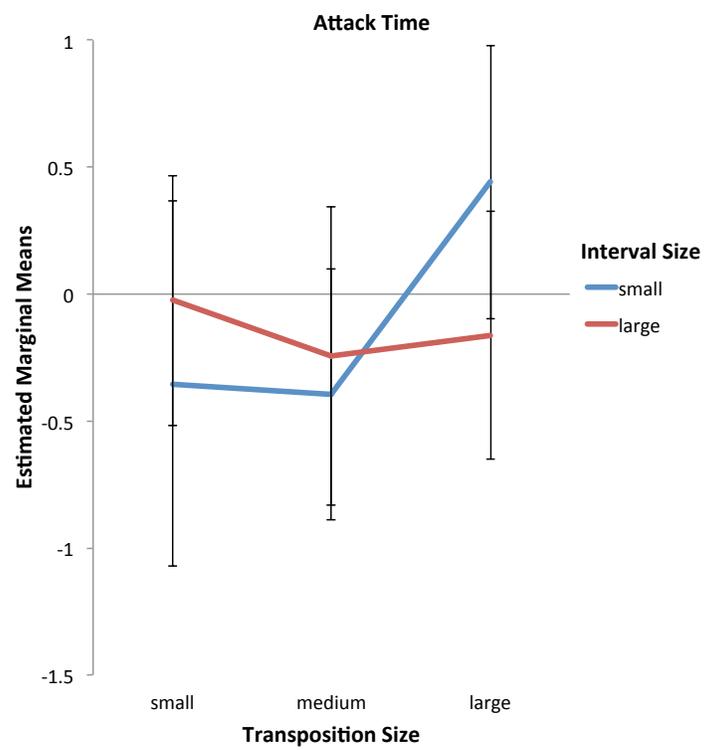


Fig. 4.3 Transposition Size \times Interval Size interaction in the attack-time block.

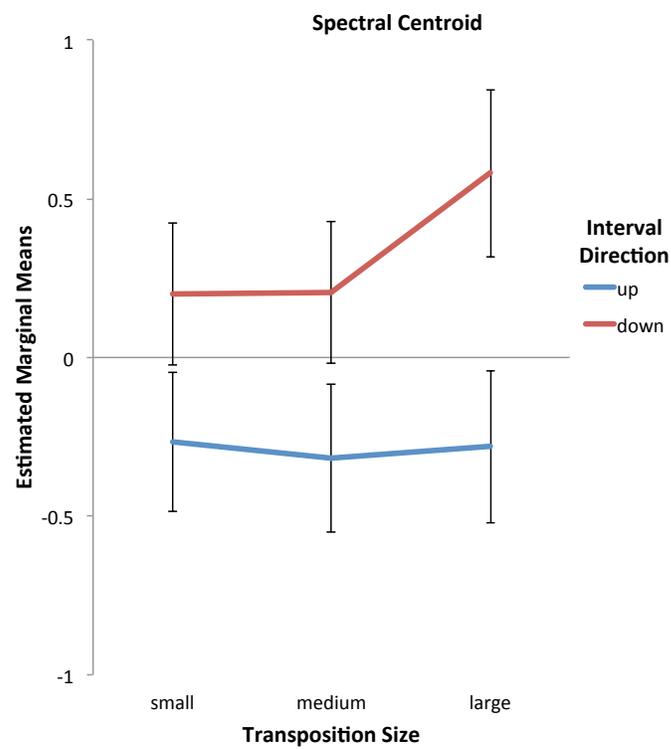


Fig. 4.4 Interval Direction \times Transposition Size interaction in the spectral-centroid block.

It is important to note that, because these interactions contain the transposition size parameter, which has **three** possible values (small, medium, large), each point on the graphs shown (Figs 4.5, 4.6, 4.11, 4.12, 4.14, and 4.15) represents the estimated marginal means of only two interval conditions from 22 subjects for a total of 44 data points.

This is especially relevant in the attack-time interaction of Interval Direction \times Transposition Size \times Interval Size, where outlying conditions disrupt the trends we would normally see (Fig 4.5). Here, we see negative means for upward interval conditions, as expected, although the effects of interval size and transposition size are largely insignificant. Downward interval condition means are mostly positive, but there is a strange interaction in the small interval case. Small downward interval conditions have negative means for small and medium transposition sizes, which is atypical because we normally see that congruent underestimations outweigh incongruent overestimations. Interestingly, a very similar interaction occurs in the centroid block as well (Fig 4.6).

To examine why this occurred, we can look specifically at the conditions that get factored into these calculations. In the case of the small transpositions, the component conditions are 13 (incongruent) and 19 (congruent), shown in Fig. 4.7. The negative values are due to an unusually small underestimation in condition 19 (mean of 1.3 and median 0.9, whereas the average underestimation in concatenated congruent attack-time conditions is 2.0). This occurs in the centroid block as well, as shown in Fig. 4.8 (mean and median of 0.5, with the average underestimation in concatenated congruent centroid conditions being 1.3). Because these congruent underestimations are smaller in magnitude, they do not outweigh the incongruent overestimations of condition 13, which results in negative estimated marginal means.

As for the downward small interval conditions with medium transpositions, this includes conditions 15 (incongruent) and 21 (congruent). In the attack-time block of condition 21 (Fig 4.9), we again see a much lower than average underestimation (0.7 mean, 0.8 median compared to 1.9 average across separated congruent attack conditions). Condition 15 has a larger than average overestimation as well, though to a lesser extent (-1.8 mean, -1.5 median, -1.2 average across separated incongruent attack conditions). Recalling that incongruent overestimation is treated as a negative value in this context, it is easy to see why this point in particular is so low on the graph.

In the centroid block of these conditions, incongruent overestimation still outweighs congruent underestimation, though not as drastically (Fig 4.10). In condition 15, the average overestimation is -1.2, with a median of -1.3, compared to a -0.8 average across separated incongruent centroid conditions. In condition 21, we see an average underestimation of 0.9 with median 0.7, compared to an average of 1.1 across all separated congruent conditions.

It is difficult to ascertain exactly why these conditions in particular deviate from the patterns found. It makes sense that small interval conditions show a less pronounced congruent underestimation because the tendency to maintain interval direction remains high. However, in some cases this difference is quite pronounced (e.g., condition 21 in the attack-time block). In addition, many of these conditions show a higher than average incongruent overestimation.

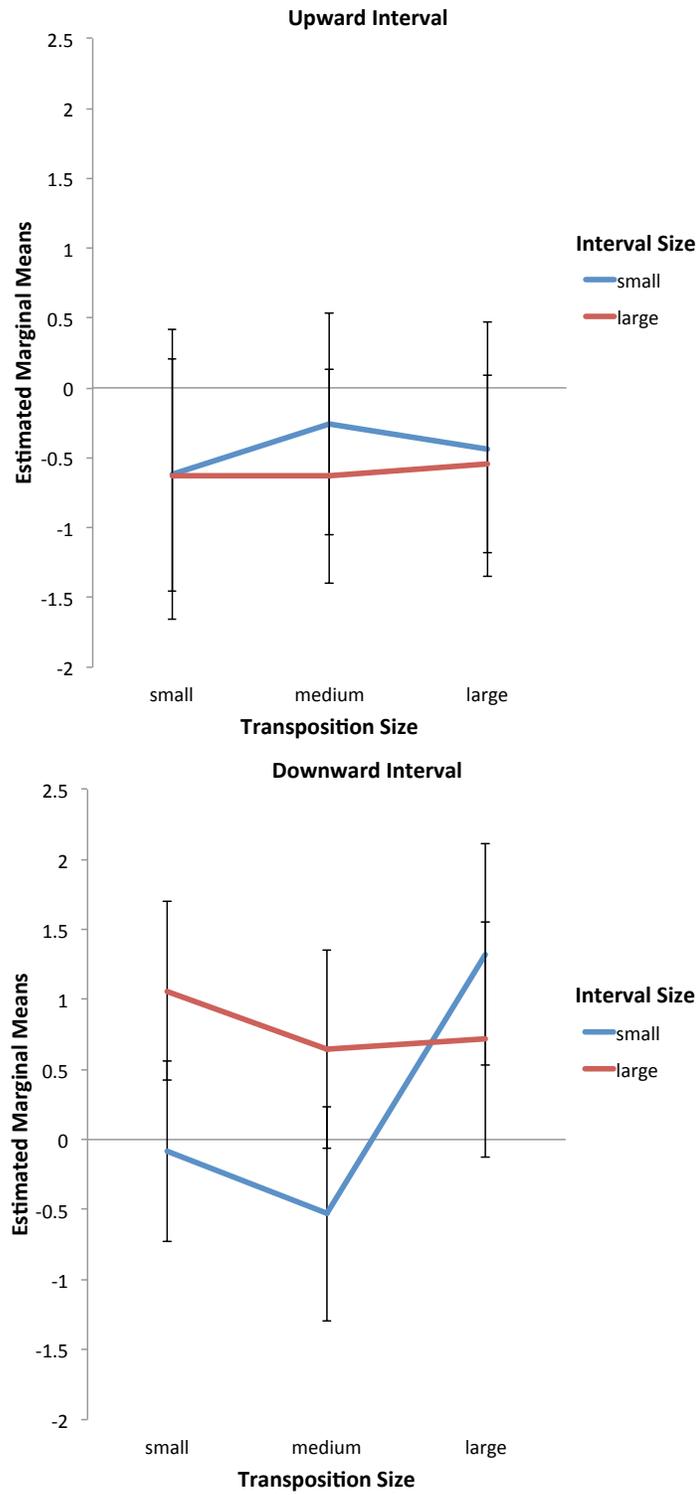


Fig. 4.5 Interval Direction×Transposition Size×Interval Size in the attack-time block.

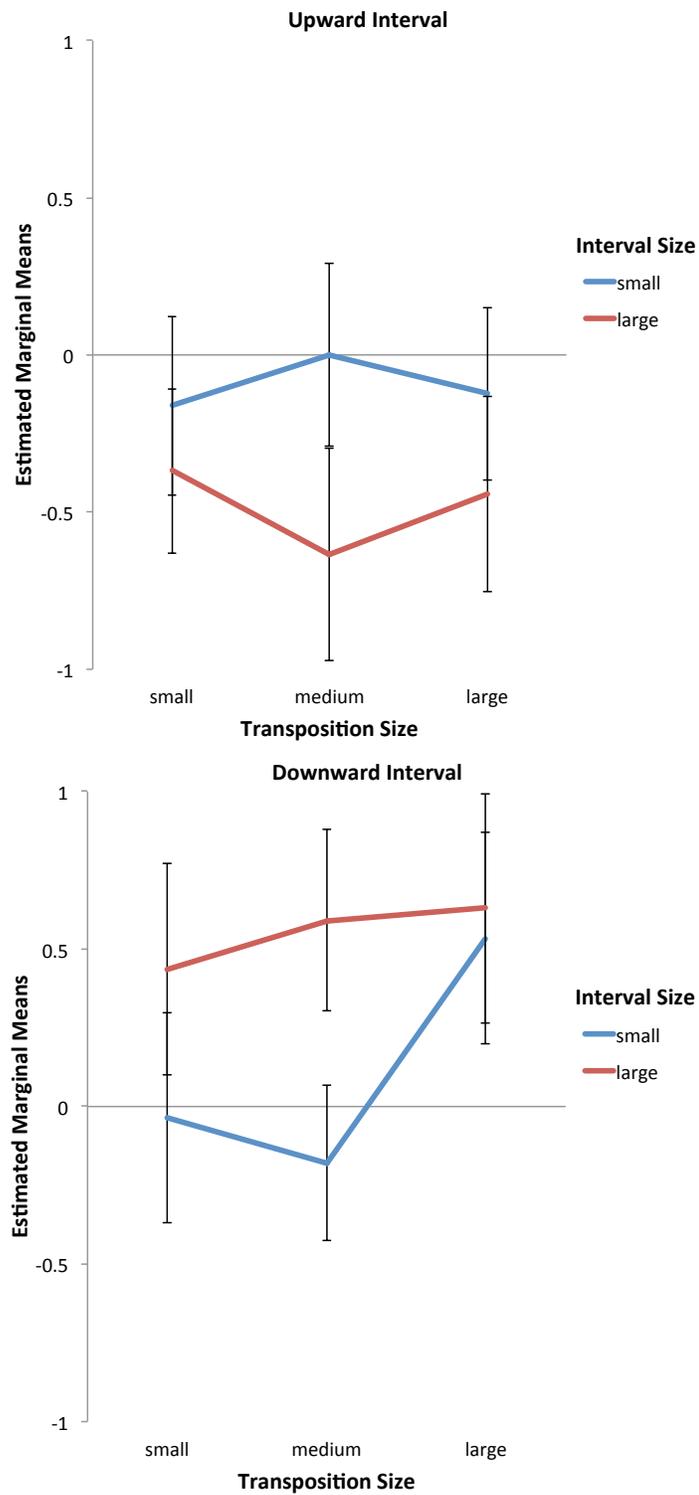


Fig. 4.6 Interval Direction \times Transposition Size \times Interval Size in the spectral-centroid block.

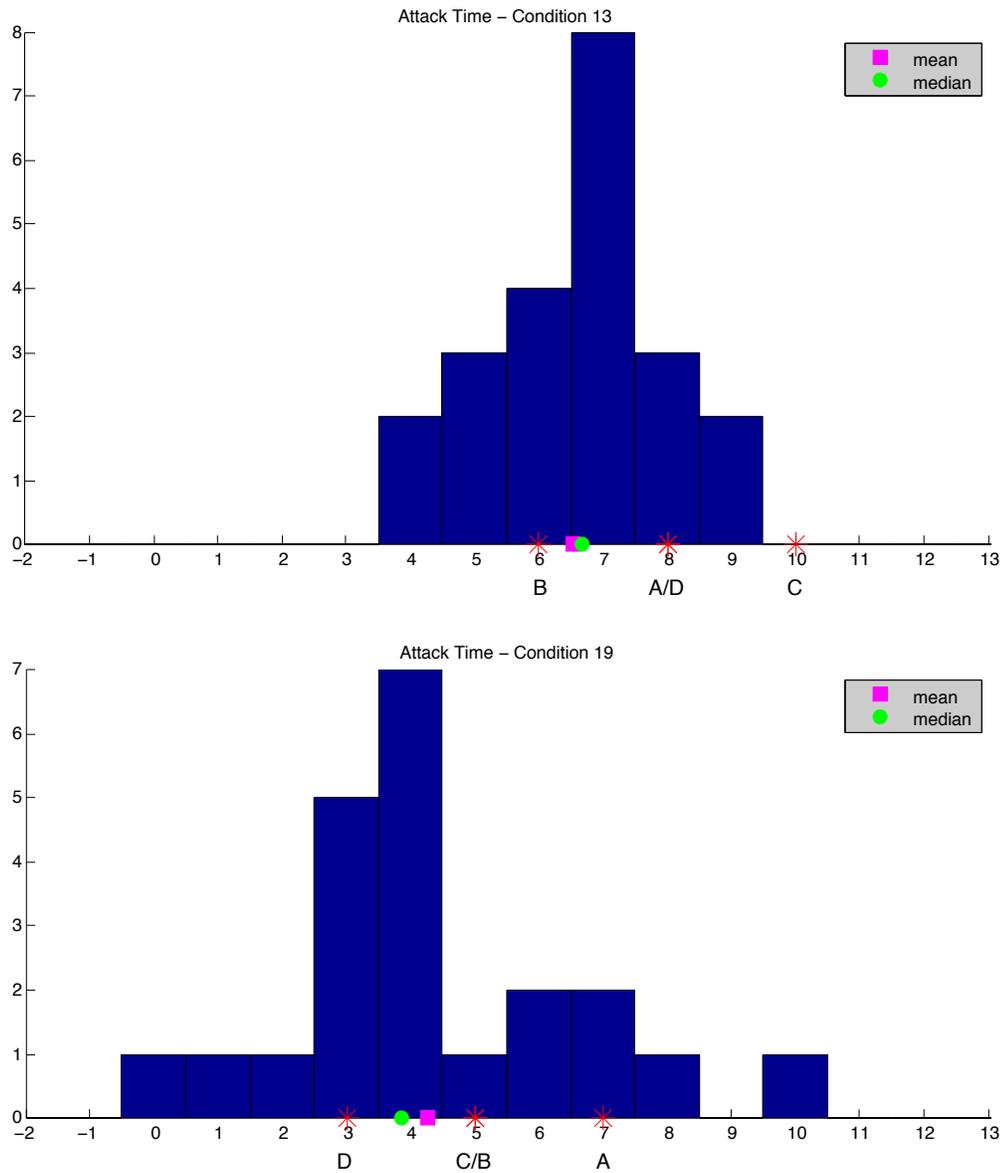


Fig. 4.7 Histograms of conditions 13 and 19 in the attack-time block. Condition 19 shows a smaller than average underestimation.

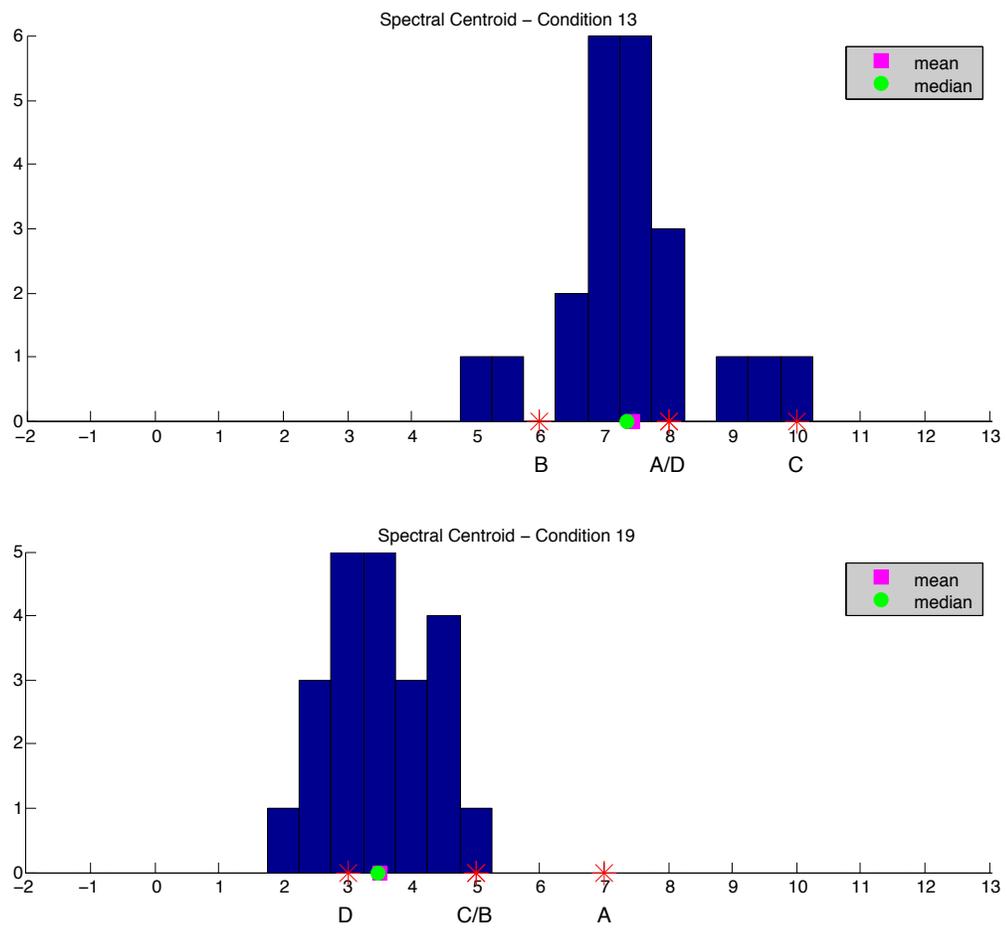


Fig. 4.8 Histograms of conditions 13 and 19 in the spectral-centroid block. Condition 19 shows a smaller than average underestimation.

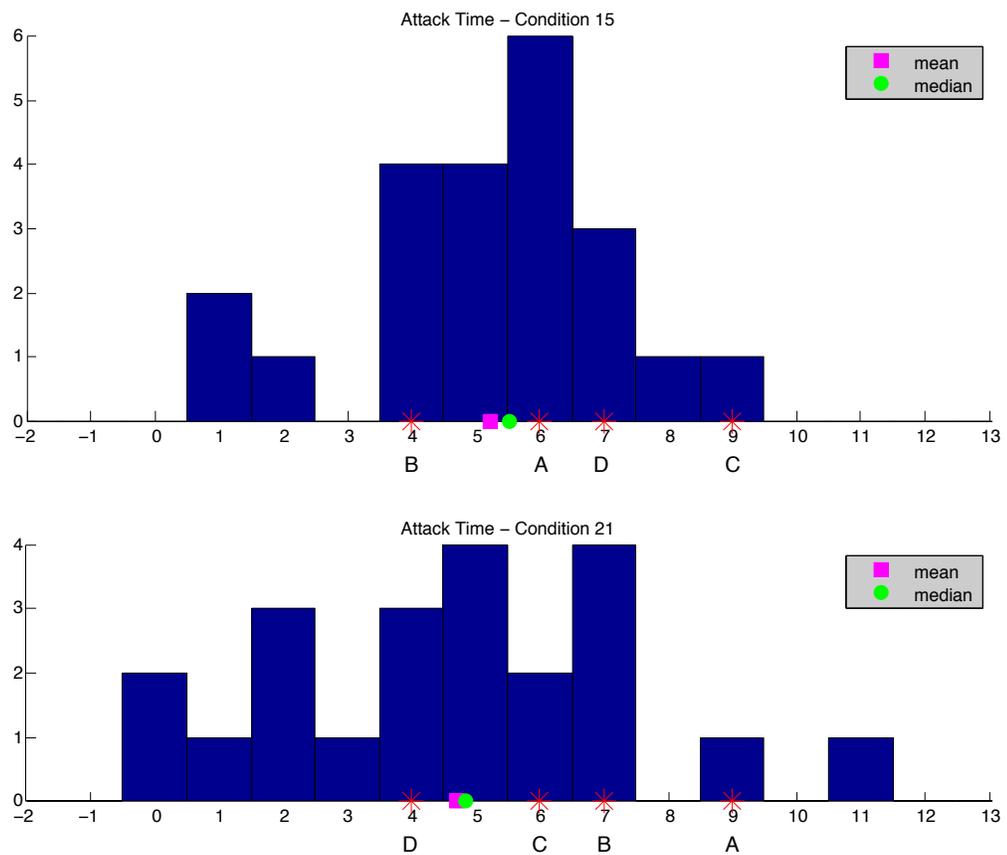


Fig. 4.9 Histograms of conditions 15 and 21 in the attack-time block. Condition 15 shows a larger than average overestimation, whereas condition 21 has a much smaller than average underestimation.

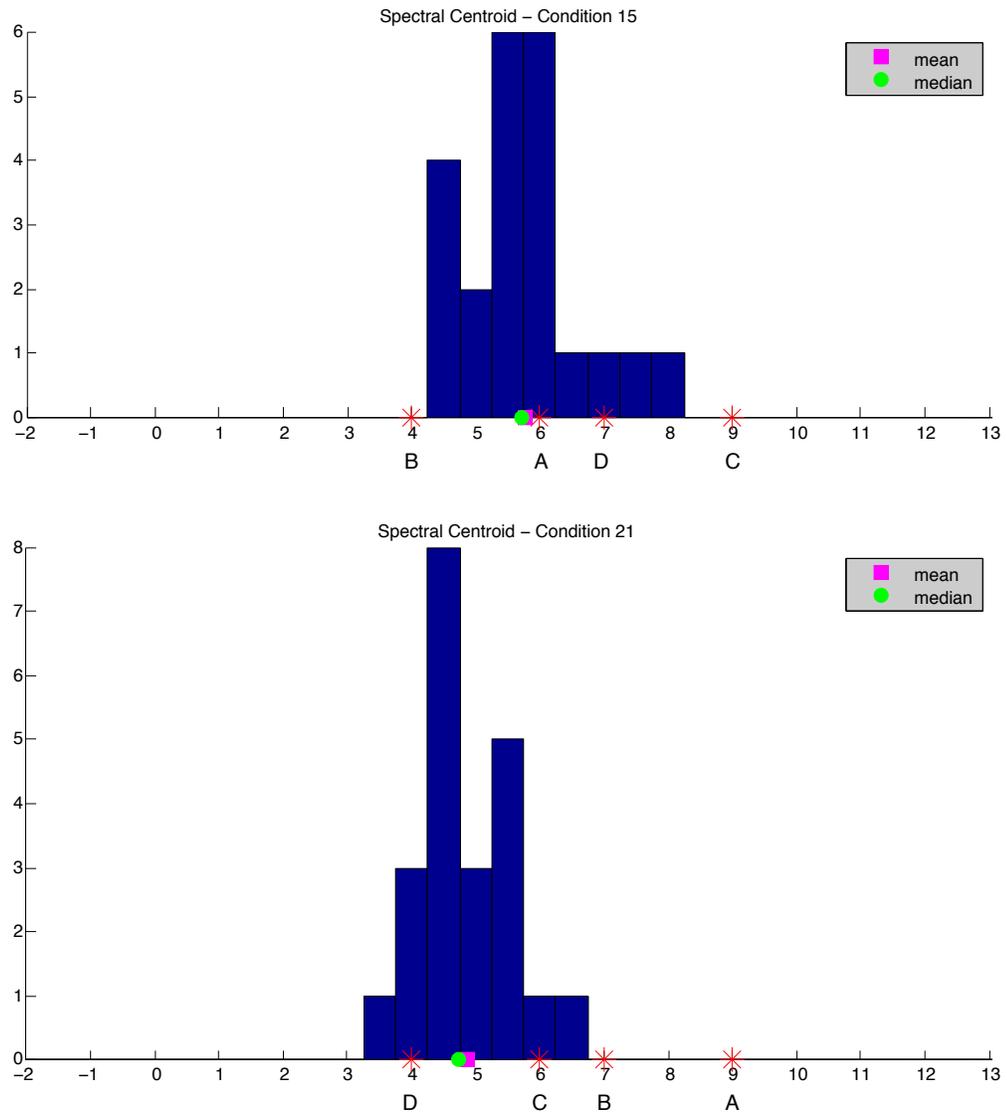


Fig. 4.10 Histograms of conditions 15 and 21 in the spectral-centroid block. Condition 15 shows a larger than average overestimation, whereas condition 21 has a smaller than average underestimation.

It is not entirely apparent why this is, though because these are among the most difficult in the entire experiment (small intervals combined with small transposition sizes makes not only the A:B interval difficult to hear, but also the jump from A to C), it is somewhat expected that the responses be a bit abnormal.

The next significant interaction in the attack-time block is Interval Direction \times Transposition Direction \times Transposition Size (Fig 4.11). For upward interval conditions in this interaction, in essence we see a summary of the major trends found in the data thus far. Congruent conditions have negative estimated marginal means, whereas incongruent conditions have positive means, due to the over/underestimation caused by B biasing. Moreover, this effect increases as the transposition size increases. In downward interval conditions, we see the B biasing effect as well, though with the signs reversed because of the interval direction being flipped. However, in these conditions we do not see a systematic effect of transposition size.

Unlike previous interactions, this one offers a “purer” look at the over/underestimation pattern, because congruent and incongruent conditions are not weighted against one another. This could also explain why we see a clear example of the effects of B biasing increasing as transposition size increases in the case of upward interval conditions. However, as for why such a trend is not seen in downward interval conditions, it might simply be that the B biasing effect is less prominent in this case. One could argue that when the A:B interval is first presented, there is a tendency to emphasize whichever tone is higher in centroid or attack. In the case of downward interval conditions this is tone A, thus undermining the subconscious anchoring of tone B. However, we still see the effects of congruent underestimation and incongruent overestimation, because the B biasing effect is not completely lost. Alternatively, it might simply be the case that transposition size does not have an extremely significant effect at all, as even in the case of upward interval conditions, the effect is slight.

The third and final significant interaction in the attack-time block is Transposition Direction \times Transposition Size \times Interval Size (Fig 4.12). In this interaction, congruent underestimations are averaged with incongruent overestimations, rather than against them, i.e., for a given transposition direction, upward and downward intervals are averaged together, so overestimation and underestimation values will share the same sign. Naturally, this leads to negative estimated marginal means for the upward transposition conditions and positive means for downward transposition.

In the large interval conditions, we see this effect increase with transposition size as well, although in small interval conditions this only occurred with downward transpositions. However, all of these effects are slight, with only the differences between the first and third points (small and large transposition sizes) being statistically significant. The deviation from this trend that we see in small interval conditions with large upward transpositions is largely due to condition 17, an outlier with no overestimation despite its incongruency (see Fig. 4.13). Since incongruent overestimation is defined as a negative value, the mean for incongruent separated conditions is -1.2 for attack time and -0.8 for centroid. However, condition 17 has a mean and median of 0.4 in the attack-time block, and a mean of 0.1 with a median of 0 in the centroid block, which

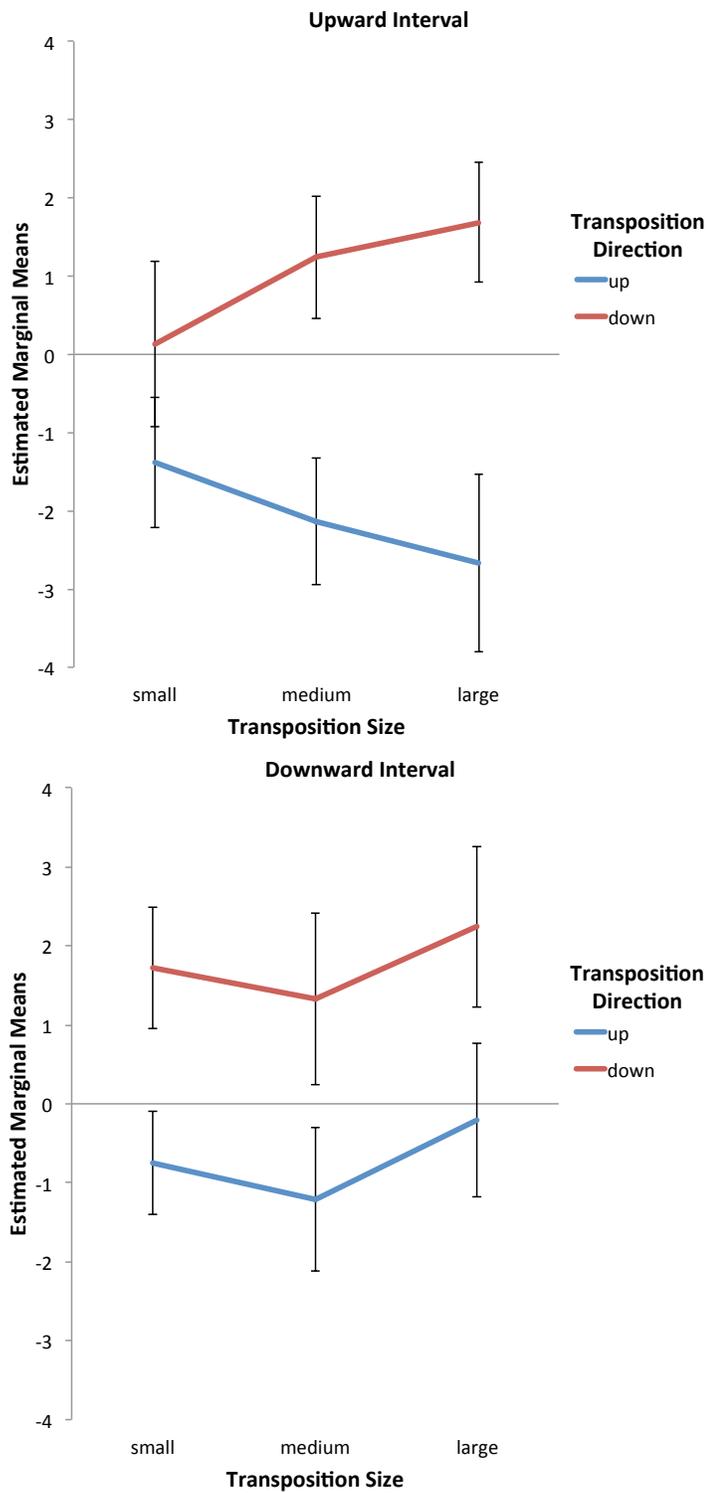


Fig. 4.11 Interval Direction \times Transposition Direction \times Transposition Size in the attack-time block.

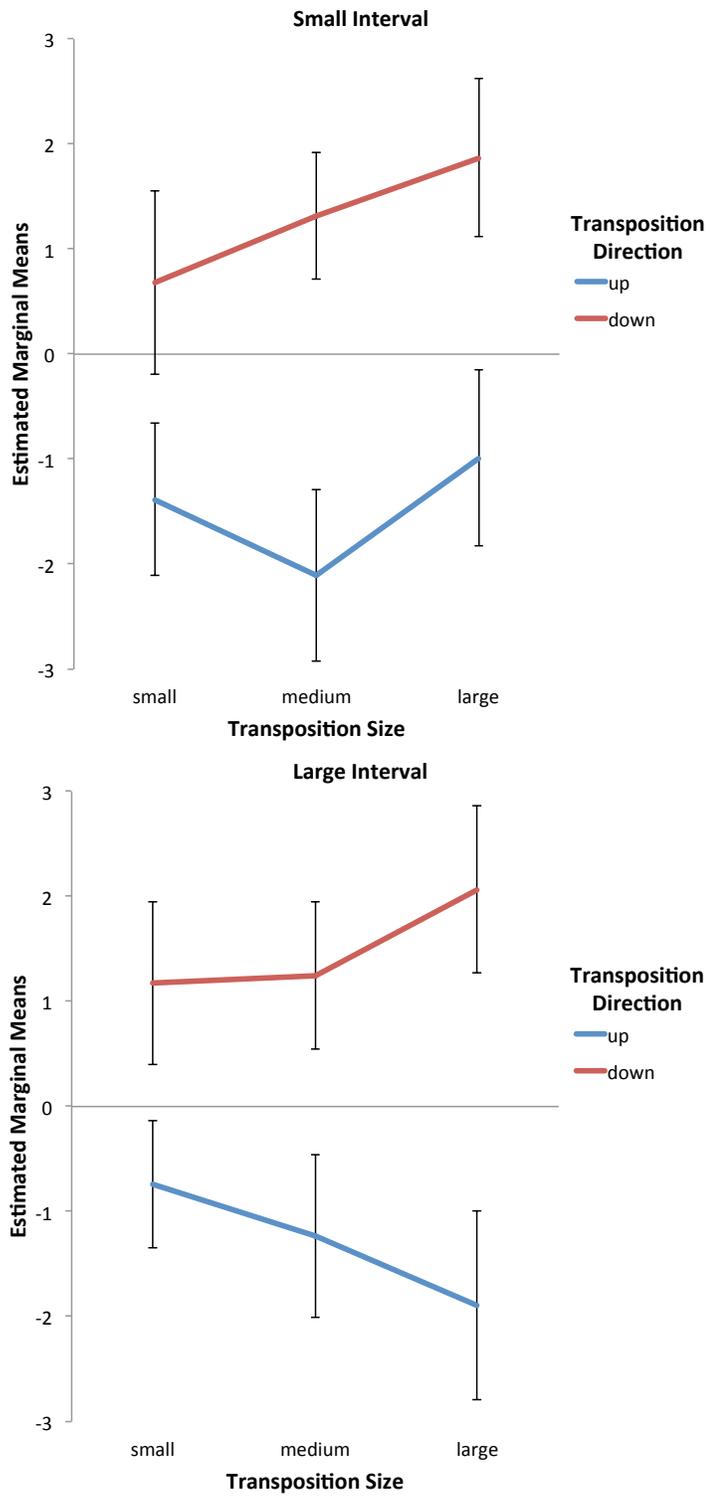


Fig. 4.12 Transposition Direction×Transposition Size×Interval Size in the attack-time block.

explains why we see the same abnormality in those interactions as well (Fig 4.14). The most likely reason for this is that the small A:B interval occurred at the lowest possible point on the scale, with a difference of only 11 ms in the attack-time block and of 0.9 in harmonic rank in the centroid block. Because the scale we used was logarithmic, raw differences between points on the low end are very small. This is likely what curbed the overestimation effect, although it is quite interesting that this resulted in significantly more accurate responses overall.

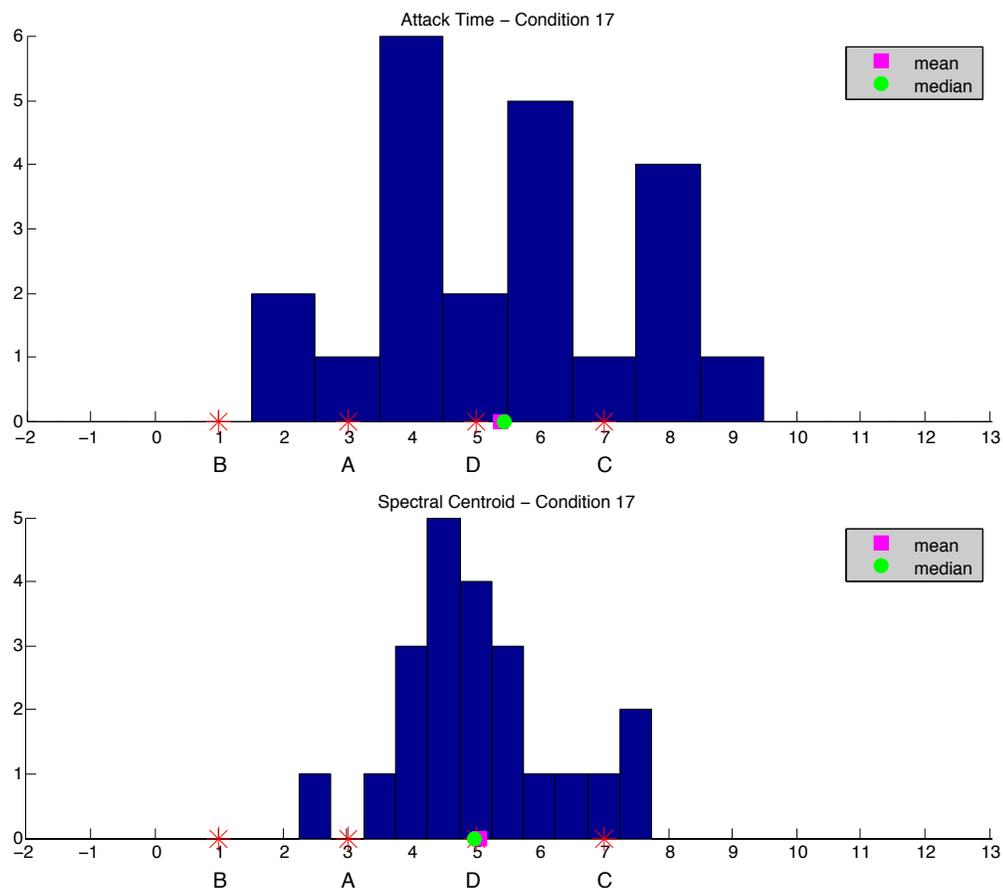


Fig. 4.13 Histogram of condition 17 in both attack-time and spectral-centroid blocks. Despite being incongruent, this condition does not show significant overestimation.

For the sake of brevity, rather than describe all 3-way interactions in the spectral-centroid block in detail, it is sufficient to point out the major ways in which they differ from attack time. Overall, the two blocks are remarkably similar in terms of the interactions seen, which is particularly fascinating as it carries the implication that the trends witnessed are based largely on the placement of the points and the four variable parameters (interval direction and size, transposition direction and size) rather than on the specific acoustic differences between the two blocks. This also suggests that similar cognitive functions are utilized in the performance

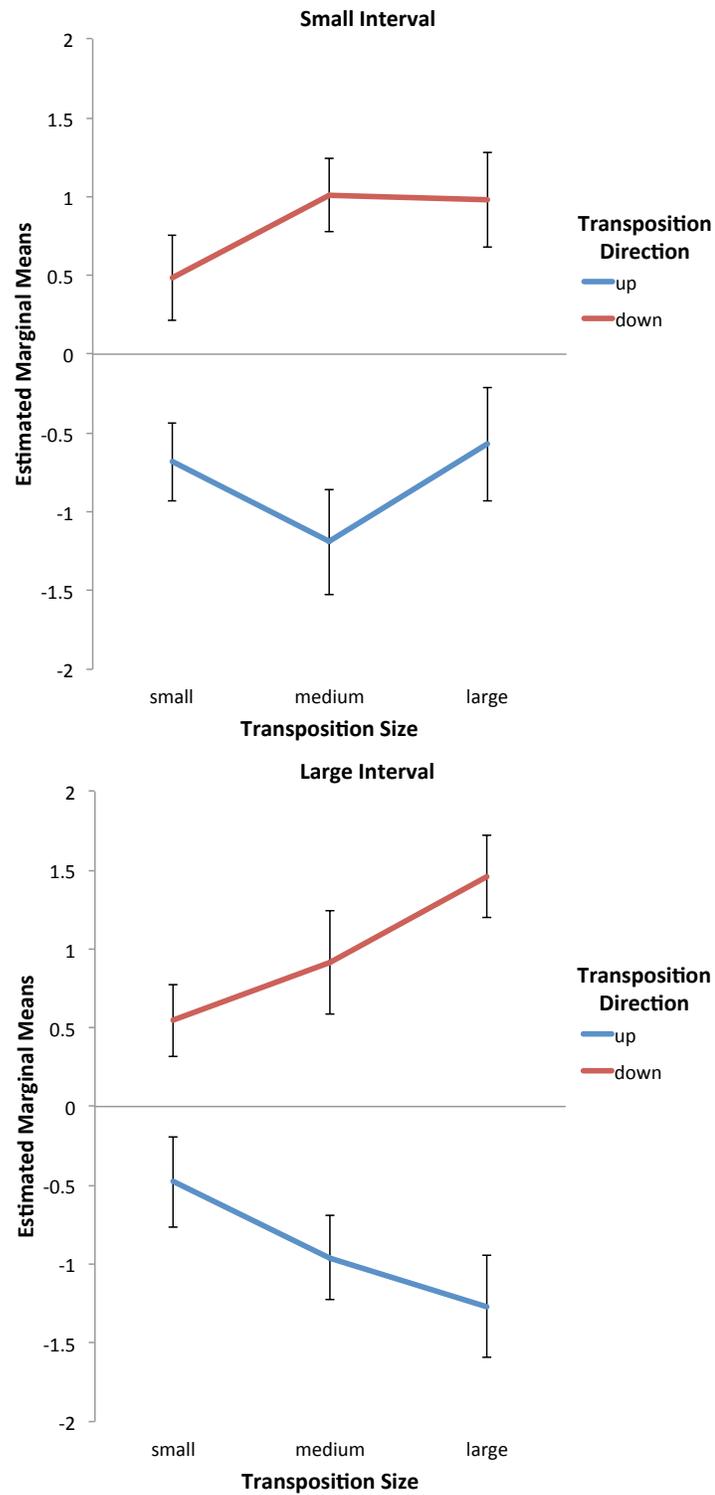


Fig. 4.14 Transposition Direction \times Transposition Size \times Interval Size in the spectral-centroid block.

of this task, regardless of dimension. It would be interesting to see whether similar trends could be observed in other contour-based dimensions such as loudness, or even whether dimensions as concrete as pitch would show similar B biasing effects.

As for differences between the two blocks, we see a different pattern in the downward transposition, downward interval conditions in the Interval Direction \times Transposition Direction \times Transposition Size interaction (compare Fig. 4.11 with Fig. 4.15). In the centroid block, there is a slight effect of transposition size on estimated marginal means, showing a positive relationship between the two. Normally we would attribute this to increasing B to D distance and the effects of B biasing, although this trend is not seen in downward interval conditions in the attack-time block, nor is it seen in downward interval, upward transposition conditions in either block. It is arguable that downward intervals naturally disrupt the tendency to anchor onto point B, in such a way that the B biasing effect does not increase with increasing transposition size. The reason why we do in fact see this trend in the downward transposition conditions is unclear, although it is clear that the small transposition conditions are responsible for the interaction in this case, as they are significantly different from both medium and large transpositions. The conditions that make up this estimated marginal mean are conditions 19 and 20, shown in Fig. 4.16. Condition 19 in particular has a mean and median underestimation of 0.5, which is considerably smaller than the average for concatenated congruent conditions in the centroid block (1.3). Underestimations in condition 20 (mean 0.8, median 0.6) are less than the average for interleaved congruent conditions (1.2), although not by as much. Because both of these component conditions have smaller than average underestimations, this explains why the estimated marginal means are significantly smaller in the small transposition conditions of this interaction in the spectral-centroid block.

Excepting this particular case, unusual or unexpected trends in the interactions as a result of outlier conditions tend to be somewhat consistent regardless of the block. This may lend even more support to the idea that similar cognitive functions were used throughout the experiment. The fact that small and medium transposition, downward interval conditions have much lower means than expected in the Interval Direction \times Transposition Size \times Interval Size interaction is a great example of this parallelism. We have already reviewed the cases where outlier conditions disrupt predicted trends in downward interval conditions, for both attack time and spectral centroid. However, it is worth noting that the upward interval condition graphs look remarkably similar as well. The main difference is that whereas none of the differences between data points are significant in the attack-time block, there is a significant difference between small and large intervals in medium and large transposition conditions in the spectral-centroid block. It is unclear whether this is the result of transposition size influencing the B biasing effect, given that this trend is not typically present in interactions that weigh congruent underestimation against incongruent overestimation. Regardless of the cause, the difference is just barely significant, and overall this interaction is remarkably similar across attack-time and spectral-centroid blocks.

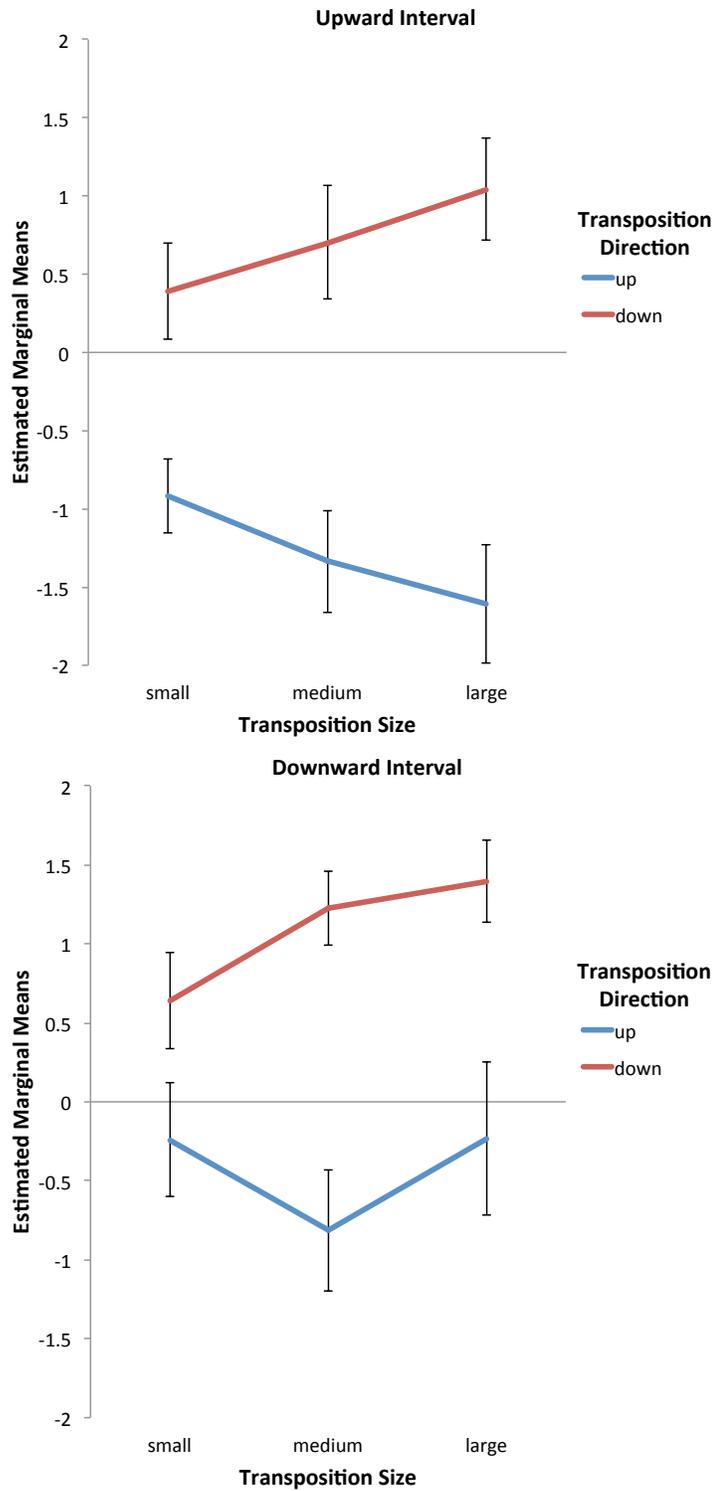


Fig. 4.15 Interval Direction \times Transposition Direction \times Transposition Size in the spectral-centroid block

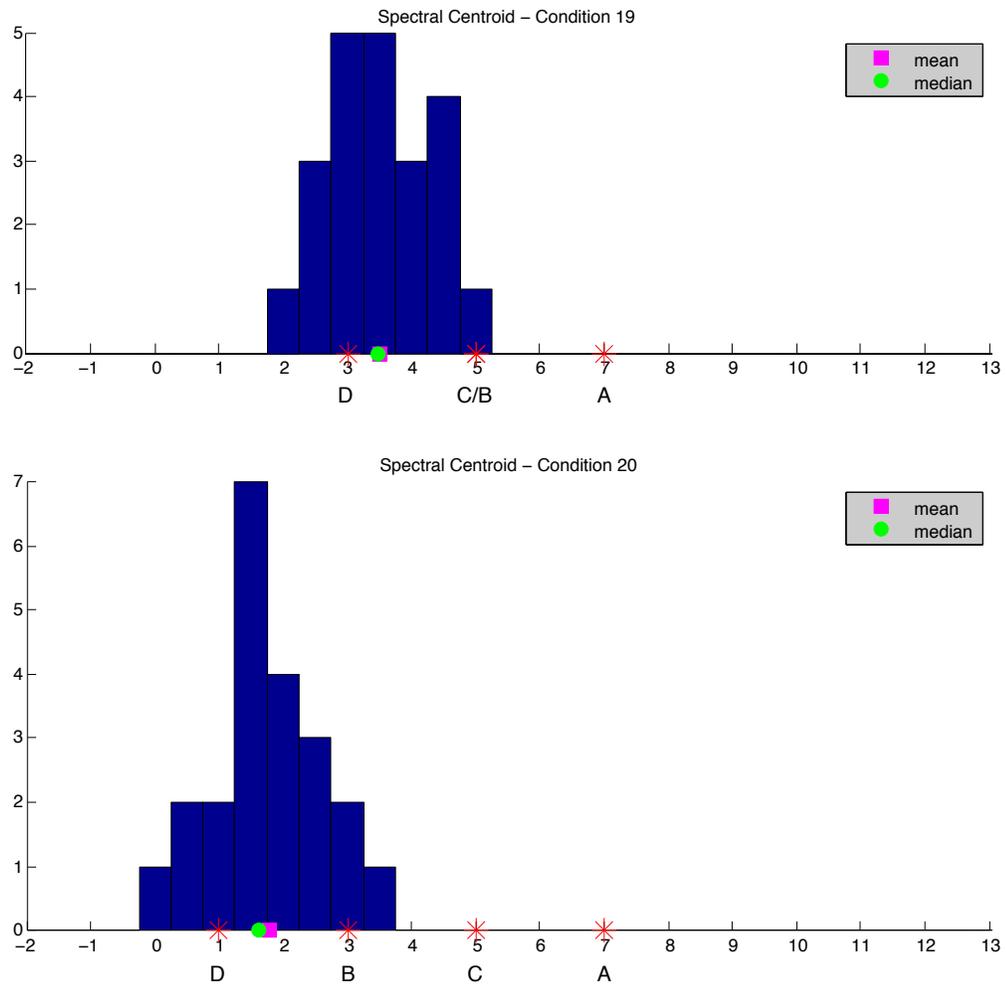


Fig. 4.16 Histogram of conditions 19 and 20 in the spectral-centroid block. Both conditions show smaller than average underestimations.

Chapter 5

Conclusion

The experiment in this paper produced a number of important results about the nature of timbre interval perception, specifically along the dimensions of attack time and spectral centroid. Overall, our findings suggest that people are largely unable to precisely reproduce specific amounts of change along these dimensions, as they can with pitch. However, we found that contour is an important cognitive schema used in the performance of timbre transposition tasks and that participants correctly reproduce the direction of the A:B interval with high frequency. This is especially true in incongruent conditions or conditions with large intervals, as these conditions present higher relative amounts of change along the dimension in question.

Another important finding is that performance of this task increased significantly in the spectral-centroid block—responses were more accurate and showed a greater degree of consensus as well. It appears that the perceptual resolution of this dimension is higher than that of attack time. Lower standard deviations in the spectral-centroid block suggest that participants relied on notions of position at least to a certain extent in their completion of the task, instead of relying purely on contour judgments. The reason for this may have to do with the fact that centroid and pitch both share the same underlying sensory continuum of the tonotopic representation of frequency in the auditory system (McAdams & Bregman, 1979).

The most surprising finding in this experiment was the tendency to underestimate congruent conditions and overestimate incongruent ones, a trend which is most likely caused by biasing as a result of cognitive anchoring to the B stimulus. This also explains the tendency for this effect to increase as a function of the distance from B to D, though the fact that this relationship is weaker for downward interval conditions suggests that participants are less likely to anchor to B if it is the “lesser” of the first two presented stimuli (A:B).

Another interesting result is that the interactions found were remarkably similar across the attack-time and spectral-centroid blocks. This result may indicate that similar cognitive functions are responsible for the recognition and reproduction of intervals, regardless of dimension. Indeed, it would be interesting to conduct the present experiment with other dimensions such as pitch or loudness, to see if the same biases and interactions would be present. On the other hand, dimensions such as even harmonic attenuation, that do not necessarily have an inherent up or

down association, could be revealing as well.

It is not entirely clear whether the similarities between interactions in this study are purely the result of the same cognitive mechanisms being used to process our interval condition parameters (interval direction and size, transposition direction and size), or whether they are simply due to biases associated with the method employed. One possible way to address the B biasing effect in particular might be to present the intervals as A:B:A and C:D:C. Sandwiching the B stimulus between two A stimuli might reduce the tendency to subconsciously anchor to it, similar to how downward A:B intervals appear to have this effect in the present study.

It would also be beneficial to see a larger study performed using more interval and transposition sizes, or using non-musician participants as well. The reason we chose musicians in the current study was to see whether people who presumably already possess sufficient musical training to transpose pitch intervals could do so with timbre as well. Introducing more participants or more interval and transposition sizes might allow us to further refine the perceptual scales of attack time and spectral centroid. In the present study it is difficult to offer any conclusive statements on the accuracy of the logarithmic scales used, in light of relatively high standard deviations in responses overall. Ideally, however, it would be possible to create a scale in which the root-mean-squared error is closer to the standard deviation, or for which congruent underestimations and incongruent overestimations are distributed more evenly around the expected point. One interesting method might be to determine the just-noticeable difference (JND) of attack time and spectral centroid, and to repeat the current experiment using a scale with interval and transposition sizes expressed in JNDs.

Much remains to be explored with respect to transposition of timbre intervals, both in terms of refining our current understanding of attack time and spectral centroid, and in terms of exploring other dimensions or combining multiple dimensions to create 2D or 3D timbre spaces. The simple 1D experiment in this paper proved to be an effective tool for focusing on the perceptual properties of a single dimension, and it is our hope that the findings in this study will provide a foundation or a basis of comparison for future research in this area.

References

- Attneave, F. & Olson, R. K. (1971). Pitch as medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 84(147-166).
- Bartlett, J. C. & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6(3), 501–515.
- Bey, C. & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 175–191.
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, 118(1), 471–482.
- Cuddy, L. L. & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 28, 255–270.
- Ehresman, D. & Wessel, D. (1978). Perception of timbral analogies. *IRCAM Technical Report*, 13.
- Gescheider, G. A. (1976). *Psychophysics: Method and Theory*. Lawrence Erlbaum Associates.
- Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 751–763.
- Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In Nielzén, S. & Olsson, O. (Eds.), *Structure and perception of electroacoustic sound and music*, (pp. 43–54). Excerpta Medica.
- Krumhansl, C. L. & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 739–751.

- McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, 4(1-2), 181–198.
- McAdams, S. & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, 4(4), 26–43.
- McAdams, S. & Cunibile, J. C. (1992). Perception of timbral analogies. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences.*, 336(1278), 383–389.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58, 177–192.
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, 19(12), 1263–1271.
- Melara, R. D. & Marks, L. E. (1990a). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48(2), 169–178.
- Melara, R. D. & Marks, L. E. (1990b). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, 18(5), 477–495.
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 976–986.
- Russo, F. A. & Thompson, W. F. (2005a). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, 67(4), 559–568.
- Russo, F. A. & Thompson, W. F. (2005b). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review*, 12(6), 1068–1075.
- Smith, B. K. (1995). Psiexp: An environment for psychoacoustic experimentation using the ircam musical workstation. In *Proceedings of the Society for Music Perception and Cognition Conference*.
- Tillmann, B. & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1131–1142.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Appendix

Condition	Point Placement	Congruency	Int. Size	Int. Dir.	Trans. Size	Trans. Dir.
1	concatenated	congruent	small	up	small	up
2	interleaved	congruent	large	up	small	up
3	separated	congruent	small	up	medium	up
4	interleaved	congruent	large	up	medium	up
5	separated	congruent	small	up	large	up
6	concatenated	congruent	large	up	large	up
7	concatenated	incongruent	small	up	small	down
8	interleaved	incongruent	large	up	small	down
9	separated	incongruent	small	up	medium	down
10	interleaved	incongruent	large	up	medium	down
11	separated	incongruent	small	up	large	down
12	concatenated	incongruent	large	up	large	down
13	concatenated	incongruent	small	down	small	up
14	interleaved	incongruent	large	down	small	up
15	separated	incongruent	small	down	medium	up
16	interleaved	incongruent	large	down	medium	up
17	separated	incongruent	small	down	large	up
18	concatenated	incongruent	large	down	large	up
19	concatenated	congruent	small	down	small	down
20	interleaved	congruent	large	down	small	down
21	separated	congruent	small	down	medium	down
22	interleaved	congruent	large	down	medium	down
23	separated	congruent	small	down	large	down
24	concatenated	congruent	large	down	large	down

Table 5.1 Each interval condition used in the experiment along with its associated point placement category, congruency, interval size and direction, and transposition size and direction

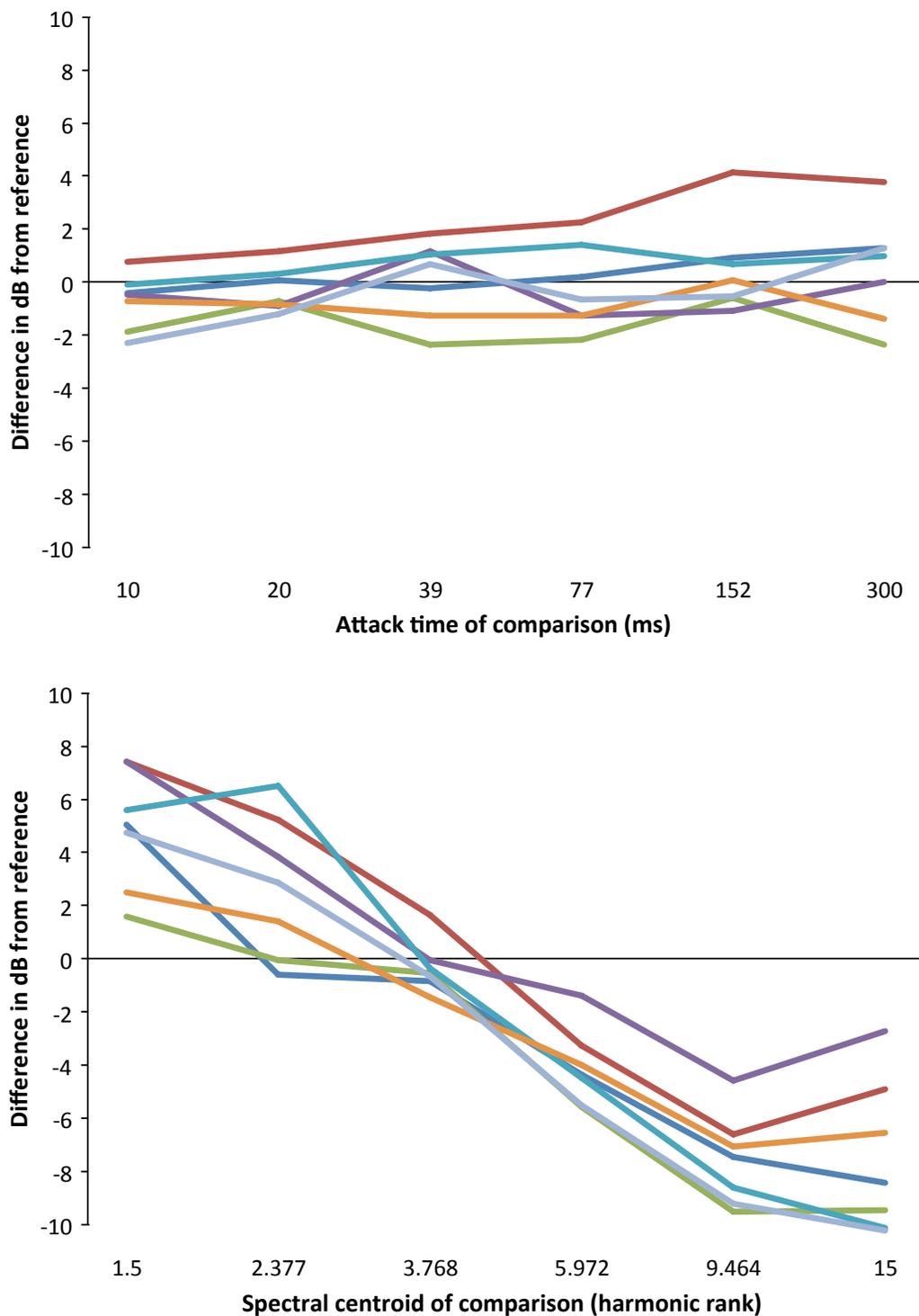
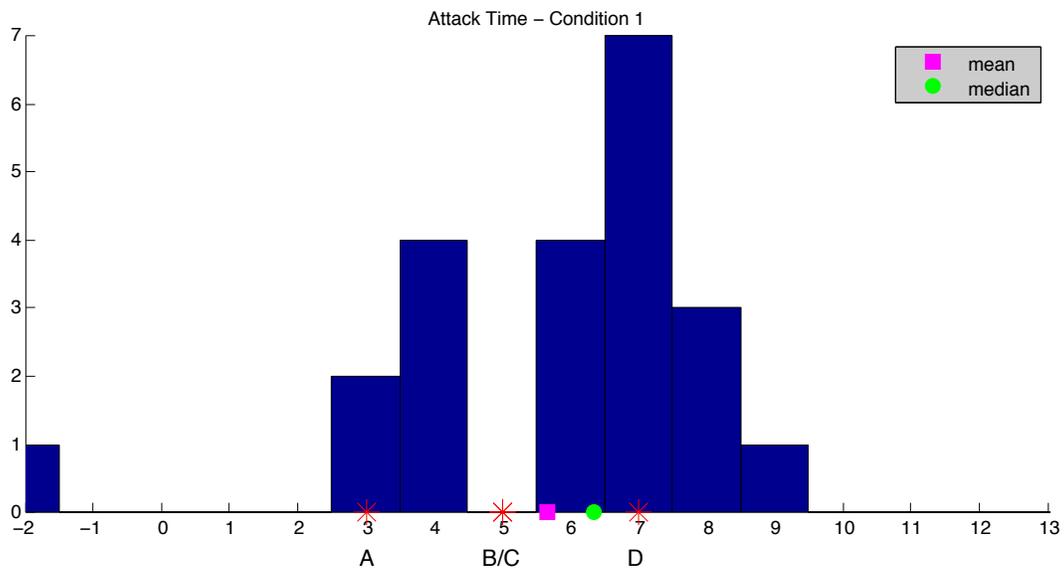


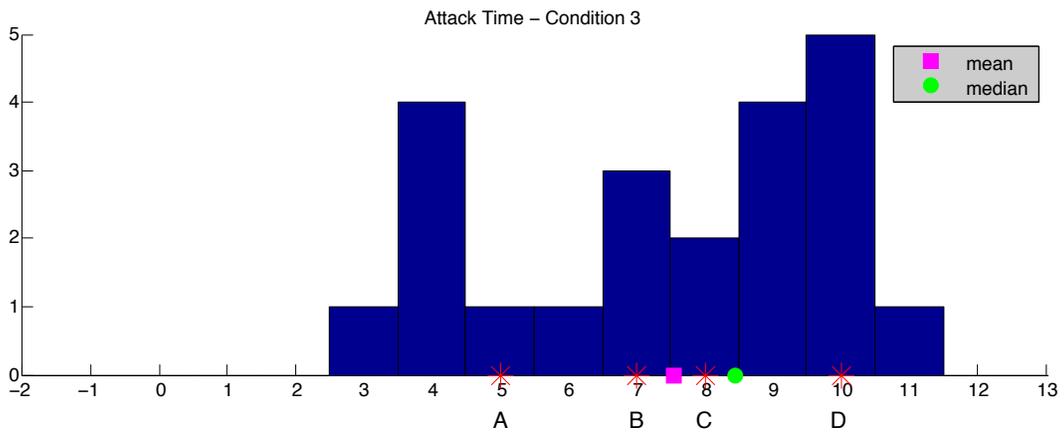
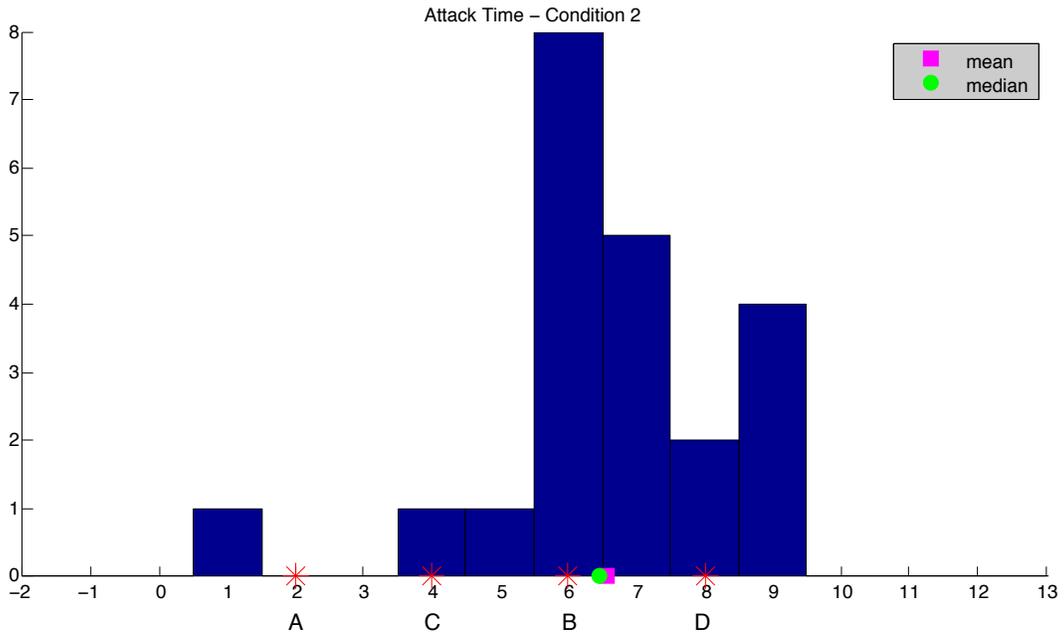
Fig. 5.1 Results of the loudness matching experiment for both attack time and spectral centroid, respectively. Each line represents a different participant.

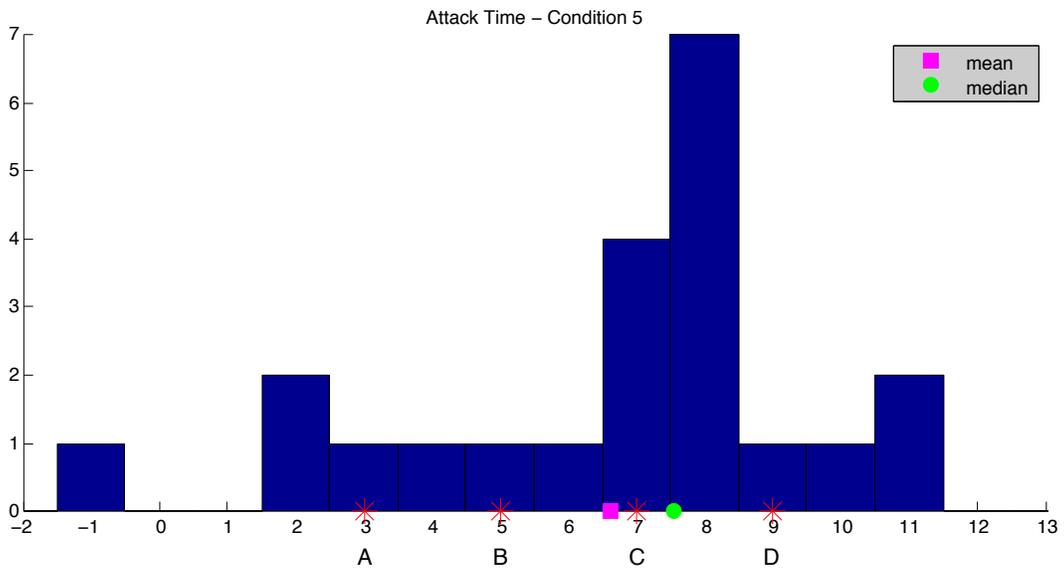
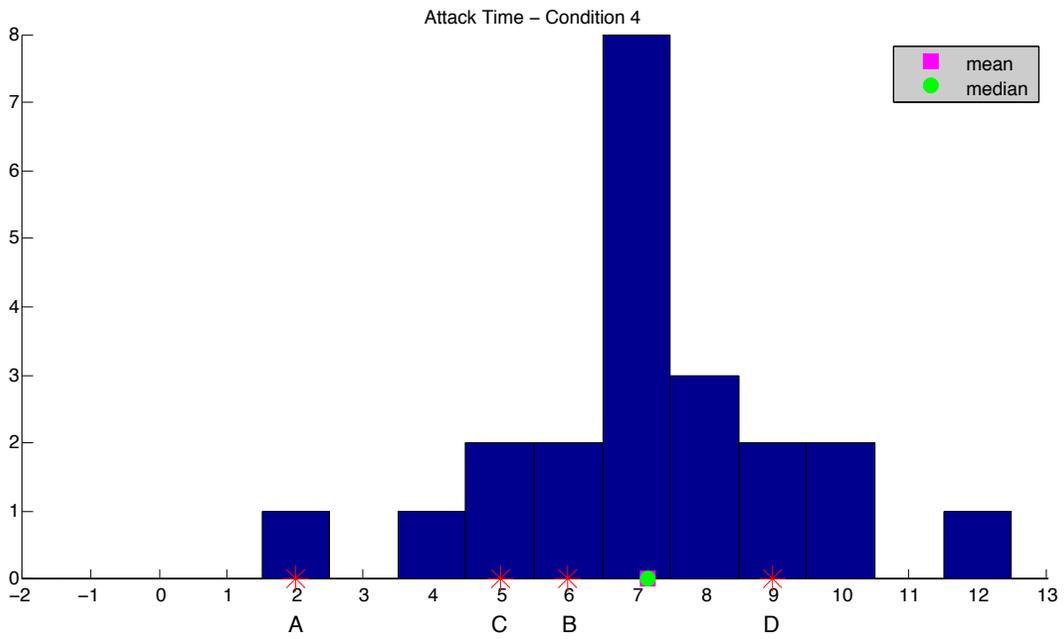
Instrument	No. of Subjects
Cello	1
Clarinet	1
Double bass	1
Flute	2
Percussion	1
Piano	6
Trombone	1
Trumpet	1
Violin	1
Voice	5

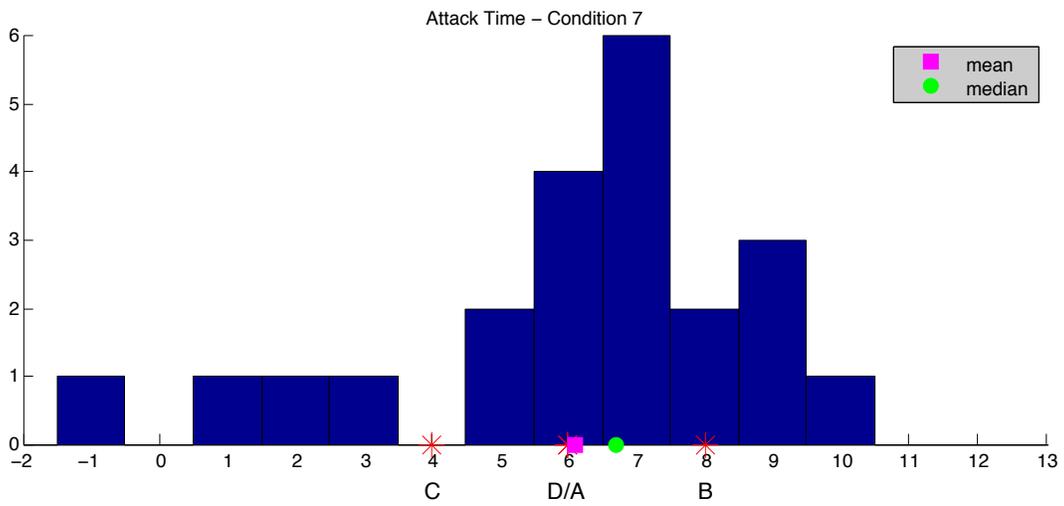
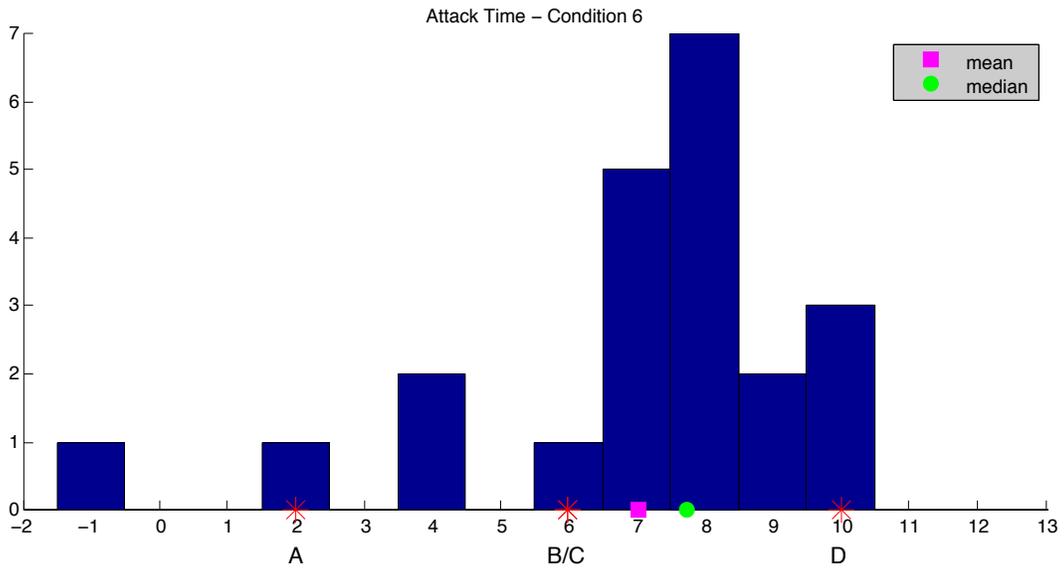
Table 5.2 The primary instruments played by participants. Only 20 of the participants provided a primary instrument on the questionnaire.

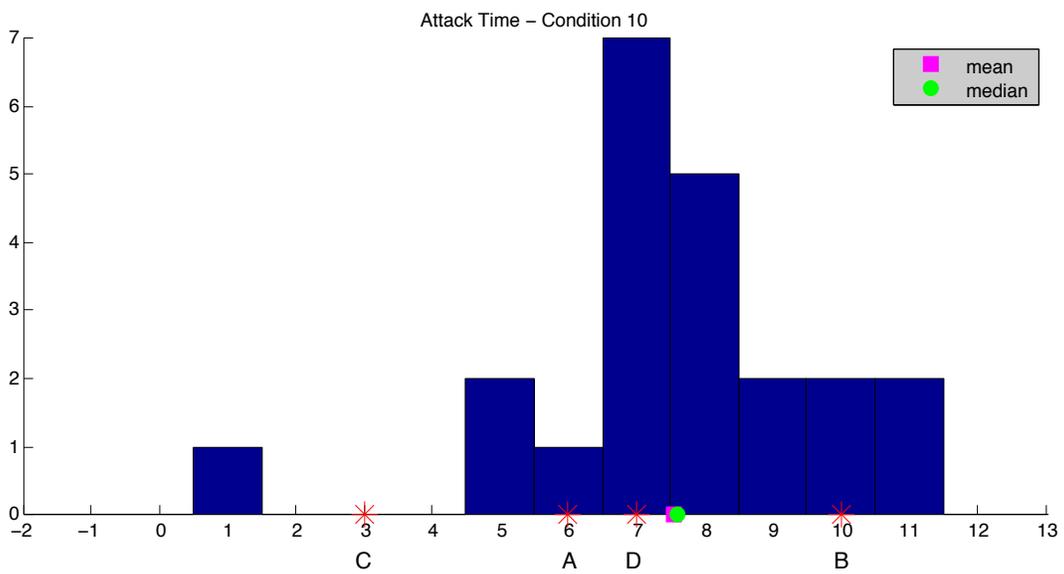
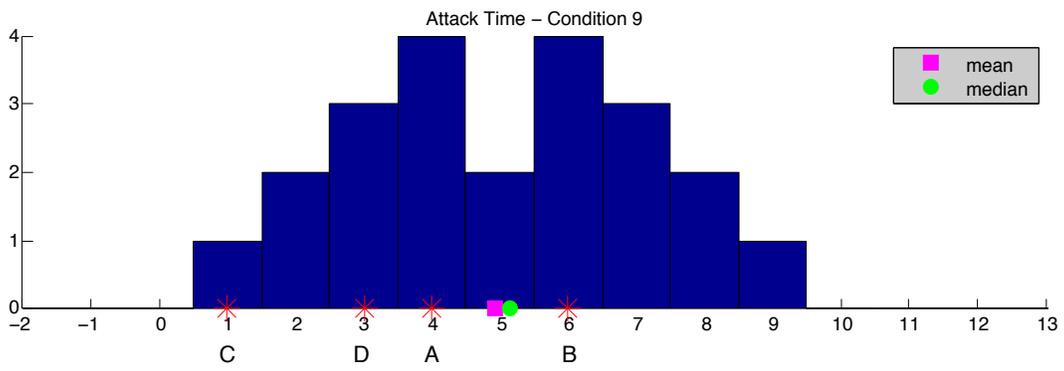
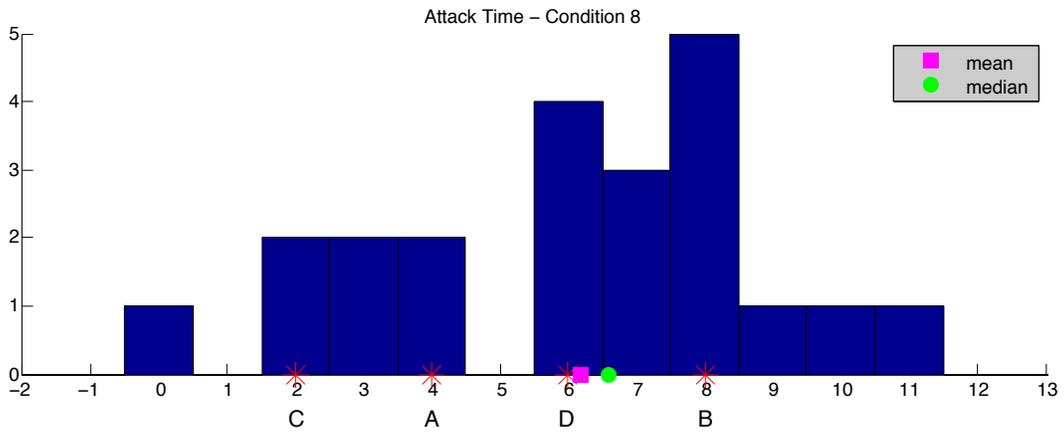
Attack-time condition histograms

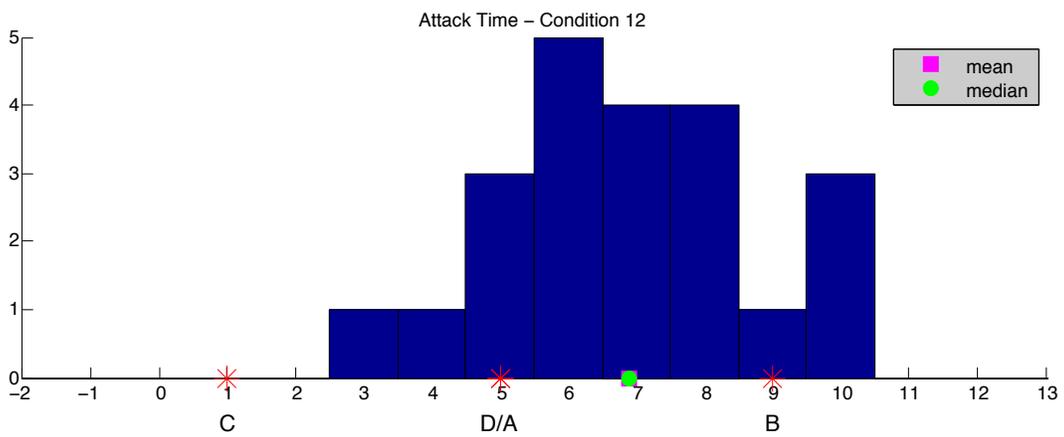
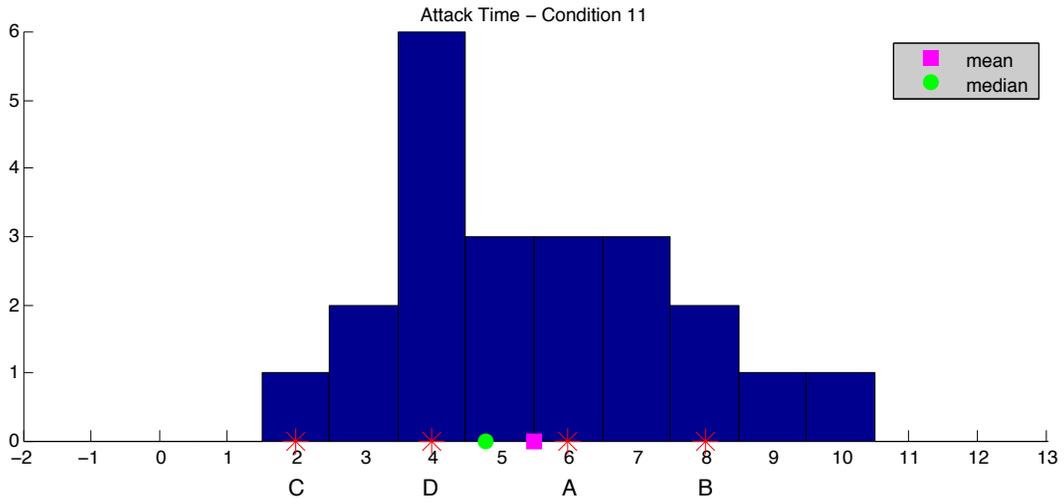


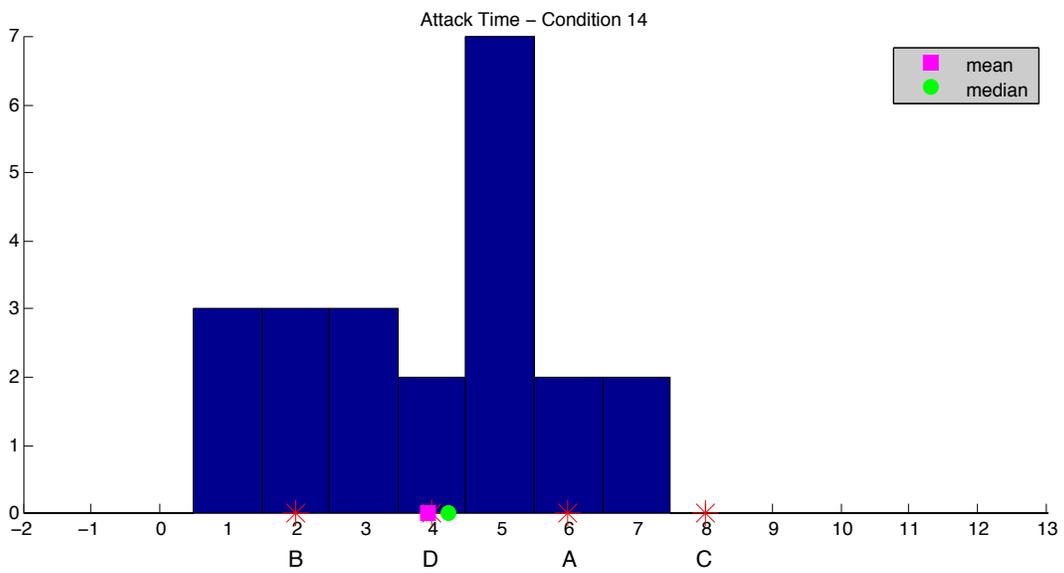
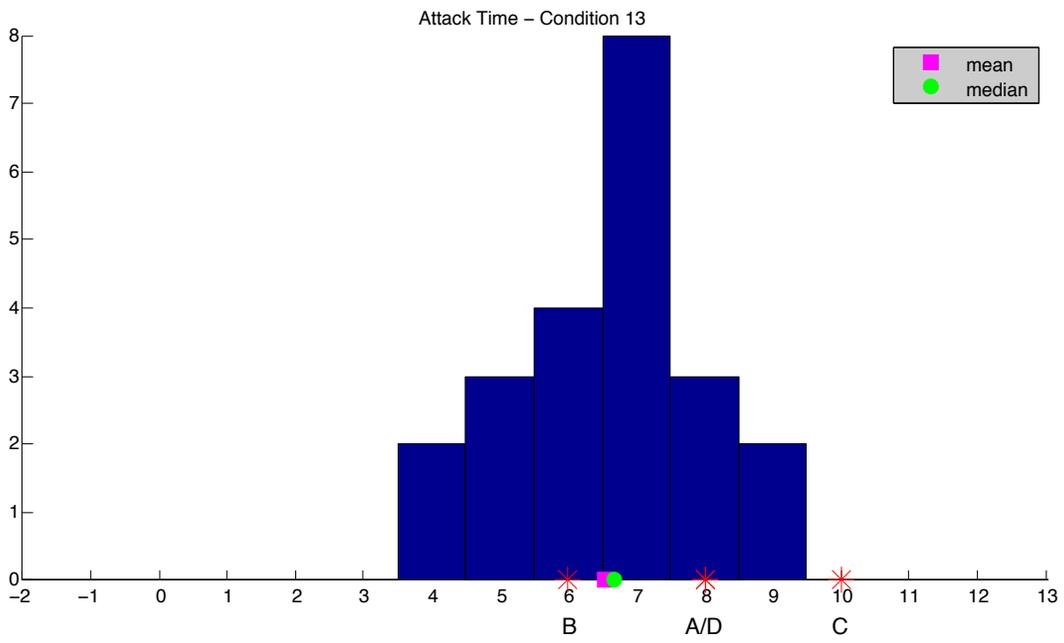


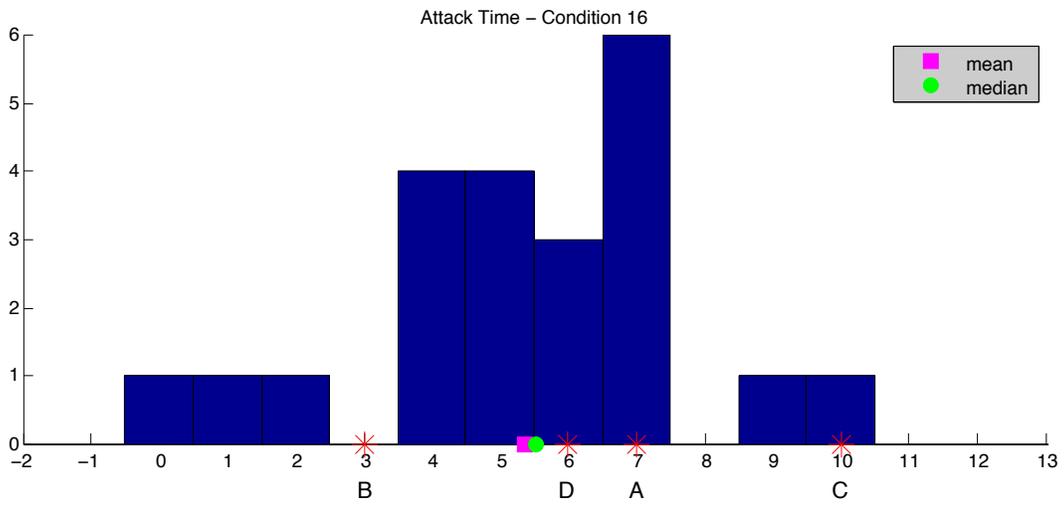
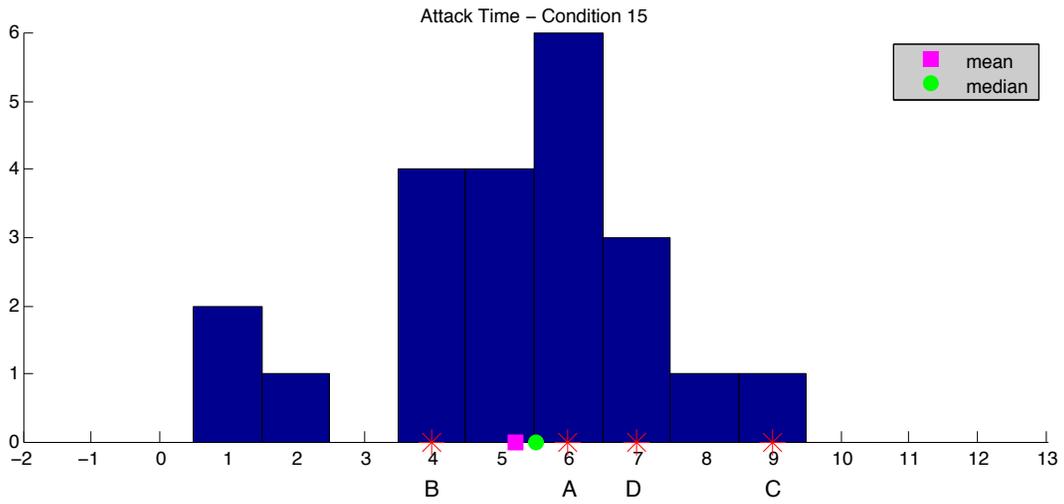


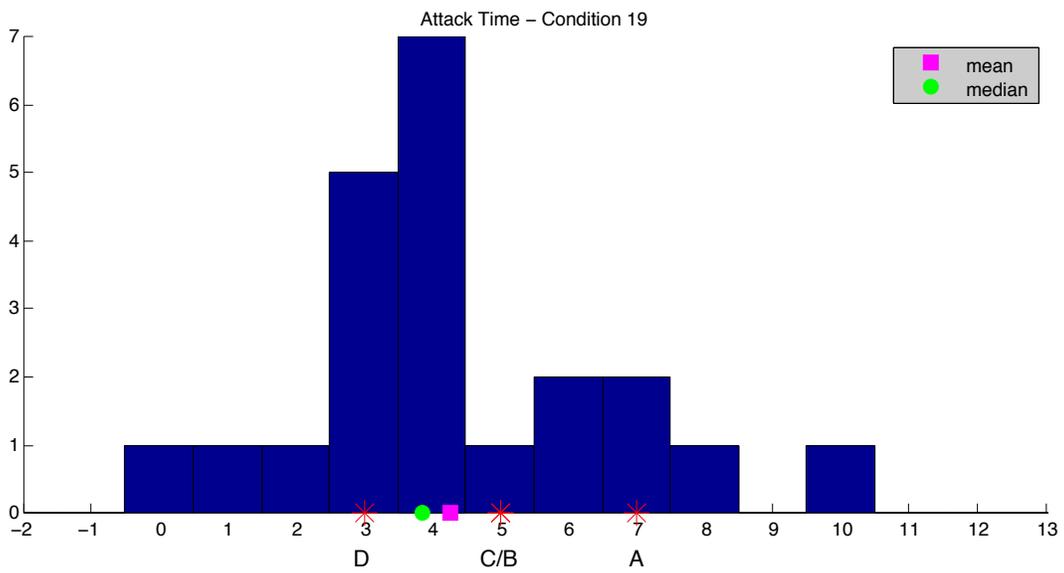
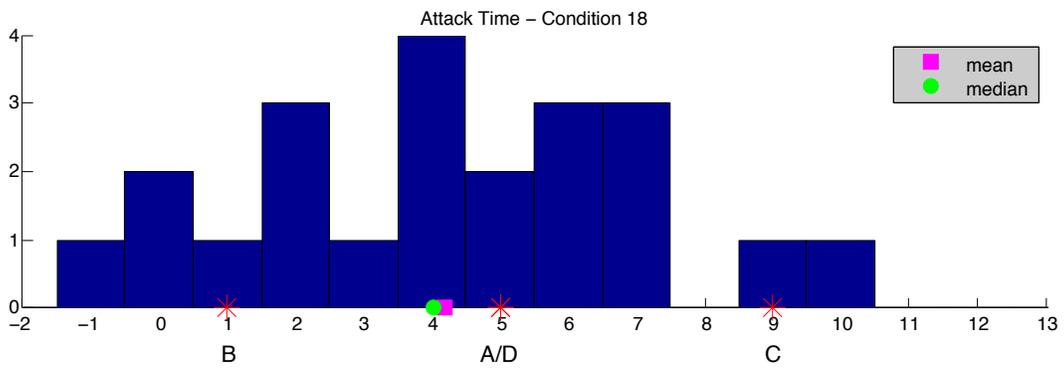
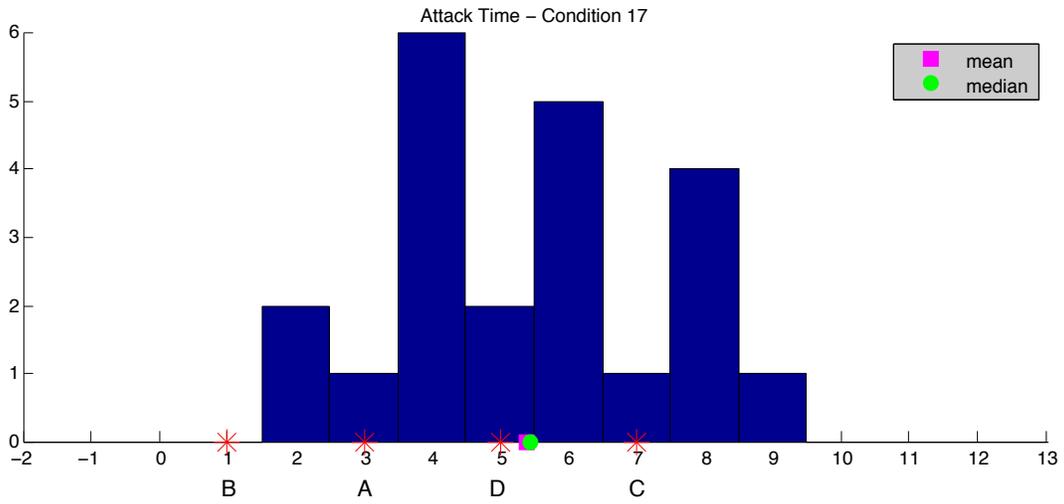


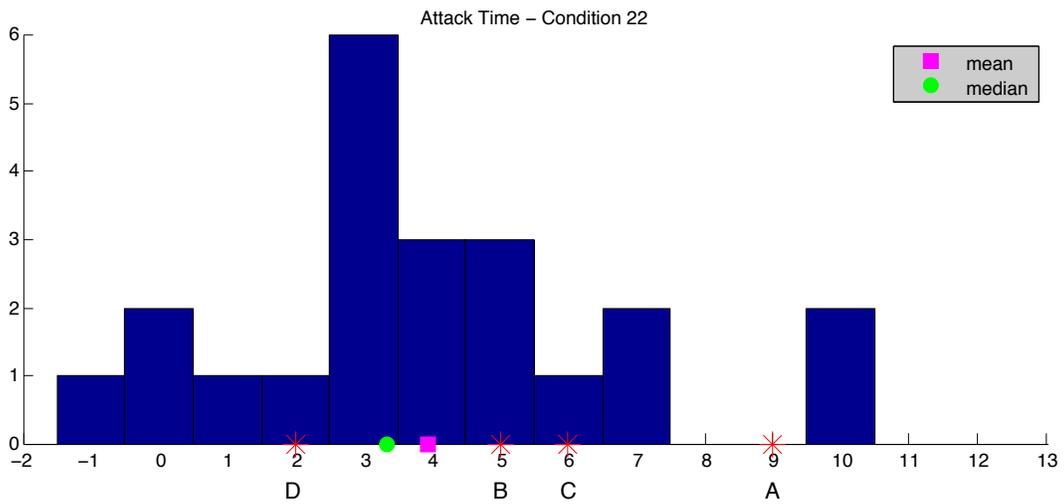
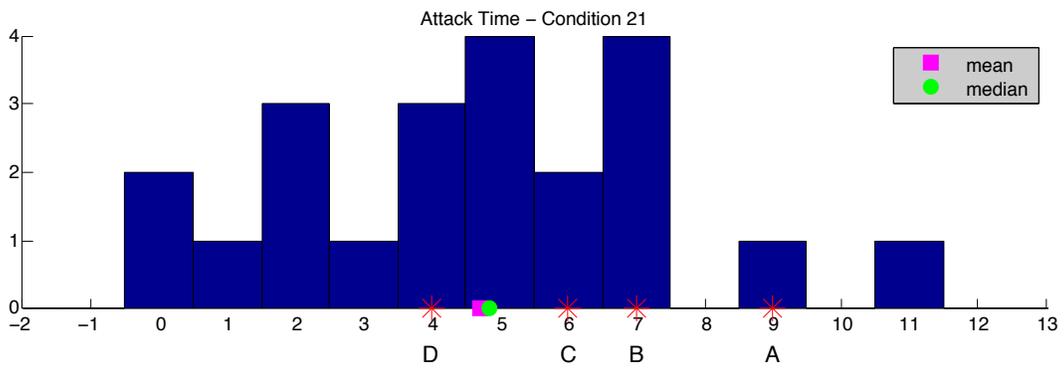
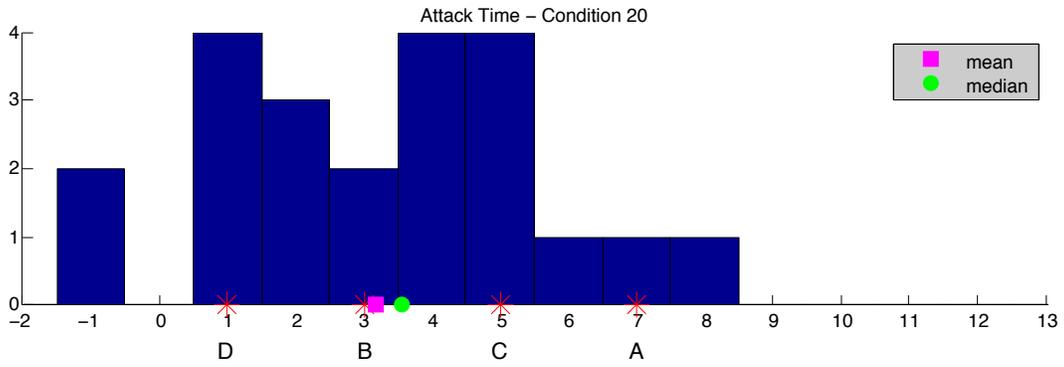


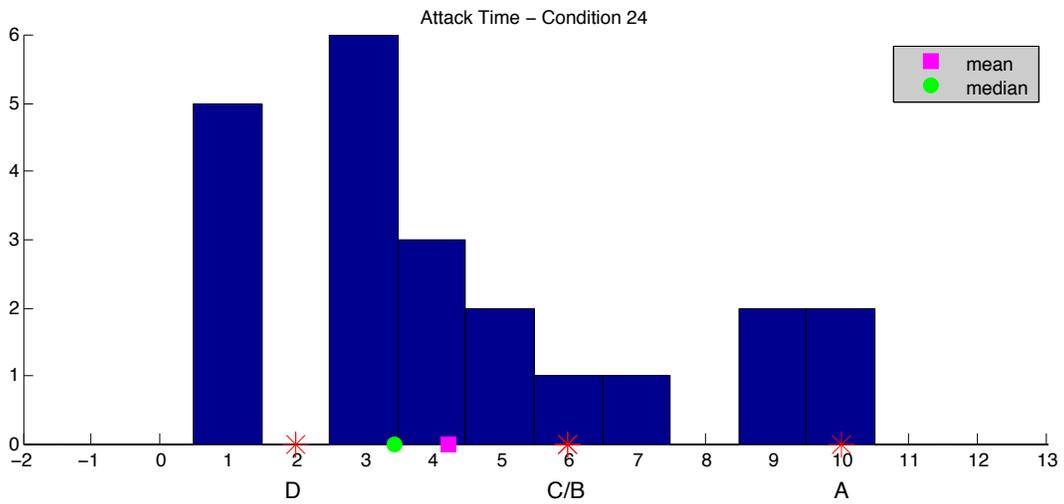
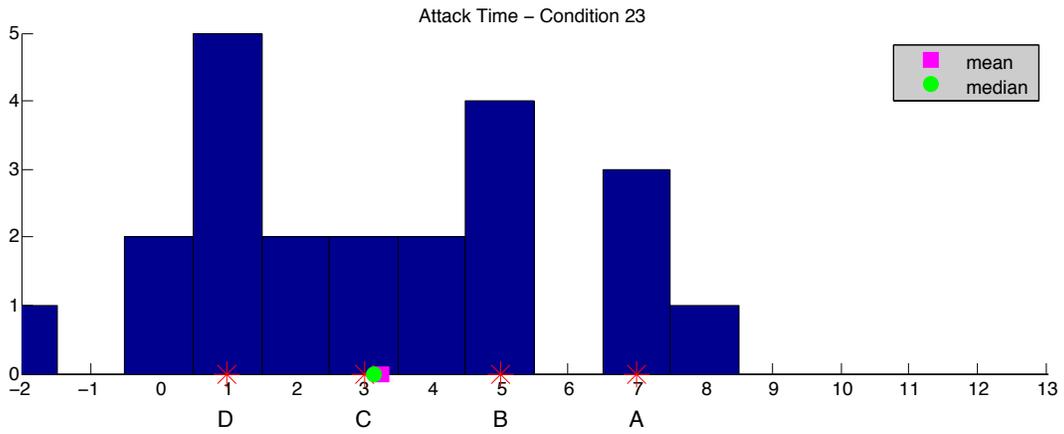












Spectral-centroid condition histograms

