Grouping by proximity and grouping by similarity in audition

Minhong Yu
Shanghai, China

B.S., Zhejiang University, 2005
M.Ed., Zhejiang University, 2007
M.A., University of Virginia, 2010

A Dissertation presented to the Graduate Faculty
of the University of Virginia in candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia

June, 2013

Michael Kubovy

William Epstein

Timo von Oertzen

Judith Shatin

Abstract

By using visual dot lattices as a tool, Kubovy and van den Berg (2008) studied two classic Gestalt grouping principles — grouping by proximity and grouping by similarity — in vision and found surprising additive effects between the two principles.

This dissertation is aimed at building an auditory analogy to explore the effects of these two grouping principles in audition. We used auditory necklaces — ambiguous auditory patterns that we have been developing (Yu & Kubovy, submitted) — as a tool and studied grouping by temporal proximity, grouping by loudness similarity, and grouping by pitch similarity in audition. In Experiment 1, we examined the effect of grouping by temporal proximity alone. In Experiment 2, we examined the separate and conjoint effects of grouping by temporal proximity and grouping by loudness similarity. In Experiment 3, we examined the separate and conjoint effects of grouping by temporal proximity and grouping by pitch similarity.

The results showed that as individual grouping principles, grouping by proximity and grouping by similarity perform as lawfully in audition as they do in vision. We can predict the probability that a note is perceived as the starting point of a circular auditory pattern by using the strength of grouping by temporal proximity, grouping by loudness similarity or grouping by pitch similarity. When grouping by temporal proximity and grouping by loudness similarity were conjointly applied to the same stimulus, their effects were additive, as was found in vision. However, when grouping by temporal proximity and grouping by pitch similarity were conjointly applied to the same stimulus, people only relied on pitch similarity for grouping auditory necklaces and ignored temporal proximity.

Acknowledgement

Six year ago, on August 13th, 2007, I put my steps onto this new country the first time and started a long personal journey. The completion of this dissertations is an achievement that marks the end of these six wonderful years at the University of Virginia. There are so many people, to whom I owe large debt of gratitude, and without whom I would not make this achievement.

First and foremost, I would like to thank Michael Kubovy, my graduate advisor. He has supported me both financially and emotionally since my arrival in US. He has been a role model and showing me how to do good science throughout my time at UVa. His passion for serious science, strict methodology and appropriate data analyst has influenced me so much and changed my whole career path.

I would like to acknowledge my committee members (Bill Epstein, Timo von Oertzen, and Judith Shatin), who provided great advice and feedback. Bill Epstein has been helping me since my first year. He provided me numerous comments and advices for my writings, presentations, researches, and even life. I thank Timo von Oertzen for inspiring me and offering insightful suggestions on the methods of my dissertation work. I thank Judith Shatin for bringing a musical perspective to my research, raising questions that made me think more deeply, and offering great suggestions for future projects.

I acknowledge Mowei Shen and Ailun Liu, my graduate and undergraduate advisor at Zhejiang University. They were my first mentors to psychology and led me into this interesting and mysterious world.

I also would like to thank the current and past fellow students in the Kubovy lab for offering their assistance and feedback on my work, especially Laura Getz, Steve Scheid, Michael Schutz and Holly Earls. As a student from another country, their help is essential

to my life in Gilmer. Additional, I thank the wonderful research assistants and DMP students that I have worked with over the years, for their assistance and hard work.

I am thankful to my wonderful friends at UVa, who provide encouragement to my work and bring happiness to my life. Ming Mao, Mingyi Hong, Jingyuan Li, Jian Lu, Yongjin Lu, Lu Wang, and Ruwei Wang, I doubt I would have finished without you.

I dedicate this dissertation to my mother Xiaohong Yu, my father Zhongming Yu, and my wife Ming Lou, who have always believed in me and have always offered endless love and support to me. Last, I also dedicate this dissertation to my upcoming baby daughter Charlotte Zijiang Yu, who brings luck and joyfulness to my whole family.

Charlottesville, VA, USA

May 8th, 2013

Table of Contents

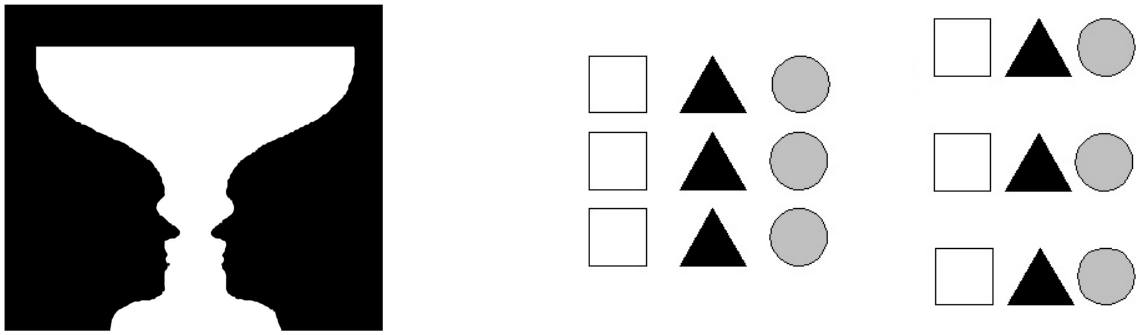
List of Figures

List of Tables

1 Introduction

The striking examples of Gestalt phenomena are illustrated in almost all psychology textbooks. These illustrations demonstrate various Gestalt laws and one of the most influential statements of Gestalt psychology, "the whole is other than the sum of the parts". For example, Figure 1a shows the classic Rubin vase/faces illusion. People can either see a white vase in the middle of the picture or two black faces on the sides of the picture. The picture represents an important subfield of Gestalt psychology — figure-ground segregation. People may use various perceptual cues to segregation figure from the background and our perceptual system is very flexible in doing this task. Figure 1b demonstrates two important grouping principles in Gestalt psychology — grouping by similarity and grouping by proximity. For the picture on the left, one is very likely to group the objects by columns because the objects in each column are the same. However, if we make the distances between rows larger, one may group the objects by columns using similarity cue or one may group the objects by rows using the principle grouping by proximity.

However, studies in Gestalt psychology were often vulnerable to criticism for their



(a) Rubin vase/faces.

(b) Demonstration of grouping principles.

Figure 1: Examples of Gestalt phenomena.

subjectivity and insufficient quantification. To overcome these weaknesses, researchers have been developing quantitative methods to study Gestalt phenomena. The dot lattice is one of the most important tools that have been used to quantify Gestalt phenomena. Researchers have been using this tool to study the classic grouping principles and have found important results.

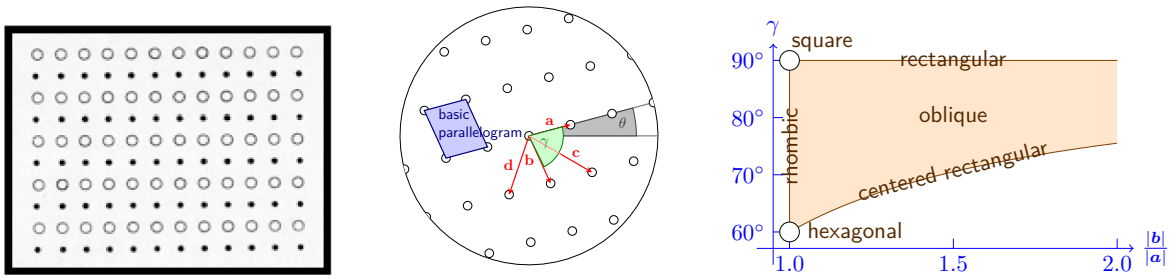
Even though progress has been made in Gestalt quantification, these breakthroughs were by and large based on studies of visual Gestalt phenomena (Kubovy, Holcombe, & Wagemans, 1998; Kubovy & van den Berg, 2008; Peterson & Gibson, 1994; Peterson & Lampignano, 2003). We will not have a proper understanding of the general effects of Gestalt principles until we generalize the findings in vision to other modalities (for a review of steps taken in this direction, see Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012).

In the current dissertation, I developed an analogy between audition and vision, and asked whether findings with visual grouping also apply to auditory perceptual organization. Before I present results from the experiments, I begin with a review of the studies that inspired the current work.

1.1 Quantifying grouping principles in vision

1.1.1 Dot lattices

Dot lattices have been used as a tool to study grouping as early as by Wertheimer (1923) (translation in Ellis, 1938) (see Figure 2a for an example). Kubovy (1994) summarized the taxonomy of dot lattices and formally defined them, which provides a solid foundation for quantification. A dot lattice is defined as a collection of dots in the plane which is invariant under two translations, a vector $\boldsymbol{a}$ (with length $|\boldsymbol{a}|$) and a vector $\boldsymbol{b}$ (whose length



(a) Dot lattice used by Wertheimer (1923).

(b) Defining features of a dot lattice.

(c) Two-dimensional space and nomenclature of dot lattices.

Figure 2: Dot lattices.

is $|\boldsymbol{b}| \geq |\boldsymbol{a}|$). These two vectors, and the angle between the vectors, γ (constrained for purely geometric reasons by $60° \leq \gamma \leq 90°$), define the *basic parallelogram* of the lattice, and thus the lattice itself. The diagonals of the basic parallelogram (shown in Figure 2b) are $\boldsymbol{c}$ and $\boldsymbol{d}$ (where $|\boldsymbol{c}| \geq |\boldsymbol{d}|$). In its *canonical orientation*, $\boldsymbol{a}$ is horizontal; the angle θ (measured counterclockwise) is the measure of the *orientation* of a dot lattice ($\theta = 15°$ in fig. 2b); we call $|\boldsymbol{a}|$ the *scale* of the lattice. If we are not interested in the scale of a lattice, we can locate dot lattices in a two-dimensional space with dimensions $|\boldsymbol{b}|/|\boldsymbol{a}|$ and γ (fig. 2c). In this space we can identify six different types of lattices, which differ in their symmetry properties.

The dot lattices that are used in experiments are multistable and ambiguous. An ambiguous stimulus can produce alternations among two or more different subjective percepts. For example, the dots in Figure 2a may be grouped in either columns or rows depending on the grouping principle that one uses. It is interesting and important to study multistablility because such ambiguity occurs when the perceptual system is on an edge, as one or more Gestalt laws are competing with each other on this single stimulus to "win" the percepts. By observing the changes of percepts while we delicately change the stimulus property, we could learn a lot about how our perceptual system works.

1.1.2 Quantifying grouping by proximity

The lack of quantification in Gestalt psychology was partially due to its reliance on phenomenological demonstrations. People often equate phenomenology with subjectivism. Marr (1982) said that Gestalt psychology "dissolved into the fog of subjectivism". In response to this sort of criticism, researchers have developed paradigms to make phenomenological demonstrations measurable and reveal lawful mechanisms.

Oyama (1961) showed rectangular dots lattices at a fixed orientation to the participants, and asked them to report whether they see the vertical or the horizontal groupings. He found that the ratio of the time participants saw the vertical and the horizontal organizations was a power function of the ratio of the vertical and horizontal distances.

Kubovy and Wagemans (1995) and Kubovy et al. (1998) developed a paradigm in which they demonstrated that we can understand grouping by proximity as the outcome of a probabilistic competition among potential perceptual organizations. In their experiments, dot lattices at near-equilibrium were presented to the participants very briefly (in hundreds of milliseconds). The participants were asked to report the perceived organization in each current trial by choosing one of the four directions (a, b, c, or d) (see Figure 3 for a typical trial in their paradigm). For each dot lattice, the researchers were able to calculate the probability each of the four directions was perceived ($p(a), p(b), p(c)$ and $p(d)$). By systematically manipulating the aspect ratio ($|\boldsymbol{b}|/|\boldsymbol{a}|$) and angle γ, they showed that all the values of $\log[p(v)/p(a)]$ (where response $v \in \{b, c, d\}$) fall on the same line, which they call the *attraction function* (Figure 4). The slope of the attraction function, ξ, is a person-dependent measure of sensitivity to proximity.

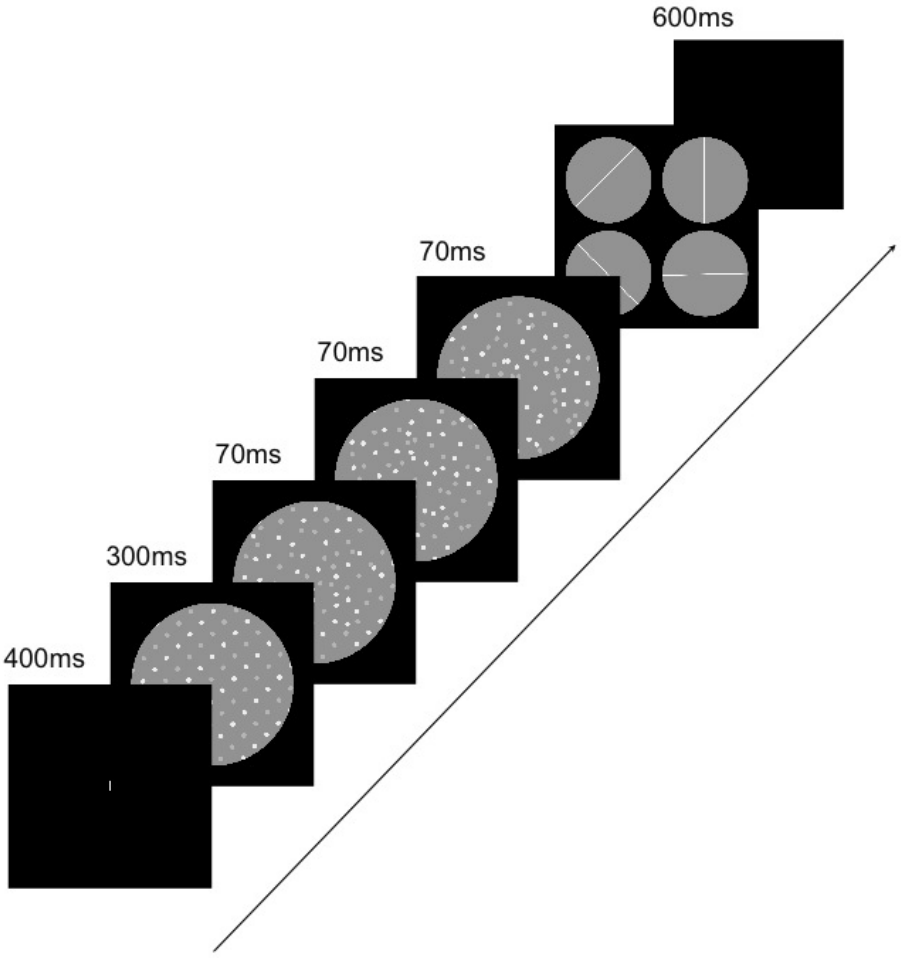


Figure 3: A typical trial in the experiments of Kubovy and van den Berg (2008).

1.1.3 Quantifying grouping by proximity and grouping by similarity

As Figure 2a shows, perceptual organizations are usually formed not by one single Gestalt principle, but by multiple Gestalt principles. An important question to answer is how those Gestalt principles work together when they are applied to the same stimulus.

To address this question, Oyama, Simizu, and Tozawa (1999) presented rectangular dimotif lattices to the participants and asked them to report whether they saw horizontal or vertical grouping by tilting a joystick to the right or left. A double-staircase procedure

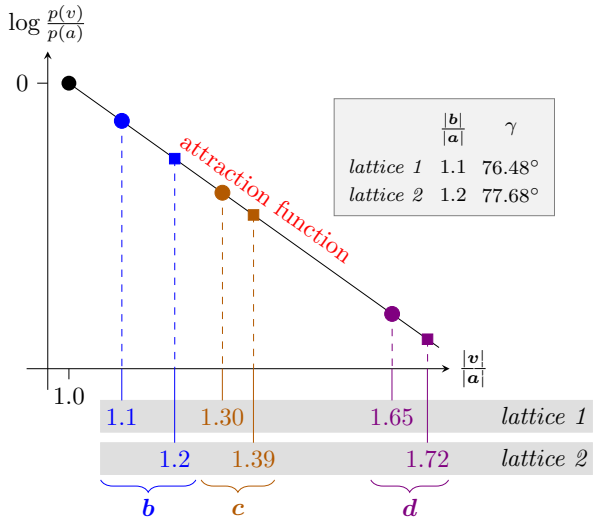
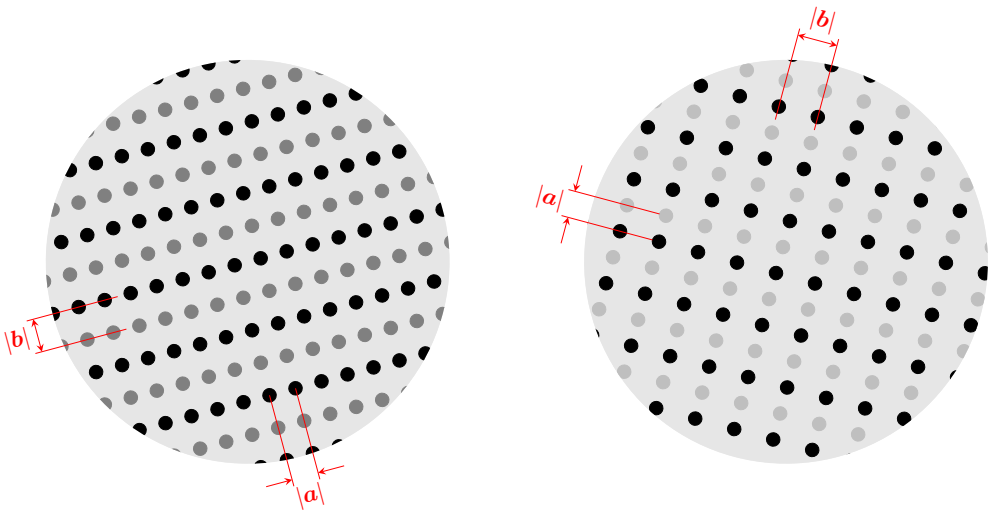


Figure 4: The attraction function of grouping by proximity. Consider two dot lattices (in which we assume that $|\boldsymbol{a}| = 1$): in the first, $|\boldsymbol{b}| = 1.1$ and $\gamma = 76.48°$; in the second $|\boldsymbol{b}| = 1.2$ and $\gamma = 77.68°$. The corresponding lengths of $\boldsymbol{c}$ are $|\boldsymbol{c}| = 1.3$ and 1.39, and the lengths of $\boldsymbol{d}$ are $|\boldsymbol{d}| = 1.65$ and 1.72.

was used to determine the ratio of vertical distance to horizontal distance $|v|/|h|$. The ratio increased after a vertical response, whereas it decreased after a horizontal response. They obtained results showing when the distance ratio was in equilibrium with different types of dissimilarity including luminance, size, color, and additional features. Although this study took a step forward to find the conjoint effect of multiple grouping principles, it did not answer whether the grouping principles work independently (additively) or not (non-additively).

To further address this question, Kubovy and van den Berg (2008) examined the conjoined effect of the two classic grouping principles — grouping by proximity and grouping by similarity. Using a paradigm similar to their previous experiments (see Figure 3) and dimotif dot lattices as stimuli (Figure 5), they obtained data to build probabilistic models and to plot a family of attraction functions to determine the relationship among these functions (parallel, suggesting additivity and independence; or unparallel, suggest-



$|\boldsymbol{b}| = 1.25|\boldsymbol{a}|$, $\delta = 2$ $|\boldsymbol{b}| = 1.5|\boldsymbol{a}|$, $\delta = -3$

Figure 5: Two dimotif dot lattices. In both, grouping by proximity favors a, but more weakly in the dot lattice on left, where $|\boldsymbol{a}|$ =1.25$|\boldsymbol{b}|$, than in the dot lattice on the right, where $|\boldsymbol{a}|$ =1.5$|\boldsymbol{b}|$. In the dot lattice on the left, grouping by similarity favors b ($\sigma > 0$, where σ is a measure of dissimilarity between two kinds of dots), whereas in the dot lattice on the right it favors a ($\sigma < 0$).

ing nonadditivity and dependence). Their results (shown schematically in Figure 6) demonstrated that grouping by proximity and grouping by similarity affect the outcome independently. The effects of these two grouping principles are additive, suggesting that "the whole is equal to the sum of the parts", which seems to be inconsistent with the traditional view of whole-parts relationships in Gestalt psychology.

1.2 An auditory Gestalt phenomena — auditory necklaces

1.2.1 Auditory necklaces

Our understanding of perceptual organizations is mostly based on studies of visual ambiguity like the dot lattices discussed above. Yet there is little doubt that Gestalt grouping principles apply to other modalities. Studies of other modalities are necessary for us to

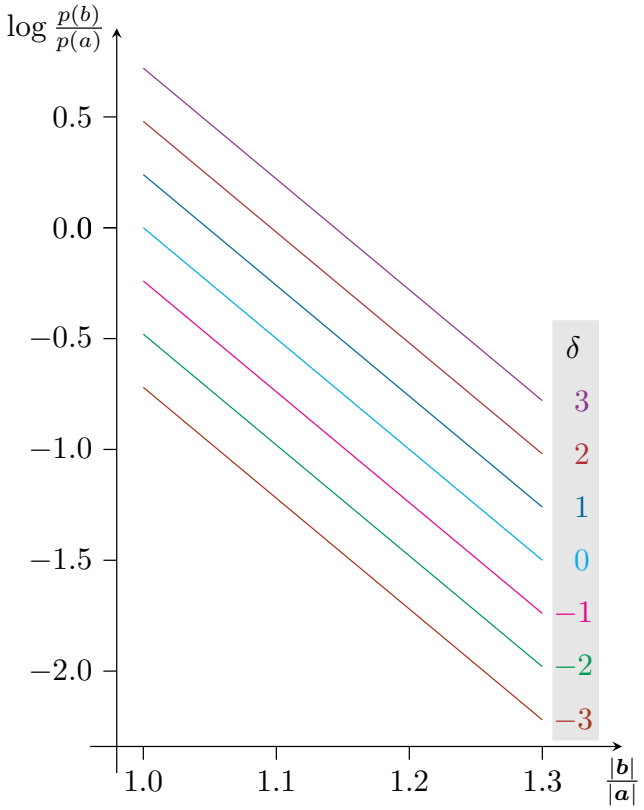


Figure 6: A schematic of the results obtained by Kubovy and van den Berg (2008). The attract functions are parallel, showing that the conjoined effects of proximity and similarity are additive.

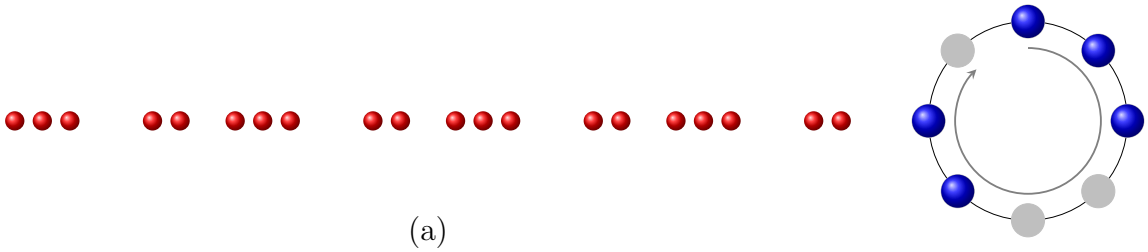
get a complete understanding of perceptual organizations. We have developed a quantifiable auditory Gestalt phenomenon — auditory necklaces — to fill in this gap in our understanding.

Imagine a repeating eight-beat auditory pattern (where ♪ represents a note and 𝄾 represents a rest):

... ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ....

It is ambiguous, because you can hear ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 as a unit and parse the pattern as:

... ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ♪ ♪ ♪ 𝄾 𝄾 ♪ ♪ 𝄾 ...,



(a) (b)

Figure 7: Grouping in vision and audition. (a) Visual grouping in space; (b) Auditory grouping in time. Each colored ball represents a note; each grey ball represents a rest. The auditory pattern plays clockwise and circularly.

or hear 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝄾 as a unit and parse the pattern as:

... 𝅘𝅥𝅯 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 𝄾 𝅘𝅥𝅯 𝅘𝅥𝅯 𝄾 ....

The segmentation of such an auditory sequence is an important auditory grouping problem. We group those notes in time in audition, whereas we usually group objects in space in vision (Figure 7). The perceptual organization of auditory patterns is essential in our daily life as it affects both the processing of speech and music (Deutsch, 1980; Longuet-Higgins & Lee, 1982; Martin, 1972).

We borrow the concept of necklace from combinatorics (Ruskey, 2011) and call those auditory repeating patterns auditory necklaces because they are best visualized when arranged on a circle. The pattern mentioned above is a binary (notes and rests) auditory necklace (in short AN) of length 8, which we code as `11100110` (where `1`s are notes and `0`s are rests). Visually, we use colored beads for notes and grey beads for rests (Figure 8).

The perceived starting beat of an AN is called its *clasp*. For each AN, people could technically perceive any beat as the clasp, but as we will see later, there are principles predicting only few notes to be the most likely clasps. For example, if one perceives `11100110` repeating itself, the underlined `1` is the clasp. A *block* is a sequence of consecutive identical events (be they notes or rests). A block of notes is called a *run* and a

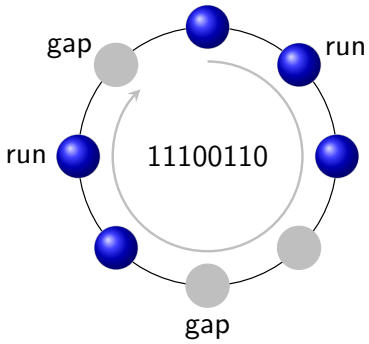


Figure 8: An auditory necklace `11100110` of length $n = 8$.

block of rests is a *gap* (Figure 8). `11100110` is a 4 block AN with two runs (`11` and `111`) and two gaps (`00` and `0`).

In the seminal work of Garner and his colleagues (Preusser, Garner, & Gottwald, 1970; Royer & Garner, 1966, 1970), they formulated two organization principles for the segmentation of ANs. The first principle is *the run principle*, which predicts that the first note of the longest run would be perceived as the beginning of a pattern (i.e. the clasp of the AN). The other principle is *the gap principle*, which predicts that the first note following the longest gap would be the clasp. For example, the run principle would predict an organization of `11100110` while the gap principle would predict an organization of `11011100`.

1.2.2 A new paradigm to study auditory necklaces

Two paradigms were used in early studies on auditory necklaces. Garner and his colleagues (Preusser et al., 1970; Royer & Garner, 1966, 1970) asked participants to report the perceived organization of ANs by pressing keys or writing down the patterns. Although these procedures recorded the participants' percepts faithfully, each trial took too long for the researchers to collect enough data for quantitative modeling.

Boker and Kubovy (1998) asked participants to strike a key on a synthesizer keyboard

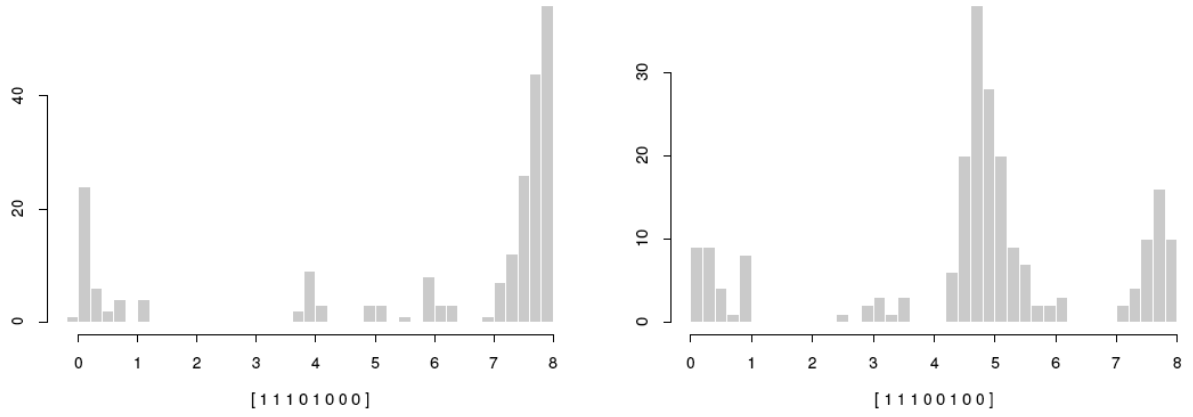


Figure 9: Response distributions for two stimulus patterns in Boker and Kubovy (1998).

at the moment they heard the clasp. This allowed them to collect a large amount of data. However, this method has two drawbacks: (a) It was an extremely hard task for participants to synchronize their responses with the tones. Figure 9 shows the response distributions for two patterns in their experiment. Most responses preceded or followed the beats that the participants perceive as the clasps. The researchers had to set an arbitrary criterion to decide which beat a response aimed for, which introduced noise into the data and complicated the analysis. (b) The task did not record pure perception. Motor control was confounded into the process.

We (Yu & Kubovy, submitted) devised a new method that (1) allows participants easily and quickly to report the clasp, thus allowing us to obtain enough data to build quantitative models; (2) Unlike some previous experiments (Boker & Kubovy, 1998), which require participants to synchronize their taps with the beat, data collected by the new method reflect perception alone.

At the beginning of each trial, a circular array of n icons (where n = the length of the AN) appeared on the screen. The computer randomly assigned icons to positions around the circle, and randomly associated the top icon with one of the beats (a tone or a rest)

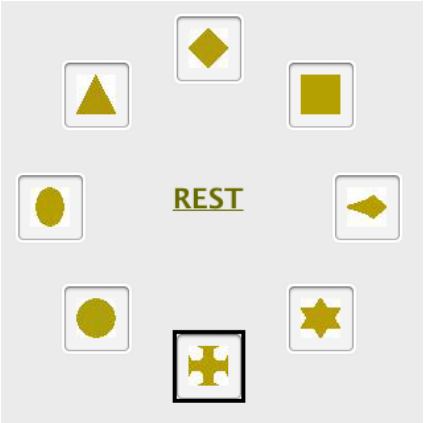


Figure 10: Screenshot of the display. At the moment depicted the cross is highlighted.

of the AN (Figure 10). While the AN was played (over headphones), a square highlighted the corresponding icon and moved clockwise. The participants were instructed to click at any time on the icon corresponding to the clasp.

1.2.3 Quantifying run principle and gap principle

Using the new paradigm, we attempted to quantify the run and gap principles proposed by Garner and his colleagues (Preusser et al., 1970; Royer & Garner, 1966, 1970) by examining the relationship between those two principles (Yu & Kubovy, submitted). In the experiment, we used a sample of ANs with two runs and two gaps, which we called the runs A and B (denoted r_{A} and rb, the gap preceding A denoted g_{A}, the gap preceding B denoted g_{B}). The lengths of the runs and the gaps were manipulated.

First of all, participants made more than 95% of their responses to the first note of a run. We treated other responses as errors so that the response variable became binomial—choosing r_{A} or r_{B}. To measure the strengths the run and gap principles, we calculated the *log run-length ratio*: $R_{\mathrm{run}} = \log(\text{length of } r_{\mathrm{A}}/\text{length of } r_{\mathrm{B}})$ and the *log gap-length ratio*: $R_{\mathrm{gap}} = \log(\text{length of } g_{\mathrm{A}}/\text{length of } g_{\mathrm{B}})$.

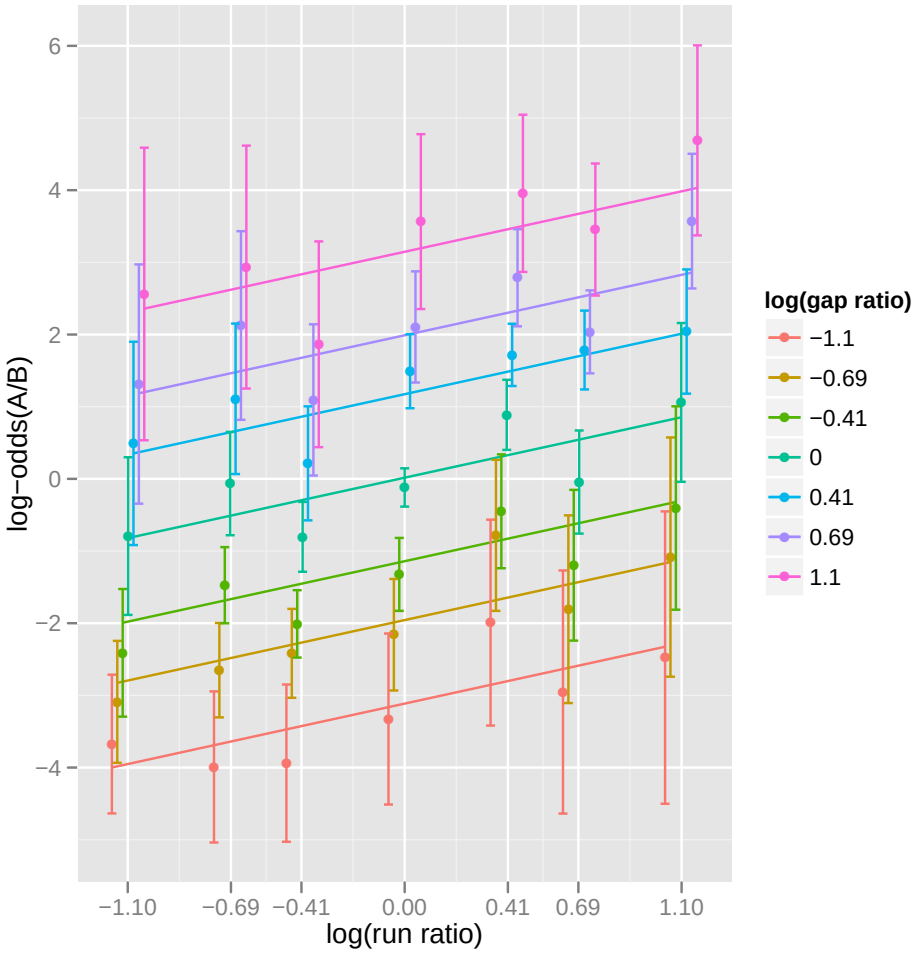


Figure 11: Log-odds(A/B) as a function R_{run} and R_{gap}. Error bars span ± 1 SE. The attract functions are parallel, showing that the conjoined effects of run and gap are additive.

The best generalized linear mixed model we fitted to the data was additive with R_{run} and R_{gap} as predictors. The results are shown in Figure 11. Each line represents an attraction function of run principle at one level of R_{gap}. The parallel nature of the lines demonstrates the additivity of the two principles.

1.3 Additivity and non-additivity

1.3.1 Additivity in figure-ground perception

The surprising additive conjoint effects of grouping principles has been found in the grouping of both visual dot lattices and auditory necklaces. Other than those two Gestalt phenomena, we also quantitatively studied another visual Gestalt phenomena — figure-ground segregation (Yu & Kubovy, 2012). We examined two classic cues in the perceptual organization of figure-ground segregation: relative area and convexity. To quantify them, we designed a set of visual stimuli as shown in Figure 12. The stimuli we designed are circles with red and green strips. The strips of one color are convex whereas the strips of the other color are concave. We manipulated the strength of relative area cue by altering the width ratio of green and red strips (w_{red}/w_{green}). We manipulated the strength of convexity cue by altering the height of the bumps (Δ_{bump}). We used a similar paradigm to the dot lattices experiments. Instead of asking the participants to report the perceived directions of dot lattices, we asked the participants to report the perceived foreground (red or green) of the visual pattern. Similar to the additivity between grouping by proximity and grouping by similarity in dot lattices, we also found that the conjoint effects of convexity and relative area in figure-ground perception are additive.

1.3.2 Non-additivity in curved dot lattices

Although several studies across sensory modalities have shown similar additivity among grouping principles, non-additive conjoint effects have also been found. Strother and Kubovy (2012) designed a special set of curved dot lattices (see Figure 13) to study the conjoint effects of aspect ratio, curvature level and density on perceptual organization. Their results showed that for lattices with sparse dots, only aspect ratio had an effect on

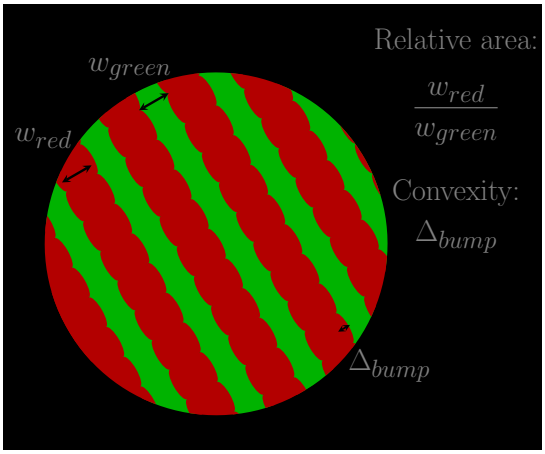


Figure 12: Figure-ground stimuli. The width ratio of two colored strips and the height of the bumps are manipulated in the experiments.

perceptual organization. For lattices with dense dots, both aspect ratio and curvature level had effects on perceptual organization. The effect of curvature level increases as aspect ratio increases, indicating the effects of those two grouping principles are not additive. They contemplated that a new emergent property may rise from the 3-D sphere-like perception of the curved dot lattices, which may have led to the non-additivity among those grouping cues.

1.4 Current work

The accumulated research on the quantification of dot lattices has enabled us to go beyond only a qualitative description of Gestalt grouping principles. The recent results (Kubovy & van den Berg, 2008) even suggested surprising additive conjoint effects between grouping by proximity and grouping by similarity. Such additivity has also been found in other Gestalt phenomena. However, we still have very limited knowledge about how similar grouping principles work in audition.

Although we successfully quantified the effects of run and gap principle in the grouping

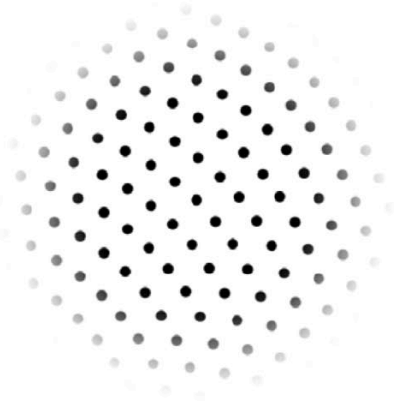


Figure 13: Curved dot lattice in Strother and Kubovy (2012).

of auditory necklaces, the difference in the characteristics of the two auditory principles make it difficult to directly compare them to the two classic and well-studied grouping principle in vision — grouping by proximity and grouping by similarity. For the run principle, there is no well-matching visual grouping principle to it. For the gap principle, it is in some sense a principle of grouping by temporal proximity. But in previous experiments (Yu & Kubovy, submitted), we only manipulated the number of beats, which had to be an integer and had a very small range (less than 5). This limited our ability to obtain more accurate results to compare it to grouping by spatial proximity in vision. Despite of the limitation of direct comparison, the establishment of auditory necklaces has provided us a useful tool to quantify auditory grouping.

In the current dissertation, I designed auditory necklaces which may be grouped by three auditory grouping principles — grouping by temporal proximity, grouping by loudness similarity, and grouping by pitch similarity. Those three auditory principles are directly analogous to the grouping by spatial proximity and grouping by similarity principles in vision. In three experiments, I examined the separate effect of grouping by temporal proximity, the conjoint effects of grouping by temporal proximity and grouping

by loudness similarity, and the conjoint effects of grouping by temporal proximity and grouping by pitch similarity.

2 Simulation studies

We used adaptive staircases as the sampling plan in the three experiments in this dissertation. The results of a pilot experiment showed that: (1) We could only afford about 100 trials per condition (per staircase). (2) The individual differences in both threshold and slope were large. Therefore, before we started collecting the data, we first conducted two simulation studies to find the optimal staircase procedure for our experiments. The results of those simulation studies can be applied not only to the experiments in this dissertation, but also to other psychophysics experiments with limited resources as well.

2.1 Adaptive staircases

As research on psychophysics grows, adaptive procedures have been developed to increase the efficiency of measurement. In the experiments using adaptive procedures, the physical characteristics of the stimuli on each trial are determined by the stimuli and responses that occurred in the previous trial or sequence of trials. Many forms of modern adaptive methods have been developed to maximize efficiency and to minimize participant and experimenter time, while preserving accuracy of the measurement.

Adaptive staircase is a series of adaptive procedures which are simple and easy to use in psychophysics experiments. The original up-down staircase method (1-1 rule) targets a point at which the probability of success is 50%. There is a fixed step size (Δ). The stimulus level would decrease Δ after a successful response and would increase Δ after an unsuccessful response. Several modified and improved staircase procedures were developed later including *adaptive staircase with up-down transformed rules (UDTRS)* and *adaptive staircase with up-down weighted rules (UDWRS)*.

In UDTRS, instead of the original 1-1 rule, we can set a different value for the

number of consecutive successful (or unsuccessful) responses that are required at the current stimulus level to bring it down (or up) by one step for the next trial. For example, a 1-3 (1-up 3-down) rule means that the stimulus level would increase Δ after 1 unsuccessful response and would decrease Δ after 3 successful responses. The procedures with different combinations of the numbers target different probability points in the psychometric function.

In UDWRS, although we use the original 1-1 rule, the size $\Delta+$ of a step up is an integer multiple (k) of the size Δ- of a step down ($\Delta+ = k \times \Delta$-). It is called as a k UDWR staircases. For example, in a 2 UDWR staircase, the stimulus level would decrease Δ- after 1 successful response and would increase 2Δ- ($\Delta+$) after 1 unsuccessful responses. Again, procedures with different ks target different probability points in the psychometric function.

Using a similar simulation method as in García-Pérez and Alcalá-Quintana (2005), we examined the efficiency and accuracy of several sampling plans. Because we found large individual differences in our pilot experiment, a single staircase procedure may not fit all participants well. In addition to using a single sampling plan for all 200 trials, we also simulated conditions using a combination of two sampling plans. We tested 6 sampling plans in total: 1-2 UDTRS, 1-3 UDTRS, half 1-2 half 1-3 UDTRS, k=2 UDWRS, k =3 UDWRS, and half k=2 half k=3 UDWRS. For each sampling plan, we set two step size levels.

2.2 Simulation 1

2.2.1 Method

In Simulation 1, we simulated the experimental settings of Experiment 1: the threshold of the psychometric function was known, we needed to estimate slope b and guessing/lapse rate γ. Responses from 4 virtual participants were simulated. We assumed that all of those 4 virtual participants' psychometric functions were logistic as defined in Equation (1).

$$\Psi(x) = \gamma + \frac{1 - \gamma \times 2}{1 + \exp[-b(x - \theta)]} \tag{1}$$

The parameters of those participants' psychometric functions are shown in Table 1. In each simulated staircase run, the virtual participants completed 100 trials. The stimulus level started from 1.5 in down direction. The minimum stimulus level is 1 and the maximum stimulus level is 2. For all sampling plans, we examined two levels of step size down (Δ-), one was 0.025 and the other was 0.05. The levels of step size up (Δ+) were calculated based on Δ- and staircase procedure of each sampling plan. We tested 12 sampling plans in total, and simulated 1000 runs for each sampling plan. The simulated responses were generated by a custom program written in Python.

Table 1: The parameters of 4 virtual participants' psychometric functions

	P1	P2	P3	P4
θ	1	1	1	1
b	5	5	10	10
γ	0.1	0.2	0.1	0.2

Table 2: Root mean squared errors (RMSEs) of estimated bs for each pair of method and virtual participant in Simulation 1. The lowest 3 errors of each column are in bold font. The lowest errors of each column are underlined.

Method	P1 $b = 5$ $\gamma = .1$	P2 $b = 5$ $\gamma = .2$	P3 $b = 10$ $\gamma = .1$	P4 $b = 10$ $\gamma = .2$
1-2, Δ-=0.025	5.50	4.18	6.14	**6.90**
1-2, Δ-=0.05	5.57	5.37	5.97	7.41
1-3, Δ-=0.025	**4.16**	**3.64**	**5.42**	7.25
1-3, Δ-=0.05	**3.80**	**3.87**	**5.20**	**6.95**
k=2, Δ-=0.025	5.65	6.71	**5.90**	7.43
k=2, Δ-=0.05	6.30	6.99	6.92	8.24
k=3, Δ-=0.025	**3.85**	**4.14**	6.63	**6.50**
k=3, Δ-=0.05	5.65	6.60	6.63	8.21
1-2/1-3, Δ-=0.025	5.36	6.03	6.72	8.05
1-2/1-3, Δ-=0.05	5.60	6.10	6.75	8.14
k=2/k=3, Δ-=0.025	6.27	6.54	6.51	8.26
k=2/k=3, Δ-=0.05	6.16	7.20	6.87	8.78

2.2.2 Results

Histograms of estimated bs and γs from those runs for each pair of 4 virtual participants and 12 sampling plans are listed in Appendix A. For each pair of virtual participant and method, we calculated the root mean squared error (RMSE) of the 1000 estimated bs and γs. RMSE is a widely used for measuring the magnitude of the estimation error. The RMSE of our simulations are defined in Equation (2) and Equation (3). Table 2 and Table 3 list RMSEs of Simulation 1.

$$RMSE_b = \sqrt{(b_{\text{estimated}} - b_{\text{actual}})^2} \tag{2}$$

$$RMSE_\gamma = \sqrt{(\gamma_{\text{estimated}} - \gamma_{\text{actual}})^2} \tag{3}$$

Table 3: Root mean squared errors (RMSEs) of estimated γs for each pair of method and virtual participant in Simulation 1. The lowest 3 errors of each column are in bold font. The lowest errors of each column are underlined.

Method	P1 $b = 5$ $\gamma = .1$	P2 $b = 5$ $\gamma = .2$	P3 $b = 10$ $\gamma = .1$	P4 $b = 10$ $\gamma = .2$
1-2, Δ-=0.025	0.093	0.183	0.077	0.160
1-2, Δ-=0.05	0.096	0.178	0.084	0.152
1-3, Δ-=0.025	0.095	0.186	0.082	0.185
1-3, Δ-=0.05	0.093	0.174	0.084	0.171
k=2, Δ-=0.025	0.095	0.172	0.078	0.139
k=2, Δ-=0.05	0.098	0.167	0.083	0.142
k=3, Δ-=0.025	**0.067**	**0.129**	**0.052**	**0.125**
k=3, Δ-=0.05	**0.089**	**0.148**	0.076	**0.131**
1-2/1-3, Δ-=0.025	0.093	0.171	**0.072**	0.153
1-2/1-3, Δ-=0.05	0.093	0.163	0.076	0.144
k=2/k=3, Δ-=0.025	**0.090**	0.164	**0.070**	0.132
k=2/k=3, Δ-=0.05	**0.090**	**0.152**	0.074	**0.131**

Similar to what was found in previous simulation studies (García-Pérez & Alcalá-Quintana, 2005), bs were usually underestimated in all methods and γs were strongly underestimated (in most runs γs were estimated to be 0) in all methods. In a 2-alternative-force-choice (2AFC) discrimination experiment like the settings of the current simulation, 1-3 UDTRS estimated bs better than other methods, whereas k=2 UDWRS estimated γs better than other methods. The combination plans did not estimate bs more accurately, but they seemed to estimate γs better. More importantly, no sampling plan could estimate both b and γ accurately at the same time.

Since guessing rate was poorly estimated in all methods and we were more interested in the slope rather than guessing rate in our experiments, so therefore we used 1-3 UDTRS in Experiment 1. The step size did not strongly affect the estimation, we used 1-3 UDTRS with $\Delta = 0.05$ in our Experiment 1 as it performed better for P1, P3 and

P4 and a pilot experiment showed that the participants' bs are around 10.

2.3 Simulation 2

2.3.1 Method

In Simulation 2, we simulated the experimental settings of Experiment 2 and 3: we needed to estimate both threshold θ and slope b. We already had the estimation of guessing rate γ. Responses from 6 virtual participants were simulated. As in Simulation 1, we assumed that all of those 6 virtual participants' psychometric functions were logistic as defined in Equation (1). The parameters of those participants' psychometric functions are shown in Table 4. In each simulated staircase run, the virtual participants completed 100 trials. The stimulus level started from 20 in down direction. The minimum stimulus level is 0 and the maximum stimulus level is 40. As in Simulation 1, we examined two levels of step size down (Δ-) for each sampling plan, one was 1 and the other was 2. The levels of step size up (Δ+) were calculated based on Δ- and staircase procedure of each sampling plan. We tested the same 12 sampling plans as in Simulation 1. We simulated 1000 runs for each sampling plan. The simulated responses were generated by a custom program written in Python.

Table 4: The parameters of 6 virtual participants' psychometric functions

	P1	P2	P3	P4	P5	P6
θ	5	5	5	10	10	10
b	0.25	0.5	1	0.25	0.5	1
γ	0.1	0.1	0.1	0.1	0.1	0.1

Table 5: Root mean squared errors (RMSEs) of estimated θs for each pair of method and virtual participant in Simulation 2. The lowest 3 errors of each column are in bold font. The lowest errors of each column are underlined.

Method	P1 θ=5 b=0.25	P2 θ=5 b=0.5	P3 θ=5 b=1	P4 θ=10 b=0.25	P5 θ=10 b=0.5	P6 θ=10 b=1
1-2, Δ-=1	2.93	1.40	0.74	2.87	1.32	0.74
1-2, Δ-=2	2.27	**1.05**	**0.52**	2.16	1.07	**0.49**
1-3, Δ-=1	5.12	2.48	1.70	4.34	2.39	1.56
1-3, Δ-=2	3.60	1.60	0.93	3.52	1.69	1.00
k=2, Δ-=1	**1.96**	**0.97**	**0.48**	**1.97**	**0.88**	**0.48**
k=2, Δ-=2	**1.72**	**0.85**	**0.52**	**1.75**	**0.83**	**0.49**
k=3, Δ-=1	2.78	1.32	0.67	2.79	1.39	0.72
k=3, Δ-=2	2.25	1.15	0.71	2.29	1.20	0.67
1-2/1-3, Δ-=1	3.92	2.22	1.63	3.67	2.10	1.20
1-2/1-3, Δ-=2	2.33	1.18	0.62	2.29	**0.98**	0.52
k=2/k=3, Δ-=1	2.60	1.33	0.67	2.54	1.22	0.65
k=2/k=3, Δ-=2	**1.98**	1.08	0.62	**2.08**	1.01	0.54

2.3.2 Results

Histograms of estimated θs and bs were listed in Appendix A. We again calculated RMSEs to measure the accuracy of the sampling plans. RMSEs are defined in Equation (2) and Equation (4). Table 2 and Table 3 lists RMSEs of Simulation 1. Table 5 and Table 6 show RMSEs of the two estimated parameters for each pair of virtual participant and method.

$$RMSE_\theta = \sqrt{(\theta_{\text{estimated}} - \theta_{\text{actual}})^2} \tag{4}$$

For both θ estimation and b estimation, three procedures performed better than other methods in most conditions: $k = 2$ UDWRS Δ-=2, $k = 2$ UDWRS Δ-=1 and 1-2 UDTRS Δ-=2. Among those three procedures, $k = 2$ UDWRS Δ-=2 estimated both parameters

Table 6: Root mean squared errors (RMSEs) of estimated bs for each pair of method and virtual participant in Simulation 2. The lowest 3 errors of each column are in bold font. The lowest errors of each column are underlined.

Method	P1 θ=5 b=0.25	P2 θ=5 b=0.5	P3 θ=5 b=1	P4 θ=10 b=0.25	P5 θ=10 b=0.5	P6 θ=10 b=1
1-2, Δ-=1	0.53	0.61	0.71	0.40	0.48	0.63
1-2, Δ-=2	**0.25**	**0.35**	**0.59**	**0.21**	0.34	**0.55**
1-3, Δ-=1	0.96	1.12	1.05	0.75	0.85	0.88
1-3, Δ-=2	0.54	0.65	0.81	0.44	0.51	0.72
k=2, Δ-=1	0.28	**0.38**	**0.55**	**0.22**	**0.31**	**0.51**
k=2, Δ-=2	**0.17**	**0.30**	**0.60**	**0.17**	**0.31**	**0.59**
k=3, Δ-=1	0.43	0.53	0.73	0.35	0.49	0.73
k=3, Δ-=2	0.33	0.49	0.78	0.32	0.44	0.82
1-2/1-3, Δ-=1	0.96	1.12	1.11	0.57	0.75	0.86
1-2/1-3, Δ-=2	0.34	0.45	0.67	0.25	**0.34**	0.62
k=2/k=3, Δ-=1	0.45	0.58	0.80	0.35	0.46	0.62
k=2/k=3, Δ-=2	**0.26**	0.44	0.68	0.23	0.36	0.73

best when actual bs were small, whereas $k = 2$ UDWRS Δ-=1 performed best when actual bs were large. The combination procedures did not provide better estimations than other procedures.

In Experiment 2 and Experiment 3, we chose $k = 2$ UDWRS Δ-=2 as our adaptive procedure. The actual step size in those two experiments were adjusted to corresponding values calculated using pilot experiment results.

2.4 Discussion

The results of the two simulation studies showed that the best adaptive procedure for the two experimental settings were different. With limited resources, a 1-3 UDTRS procedure is best when we only need to estimate the slope (b) of the psychometric function whereas a $k = 2$ UDWRS procedure is best when we need to estimate both the threshold (θ) and

the slope (b) of the psychometric function. Although the current simulation studies were designed specifically to the current experiment, the results can be generalized to similar psychophysics experiments.

In the current two simulation studies, we attempted to test whether a combination of two procedures would yield more accurate estimations. None of those combination procedures performed best. In those combination procedures, we simply used two staircases with different procedures for half of the trials. It usually took several trials (usually 5-10 trials) of each staircase for the stimulus levels to converge to the sensitive range, and those those trials were not very useful for accurate estimation. Therefore, the combination procedures wasted 5-10 more trials than simple procedures. This may be one of the reasons why the combination procedures did not perform well. In future simulation studies, we can try to design combination procedures that directly connect one to another without wasting trials during converging phases. Those procedures may perform better than simple ones.

3 Experiments

3.1 Experiment 1: Grouping by temporal proximity

In Experiment 1, we explored the effect of grouping by temporal proximity in the segmentation of auditory necklaces. Although we have previously studied the effect of the gap principle (Yu & Kubovy, submitted), which resembles an auditory grouping by temporal proximity principle, that quantification was rough. We used metric ANs in previous experiments and gap lengths could only be integer. Therefore, our manipulation of relative gap strengths was very limited. In the current experiment, we used non-metric ANs so that we could continuously manipulate the temporal distances among notes.

3.1.1 Method

3.1.1.1 Participants Eleven undergraduate students from the University of Virginia participated Experiment 1. They received introductory course credits for their participation. All of them reported normal or corrected-to-normal vision and normal hearing.

3.1.1.2 Stimuli Non-metric ANs with four notes (note 1 — note 4) were used. The stimulus onset asynchrony (SOA) among the four notes were SOA_a (in short a), SOA_b (in short b), SOA_a and SOA_b, correspondingly (see Figure 14). a was fixed to 200ms through the experiment. b was manipulated and the ratio of the two SOAs (b/a) was used to quantify grouping by temporal proximity. The range of b/a is Experiment 1 was from 1 to 2 with increments of 0.05. In Experiment 1, all notes in an AN used the same pitch and loudness. We used either a 440Hz or a 622Hz sine wave pure tone. The duration of each note was 50 ms with 5ms fade-in and 5ms fade-out.

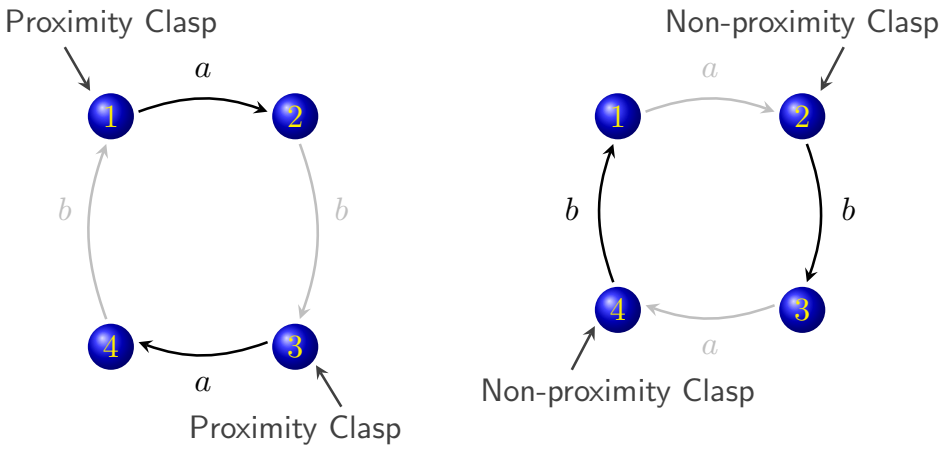


Figure 14: Non-metric ANs used in Experiment 1. Each ball represents a note.

3.1.1.3 Design Based on the results of Simulation 1, we used 1-3 UDTRS adaptive staircase procedure in Experiment 1. b/a would decrease a step (0.05) after 3 responses to proximity clasps (note 1 or 3) and would increase a step (0.05) after 1 response to a non-proximity clasp (note 2 or 4).

Each participant completed two randomly interleaved staircases. Each staircase contained 100 trials. One staircase used ANs with 440Hz tones and the other staircase used ANs with 622Hz tones. b/a started from 1.5 for both staircases.

3.1.1.4 Procedure In Experiment 1, we used a paradigm similar to the one we used in our previous AN experiments (Yu & Kubovy, submitted). On each trial of the experiment, we presented an AN to the participant over the headphone. The AN started at a random note and the playing speed decelerated during the first 4 cycles. As soon as each AN began to play, the screen showed 4 small grey squares arranged in a circle. The computer randomly associated the first note of the AN with one of the squares and the following notes were associated with other squares in clockwise order. After the tempo became steady, two red squares showed up to highlight the square corresponding to the

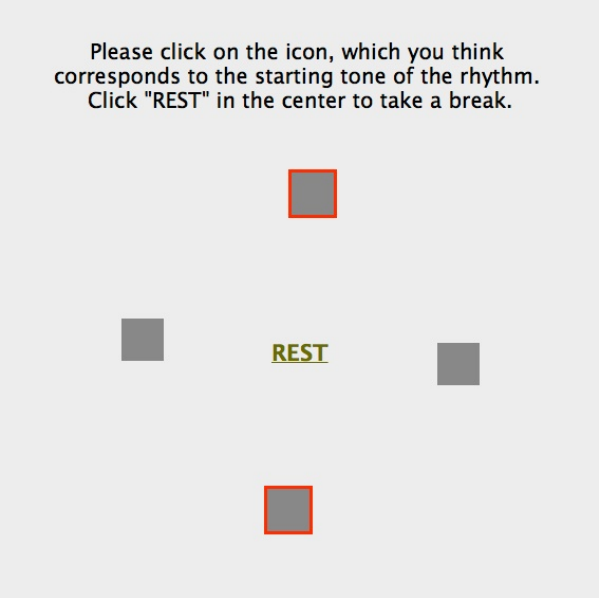


Figure 15: Screenshot of Experiment 1 display. At the moment the top and bottom squares were highlighted.

currently playing note and the square opposite to it (see Figure 19). We highlighted two squares because although we showed 4 squares on the screen the participants were listening to repetitive 2-note auditory patterns.

The participants were instructed to click on either of the two squares corresponding to the note they heard as the beginning of the pattern (the clasp) at any time. The experiment used an adaptive staircase procedure described below. There was no scheduled rest during the experiments, but the participants could click the "REST" button in the middle of the circular array to take a break anytime and they were encouraged to do so.

3.1.2 Results and discussion

Using maximum likelihood (ML) estimation, we fitted logistic psychometric functions to each participant's responses. The estimation was conducted using a custom program written in R. Because Experiment 1 used a 2-alternative force choice (2AFC) paradigm,

the logistic psychometric functions is defined as:

$$\Psi(x) = \gamma + \frac{1 - \gamma \times 2}{1 + \exp[-b(x - \theta)]} \tag{5}$$

in which γ is the guessing/lapse rate, b is the slope and θ is the 50% threshold.

We transformed b/a to be $\log(b/a)$ for all the models because it is in a ratio scale. The 50% threshold (θ) of b/a was 1 due to the nature of the stimuli. When b/a was 1, all SOAs between adjacent notes were equal. There was no physical information for the participants to deliberately select any of the notes to be the clasp. We also simply set γ to be 0.1 because the simulation studies showed that γs cannot be well estimated. We fitted two psychometric functions for each participant, one for the 440Hz staircase and the other for the 662Hz staircase. Table 7 lists the estimated bs for those psychometric functions. Figure 16 shows the two fitted psychometric functions with 95% confidence interval bands across all participants.

Table 7: Estimated bs for 440Hz and 662Hz staircases for each participant in Experiment 1.

Staircase	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
440Hz	8.53	22.40	10.12	17.74	27.19	20.98	14.93	4.09	13.83	13.54	17.10
662Hz	11.55	40.00	11.05	11.01	21.51	18.49	13.65	6.52	16.62	16.99	7.91

In addition to fitting psychometric functions, we also used the function `glmer` in the R package `lme4` (Bates, Maechler, & Bolker, 2011) and fitted Generalized Linear Mixed Models (GLMMs) using the logit link to the data. When comparing the fitted models, we used the measure AIC_c (Sugiura, 1978), which is a finite–sample correction of the Akaike Information Criterion (Akaike, 1974; Anderson, 2008). A smaller AIC_c indicates a better model.

We fitted models with b/a and/or tone pitch as independent variables and the binomial

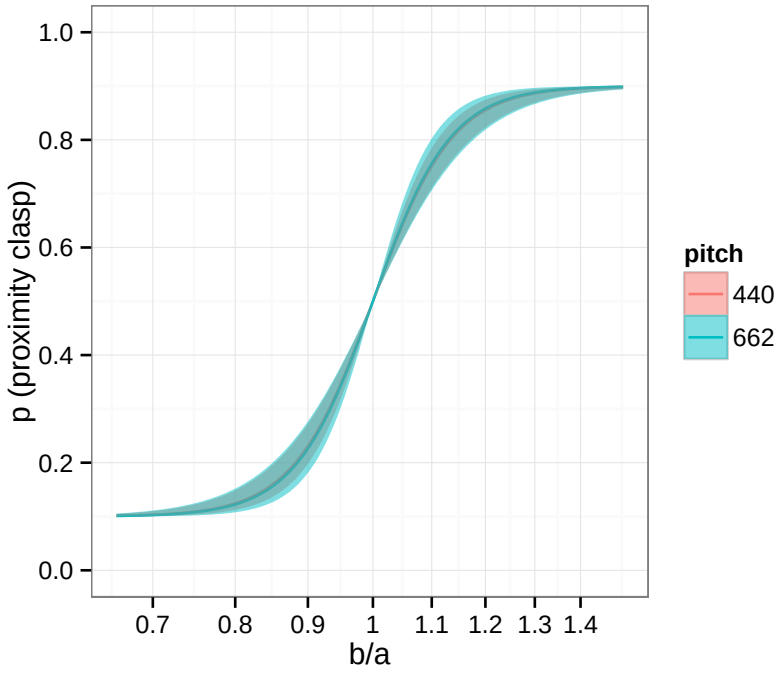


Figure 16: The fitted psychometric functions with 95% confidence intervals of Experiment 1.

responses as dependent variable. The best fitting model included only b/a without an intercept (intercept was equal to zero). Table 8 lists the AIC_c, and the ΔAIC_c for four models fitted to the data. Figure 17 depicts the results, with the line in the figure representing the fitted line of our best fitting model. The dots in the figure with 95% confidence intervals around them were predicted by another GLMM treating b/a as a categorical variable. Since we used adaptive procedure in our experiment, the number of trials are not equal across all b/a levels. The dot size in the figure represents total number of trials among all 11 participants at each level of b/a. As expected, for those large b/a levels with only few trials, the confidence intervals were very wide and were subject to a ceiling effect.

The results from both the fitted psychometric functions and the GLMMS showed that as in the visual dot lattice studies, we were able to predict the probability that

Table 8: The AIC_c, and the ΔAIC_c for four models in Experiment 1

Model	AIC_c	ΔAIC_c
Zero intercept + b/a	2089.7	0.00
Intercept + b/a	2091.5	1.80
Zero intercept + b/a + pitch level	2098.7	9.01
Intercept + b/a + pitch level	2098.7	9.01

proximity clasp was perceived by using the quantified strength of grouping by proximity. The grouping by temporal proximity principle in audition is as lawful as the grouping by spatial proximity in vision. The slope differences between the two staircases were small and were not statistically meaningful, which means the effects of b/a were the same across the two pitch levels used in the current experiments.

3.2 Experiment 2: Grouping by temporal proximity and grouping by loudness similarity

After we established the effect of grouping by temporal proximity in the perceptual organization of ANs, we took another step forward and explored the conjoint effects of two auditory grouping principles. We first examined grouping by temporal proximity and grouping by loudness similarity in Experiment 2.

3.2.1 Method

3.2.1.1 Participants Eighteen undergraduate students from the University of Virginia participated Experiment 2. None of them have participated in Experiment 1. Each of them completed two 1-hour sessions. They received introductory course credits for their participation. All of them reported normal or corrected-to-normal vision and normal hearing.

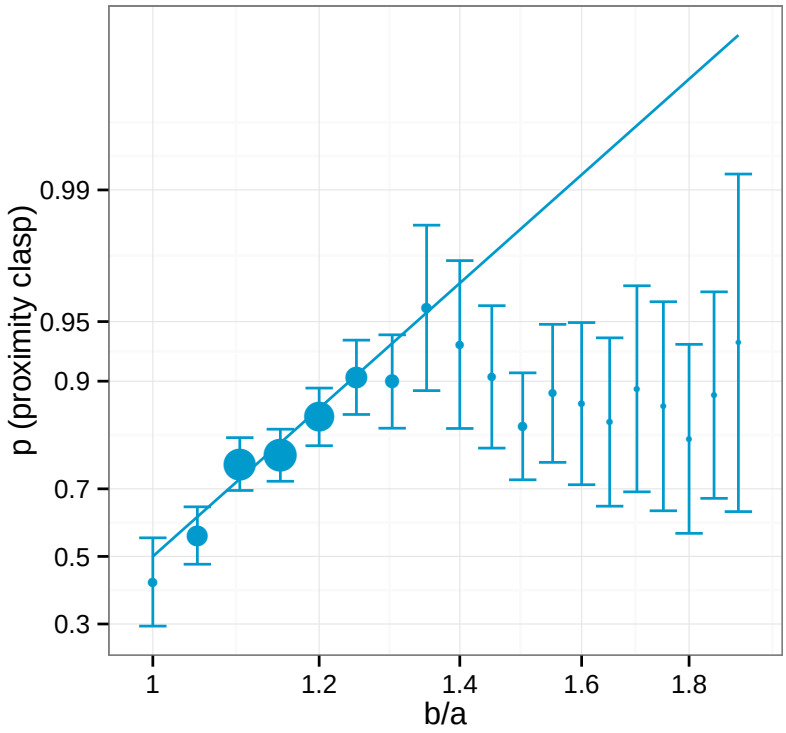


Figure 17: p (proximity clasp) as a function of b/a (SOA ratio) in Experiment 1. The size of the dot represents sample size at each level of b/a.

3.2.1.2 Stimuli Figure 18 demonstrates the stimuli used in Experiment 2. We again used non-metric ANs with four notes. SOA_a (a) was still fixed to 200ms. We used four levels of b/a in Experiment 2: 1, 1.04, 1.08, and 1.14. Based on the results of Experiment 1, these levels corresponded to 50%, 60%, 70% and 80% points in the psychometric function of an average participant.

The amplitude of note 1 (A_1) was equal to the amplitude of note 4 (A_4), while the amplitudes of note 2 (A_2) and note 3 (A_3) were equal. The amplitude difference between A_1 and A_2 (ΔA) were manipulated to quantify the principle of grouping by loudness similarity. ΔA was measured in decibels and ranged from 0 to 10 in increments of 0.5.

The duration of all notes were 50 ms with 5ms fade-in and 5ms fade-out. All notes

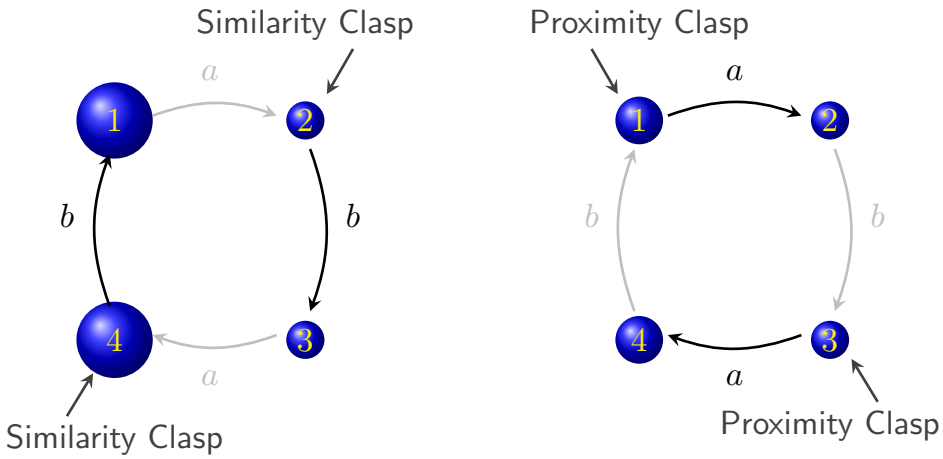


Figure 18: Auditory necklaces used in Experiment 2. Each ball represents a note. The size of the ball represents the amplitude of the note.

in an AN had the same pitch. To enhance the diversity among ANs, three pitches were used in Experiment 2 — 440Hz, 523Hz and 622Hz. Two levels of A_1 were used in the experiment (one level was 1dB higher than the other level).

In each trial of the experiment, if participants used the principle of grouping by similarity, note 2 or note 4 would be perceived as clasp. On the contrary, if participants used the principle of grouping by proximity, note 1 or note 3 would be perceived as clasp.

3.2.1.3 Design Different from Experiment 1, ΔA was the adaptively manipulated parameter in Experiment 2. Each participant completed four randomly interleaved staircases with different b/a. Each staircase contained 100 trials.

For the staircase with $b/a = 1$, we still used 1-3 UDTRS adaptive staircase procedure. ΔA would decrease a step (0.5dB) after 3 responses to similarity clasps (note 2 or 4) and would increase a step (0.5dB) after 1 response to a proximity clasp (note 1 or 3). For other 3 staircases, based on the results of Simulation 2, we used $k = 2$ UDWRS adaptive staircase procedure. ΔA would decrease a small step (0.5dB) after a response to a

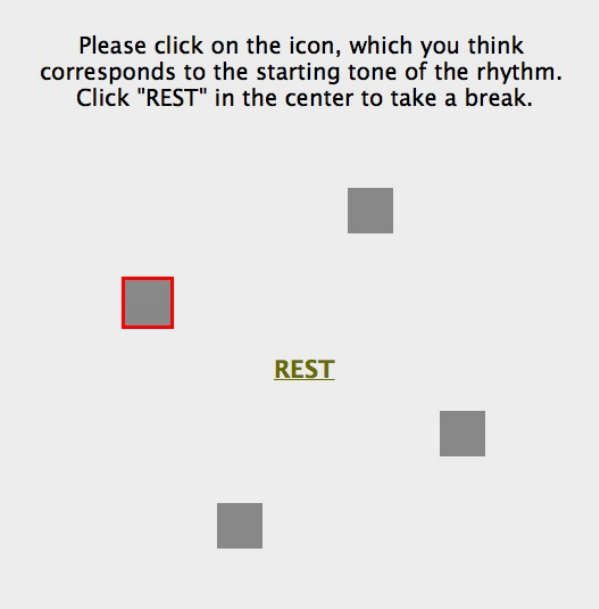


Figure 19: Screenshot of Experiment 2 display. At the moment the upper left square was highlighted.

similarity clasp and would increase a large step (1dB) after a response to a proximity clasp.

ΔA started from 5dB for all staircases. The pitch level of each AN was random among the three levels and A_1 level of each AN was random between two levels.

3.2.1.4 Procedure In Experiment 2, we used the same procedure as Experiment 1, except that only one red square showed up to highlight the square corresponding to the currently playing note after the tempo became steady (see Figure 19). This was because in Experiment 2, the participants heard repetitive 4-note auditory patterns and thus all four squares represented different notes in an AN.

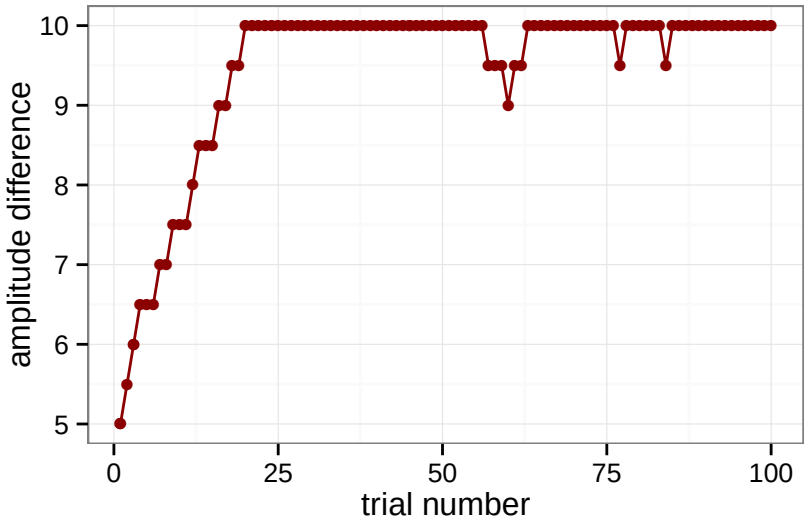


Figure 20: The change of ΔA in the staircase $b/a = 1$ of one participant. Most responses are to the proximity clasp so that ΔA increased to the maximum value very quickly.

3.2.2 Results and discussion

Among the 18 participants, 4 participants clearly did not group the ANs using the principle of grouping by loudness similarity and they were excluded from the analysis. From the responses of those participants, they seemed to consciously choose the second note of a pair of same-amplitude notes as the clasp and group the ANs as "soft–loud–loud–soft" or "loud–soft–soft–loud" repeating themselves. Figure 20 shows the change of ΔA in the staircase $b/a = 1$ of one of those 4 participants. Even though the participant cannot use the principle of grouping by temporal proximity because $b = a$, she still chose proximity clasp so that ΔA increased all the time and hit the maximum value very quickly. The reason that they grouped the ANs in this way is unknown to us. One possible explanation is that they heard the ANs as a simple melody and were grouping the ANs using some music-related principles.

In Experiment 2 and the following Experiment 3, we did not fit psychometric functions

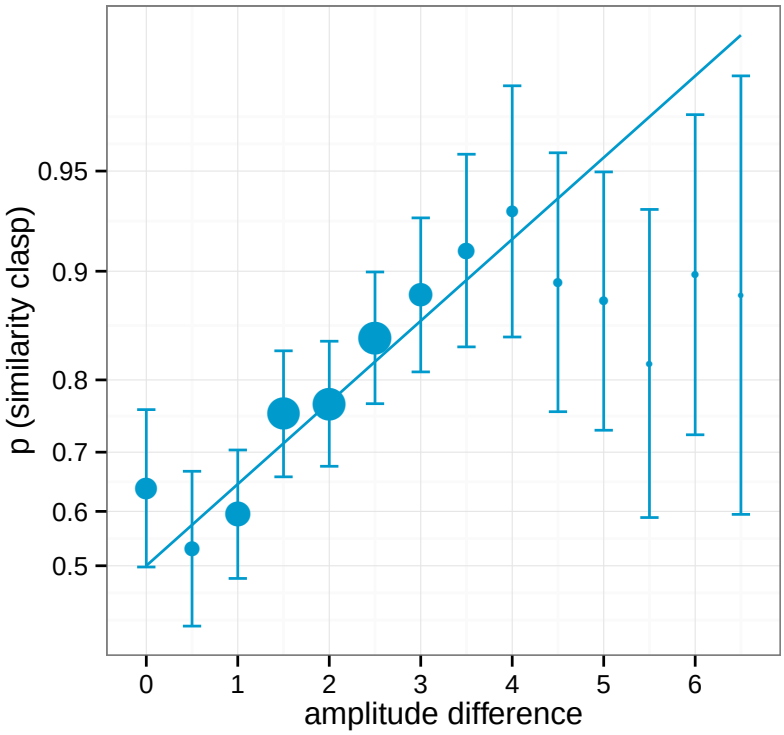


Figure 21: p (similarity clasp) as a function of ΔA for the staircase $b/a = 1$ in Experiment 2. The size of the dot represents sample size at each level of ΔA.

to the data and only fitted Generalized Linear Mixed Models (GLMMs) to the data. In the current experiment, we first established the effect of grouping by loudness similarity. We fitted a GLMM only to the responses of the staircase with $b/a = 1$ for those 14 participants. The model included zero intercept and ΔA as both fixed and random effects. Figure 21 depicts the model results. Again, the line in the figure is the fitted line. The dots with error bars representing 95% confidence intervals are predicted by another GLMM treating ΔA as a categorical variable. The dot size in the figure represents total number of trials among all 14 participants at each level of ΔA. The results showed that most participants used grouping by loudness similarity to group the ANs. The model showed that we were able to predict the probability that similarity clasp was perceived

Table 9: The AIC_c, and the ΔAIC_c for four models in Experiment 2

Model	AIC_c	ΔAIC_c
Intercept + b/a + ΔA	6161.32	0.00
Zero intercept + b/a + ΔA	6161.89	0.58
Zero intercept + b/a + ΔA + $b/a \times \Delta$A	6167.74	6.42
Intercept + b/a + ΔA + $b/a \times \Delta$A	6168.57	7.25

by using the quantified strength of grouping by loudness similarity.

After we established the effect of grouping by loudness similarity, we fitted GLMMs to the responses of all staircases. Table 9 lists the AIC_c, and the ΔAIC_c for four models we fitted. The two models without interaction performed far better than the models with interaction. The ΔAIC_c between the two models without interaction is very small so we decided to use the model with zero intercept as our final model because the intercept was theoretically zero due to the nature of stimuli. Figure 22 depicts the fitted line of this final model. The dots with error bars representing 95% confidence intervals are predicted by another GLMM treating both b/a and ΔA as categorical variables. The dot size represents the sample size at each stimulus level.

The results showed that the quantified strength of grouping by loudness similarity predicted the probability that similarity clasp would be perceived very well. When the two grouping principles — grouping by temporal proximity and grouping by loudness similarity— were applied to the same AN, participants used both grouping cues to perceive the auditory pattern. Importantly, as what was found in vision, the conjoint effect of those two grouping principles are additive. The two grouping principles work independently.

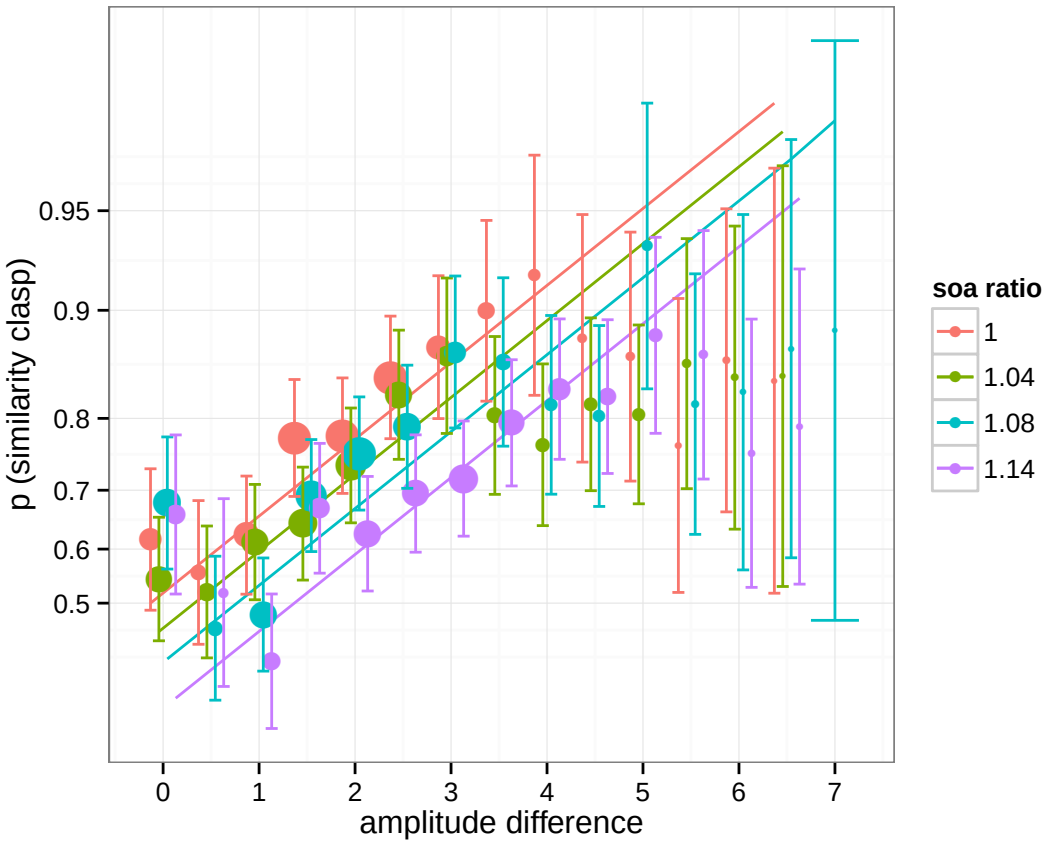


Figure 22: p (similarity clasp) as a function of ΔA for all staircase in Experiment 2. The size of the dot represents sample size at each stimulus level.

3.3 Experiment 3: Grouping by temporal proximity and grouping by pitch similarity

In Experiment 3, we explored the effect of another grouping by similarity principle in audition — grouping by pitch similarity. The separate and conjoint effects of grouping by temporal proximity and grouping by pitch similarity was examined.

3.3.1 Experiment 3a

3.3.1.1 Method

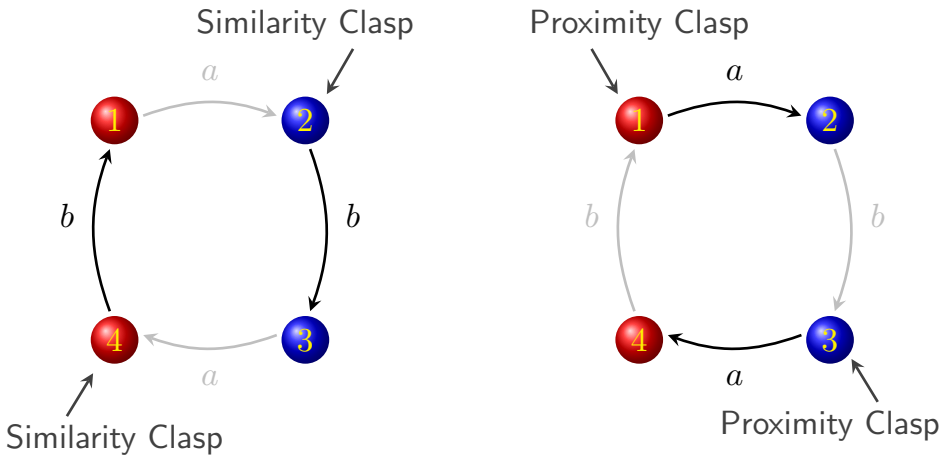


Figure 23: Auditory necklaces used in Experiment 3. Each ball represents a note. Different colors represent different pitches.

Participants Ten undergraduate students from the University of Virginia participated in Experiment 3a. None of them had participated in Experiment 1 or 2. Each of them completed two 1-hour sessions. They received introductory course credits or payment for their participation. All of them reported normal or corrected-to-normal vision and normal hearing.

Stimuli Figure 23 demonstrates the non-metric ANs used in Experiment 3a. SOA_a (a) was fixed to 200ms, and the same four levels of b/a in Experiment 2 were used in Experiment 3: 1, 1.04, 1.08, and 1.14.

The frequency of note 1 (f_1) was equal to that of note 4 (f_4), while the frequency of note 2 (f_2) and note 3 (f_3) were equal. The frequency difference between f_1 and f_2 (Δf) was manipulated to quantify the principle of grouping by pitch similarity. Δf was measured in cents (1/100 of a semitone) and ranged from 0 to 80 in increments of 2.

The duration of all notes were 50 ms with 5ms fade-in and 5ms fade-out. All notes in an AN had the same amplitude. Again, to enhance the diversity among ANs, two

levels of f_1 were used in Experiment 3a — 440Hz, and 622Hz. We also used two levels of amplitude in the experiment (one level was 1dB higher than the other level).

Similar to Experiment 2, if participants used the principle of grouping by similarity, note 2 or note 4 would be perceived as clasp. On the contrary, if participants used the principle of grouping by proximity, note 1 or note 3 would be perceived as clasp.

Design and procedure The design was similar to Experiment 2. Δf was the adaptively manipulated parameter in Experiment 3a. Each participant completed four randomly interleaved staircases with different b/a levels. Each staircase contained 100 trials.

We again used 1-3 UDTRS adaptive staircase procedure for the staircase with $b/a = 1$, and used $k = 2$ UDWRS adaptive staircase procedure for other 3 staircases. The downward step size was 2 cents for all staircases. Δf started from 20 cents for all staircases.

The procedure was exactly the same as Experiment 2.

3.3.1.2 Results and discussion

As in Experiment 2, we first attempted to establish the effect of grouping by pitch similarity by fitting a GLMM only to the responses of the staircase with $b/a = 1$. The model included zero intercept and Δf as both fixed and random effects. Figure 24 depicts the fitted line and the dots with 95% confidence intervals predicted by another GLMM treating Δf as a categorical variable. Although the model showed that there is a significant effect of Δf, suggesting that participants sometimes do use grouping by pitch similarity to group the ANs, the effect of Δf is very weak. The weak effect of Δf was probably because of the small increments and range of Δf that we used in the current experiment. The largest Δf in the current experiment was 80 cents, which was smaller than a semitone. It might have been very hard for the

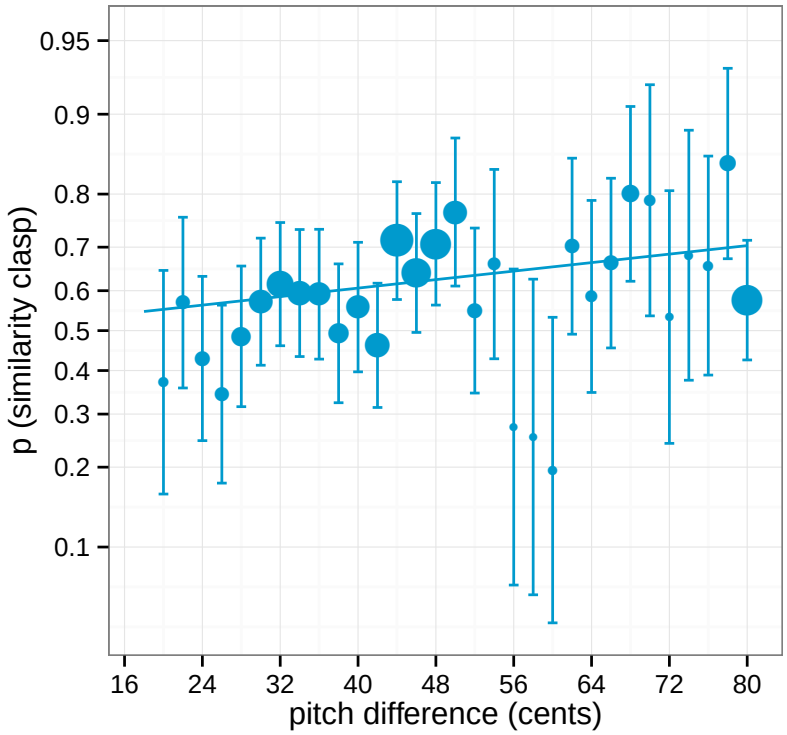


Figure 24: p (similarity clasp) as a function of Δf for the staircase $b/a = 1$ in Experiment 3a. The size of the dot represents sample size at each level of Δf.

participants to discriminate the pitch difference among the tones in the ANs. To account for this, we increased the increments and range of Δf in Experiment 3b to obtain more accurate results.

Despite of the small effect of Δf, we still established a significant effect of grouping by pitch similarity. Therefore, we then fitted GLMMs to the responses of all staircases. Table 10 lists the AIC_c, and the ΔAIC_c for the six models we fitted. The model with only zero intercept and Δf was the best fitting model in the current experiment. In those models including b/a as a predictor, its effect was not statistically significant. Figure 25 depicts the fitted line of the best fitting model, and the dots with 95% confidence interval predicted by another GLMM treating both b/a and Δf as categorical variables.

Table 10: The AIC_c, and the ΔAIC_c for six models in Experiment 3a

Model	AIC_c	ΔAIC_c
Zero intercept + Δf	5248.52	0.00
Intercept + Δf	5250.08	1.56
Zero intercept + b/a + Δf	5254.00	5.49
Intercept + b/a + Δf	5254.41	5.89
Zero intercept + b/a + Δf + $b/a \times \Delta$f	5263.88	15.36
Intercept + b/a + Δf + $b/a \times \Delta$f	5263.89	15.37

The small but significant effect of Δf showed that as with grouping by loudness similarity, we can also predict the probability that the similarity clasp was perceived by the strength of grouping by pitch similarity. However, when the the principle of grouping by pitch similarity was combined with grouping by temporal proximity, participants ignored temporal proximity and only used pitch similarity to group the ANs.

3.3.2 Experiment 3b

In Experiment 3b, we increased the interval and range of Δf to confirm our findings in Experiment 3a.

3.3.2.1 Method

Participants Seven undergraduate students from the University of Virginia participated Experiment 3b. Two of them have participated in Experiment 3a. They received introductory course credits or payment for their participation. All of them reported normal or corrected-to-normal vision and normal hearing.

Stimuli The ANs used in Experiment 3b were the same as those used in Experiment 3a except that Δf ranged from 0 to 400 cents in increments of 10 cents.

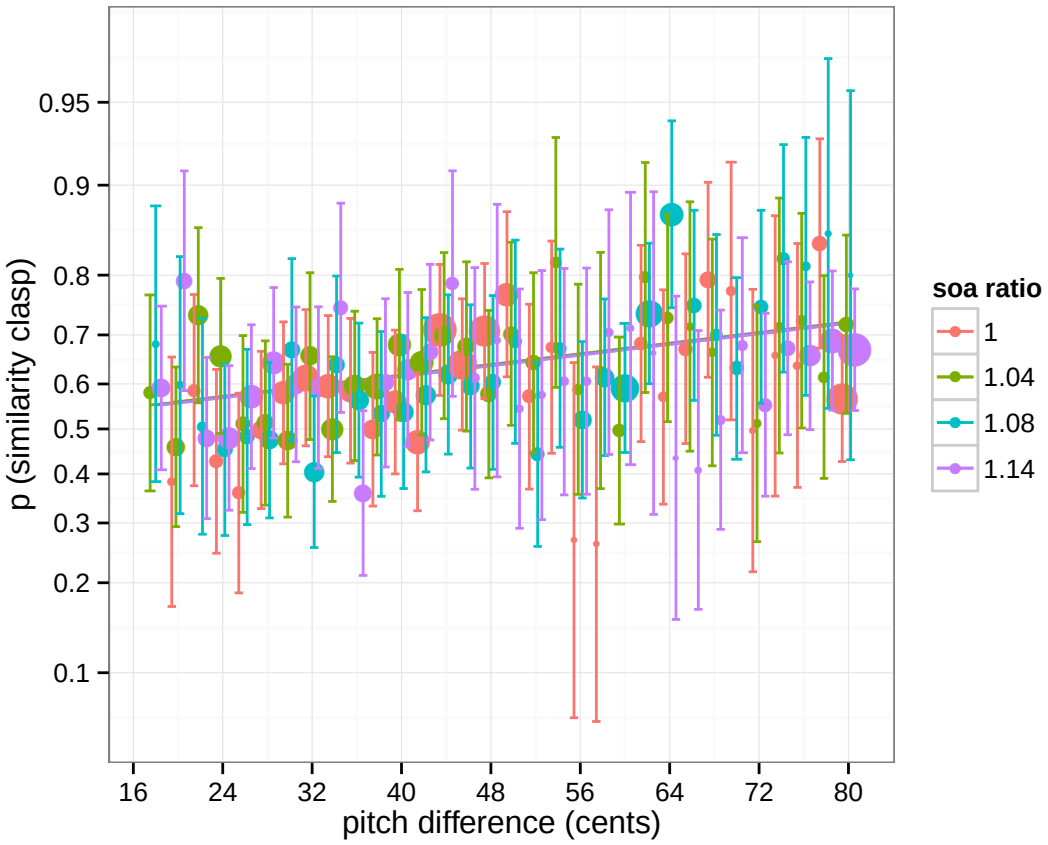


Figure 25: p (similarity clasp) as a function of Δf for all staircase in Experiment 3a. The size of the dot represents sample size at each stimulus level.

Design and procedure The design and the procedure was the same as Experiment 3a except that each staircase contained 80 trials and the participants completed the whole experiment in a single 1.5-hour session. Δf started from 100 cents for all staircases.

3.3.2.2 Results and discussion Among the 7 participants, 2 participants obviously did not group the ANs using the principle of grouping by pitch similarity and they were excluded from further analysis. For reasons unknown to us, similar to the 4 participants in Experiment 2, those 2 participants seemed to consciously choose the second note of a pair of same-pitch notes as the clasp and group the ANs as "low–high–high–low" or

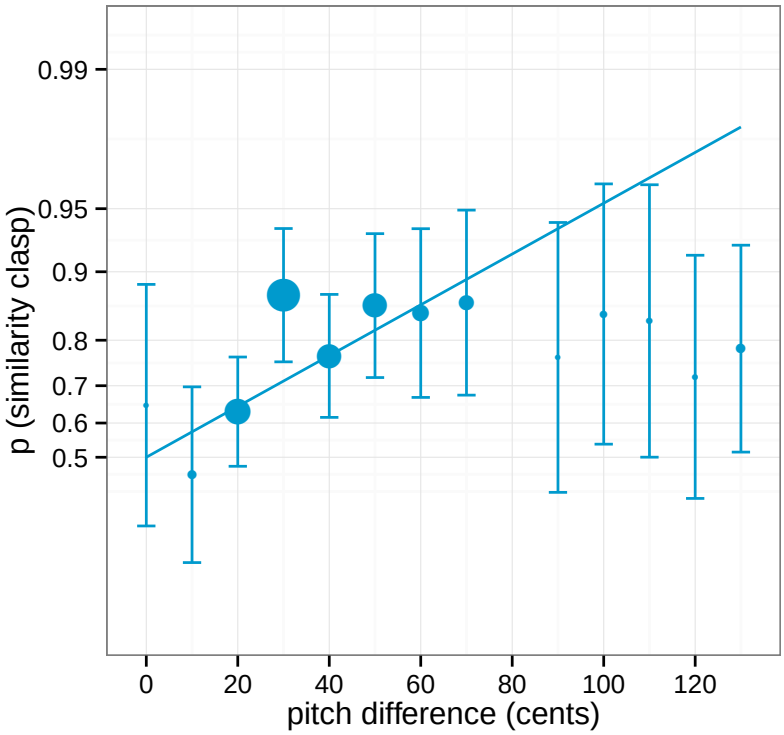


Figure 26: p (similarity clasp) as a function of Δf for the staircase $b/a = 1$ in Experiment 3b. The size of the dot represents sample size at each level of Δf.

"high–low–low–high" repeating themselves.

We then fitted the same GLMM as in Experiment 3a to only the responses of the staircase with $b/a = 1$ for the remaining 5 participants. Figure 26 depicts this model with zero intercept and Δf as predictor. The effect of Δf was much clearer than it was in Experiment 3a due to our new manipulation of Δf. The current results clearly showed that the participants could use grouping by pitch similarity to group the ANs, and we were able to predict the probability that the similarity clasp was perceived by using the quantified strength this grouping principle.

The same GLMMs as in Experiment 3a were then fitted to the responses of all staircases. Table 11 lists the AIC_c, and the ΔAIC_c for the six models we fitted. The results

Table 11: The AIC_c, and the ΔAIC_c for six models in Experiment 3b

	AIC_c	ΔAIC_c
Intercept + Δf	1764.00	0.00
Intercept + b/a + Δf	1765.24	1.23
Zero intercept + Δf	1770.72	6.72
Intercept + b/a + Δf + interaction	1771.30	7.30
Zero intercept + b/a + Δf	1771.40	7.39
Zero intercept + b/a + Δf + interaction	1772.62	8.61

of model fitting in the current experiment were very similar to those in Experiment 3a. Among models with intercept, the model with only Δf as a predictor was better than the other models, which was the same among models with zero intercept. For the models with both b/a and Δf as predictors, the effects of b/a were not statistically significant. Because the model with intercept was far better ($\Delta AIC_c = 6.72$) than the model with zero intercept, we used that model as our final model. Figure 27 depicts the fitted line of this model, and the dots with 95% confidence intervals predicted by another GLMM treating both b/a and Δf as categorical variables.

After we increased the range of Δf in the current experiment, the results showed that the participants used the principle of grouping by pitch similarity very well, as they used the principle of grouping by loudness similarity and grouping by temporal proximity. We can predict the probability a clasp was perceived by the quantified strength of grouping by pitch similarity very well. The results also confirmed another important finding in Experiment 3a: when the two grouping principles — grouping by temporal proximity and grouping by pitch similarity— were applied to the same AN, participants only used pitch similarity to group ANs and they totally ignored temporal proximity.

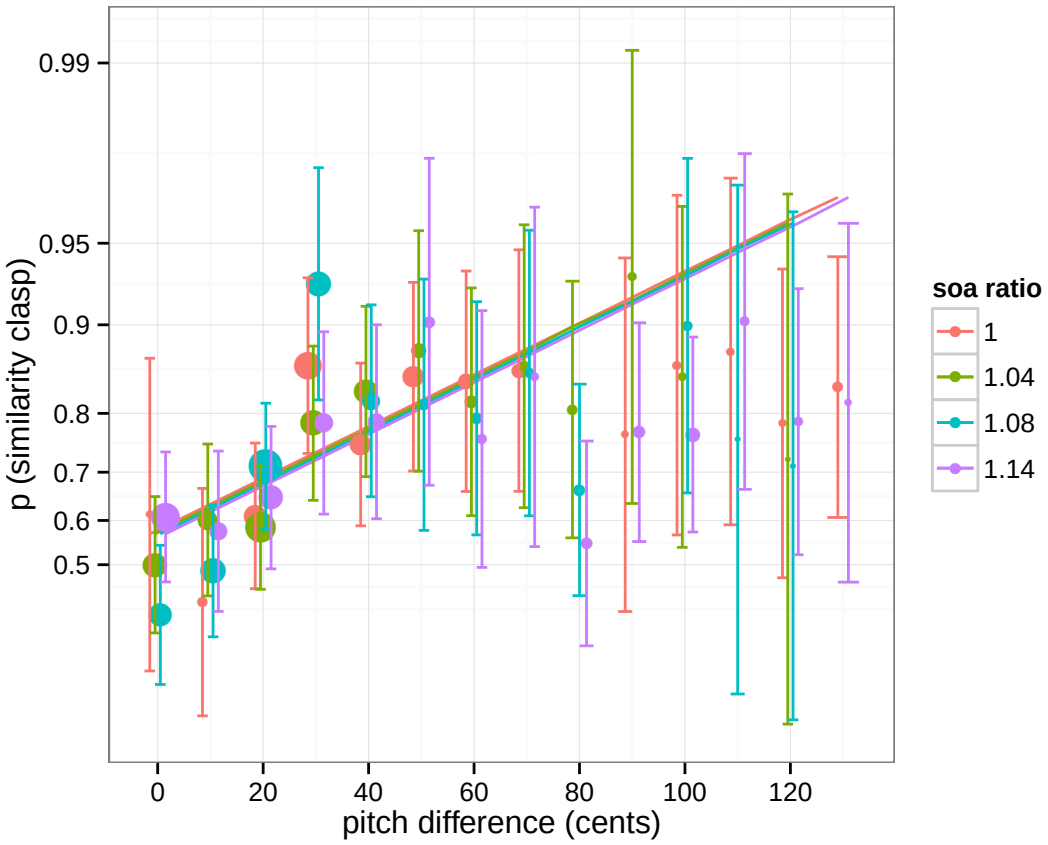


Figure 27: p (similarity clasp) as a function of Δf for all staircase in Experiment 3b. The size of the dot represents sample size at each stimulus level.

4 General Discussion

4.1 Grouping in audition

Grouping by proximity and grouping by similarity are two classic grouping principles that has been intensively studied in vision (e.g. Rock & Brosgole, 1964; Beck, 1966; Oyama et al., 1999; Kubovy & van den Berg, 2008). A number of studies and demonstrations have shown that participants depend on spatial proximity, luminance similarity, color similarity and shape similarity to group visual objects. Moreover, if the strengths of

these grouping principles are quantified appropriately, it is possible to predict the probability that each competitive perceptual organization is perceived by those quantitative measurements (Kubovy, 1994; Kubovy et al., 1998; Oyama et al., 1999).

In this dissertation, we attempted to build a direct analogy between the two classic visual grouping principles of similarity and proximity and three auditory grouping principles: grouping by temporal proximity, grouping by loudness similarity, and grouping by pitch similarity. The results of our three experiments showed that the participants depended on all of these grouping principles to group auditory patterns: they successfully used temporal proximity, loudness similarity and pitch similarity to identify the starting point of ambiguous auditory patterns and group sequences of repetitive notes into meaningful units.

Similar to studies in vision, we quantified the strengths of these principles and built models to predict the auditory perceptual organizations. For proximity, Kubovy et al. (1998) used distance aspect ratio as the quantification of the strength of spatial proximity. We used similar time aspect ratio (SOA ratio) to quantify the strength of temporal proximity in audition. For similarity, we also used similar measurements to those used in vision (e.g. luminance difference). We quantified loudness similarity by taking the amplitude difference between notes (measured in decibel) and quantified pitch similarity by taking the frequency difference between notes (measured in cents). These quantitative measurements predicted the probability of each perceived organization very well in our models.

Before the current dissertation, we studied the run principle and the gap principle proposed by Garner and his colleagues (Yu & Kubovy, submitted) and have shown that for metric ANs, we were able to predict the probability of perceived clasps by run length ratio and gap length ratio — the quantitative measurements of the two principles. The

run principle is a perceptual organization principle specific to the auditory system, which we discuss later. The gap principle is similar to the principle of grouping by temporal proximity for metric auditory patterns. In the current dissertation, we generalized the lawful effect of gap principle for metric ANs to the effect of a general grouping by temporal proximity principle for non-metric ANs.

The paradigm and modeling framework of the current work has been used to quantitatively study many Gestalt problems in both vision and audition effectively. Our findings in audition indicated that as in vision, as long as we are able to appropriately quantify the physical characteristics of the strength of an auditory grouping principle, we can predict the percepts of the stimuli — the psychological output — by those measurements. Thereafter, we can apply the same paradigm to study other Gestalt problems and the mechanism of various grouping principles.

4.2 The relationship among Gestalt principles

In addition to the lawful effects of the three auditory grouping principles, a more important finding was that the conjoint effect of grouping by temporal proximity and grouping by loudness similarity in Experiment 2 was additive. We did not find similar additive effect between grouping by temporal proximity and grouping by pitch similarity in Experiment 3. However, rather than interactive conjoint effect, we found there was no effect of grouping by temporal proximity when those two principles were applied to the same stimuli. We will discuss this in the next section.

The non-additivity and non-linearity is essential to Gestalt effects as one of the most important claims of Gestalt psychology is "the whole is not equal to the sum of its parts". Nobody would question that Gestalt grouping itself is a non-linear system. We cannot just sum up the separate information provided by many auditory notes or visual dots

to get meaningful repetitive units or columns of dots grouped together. We have to use Gestalt principles on the *whole* auditory or visual patterns to perceive these meaningful groupings. However, a whole composed of non-additive Gestalts may not need to be a non-additive Gestalt itself (see also Kubovy & van den Berg, 2008). The effects of those Gestalts that form a whole may be additive.

This is not the first time that an additive conjoint effect was found. In audition, the conjoint effects of run principle, gap principle and accent principle (the note with increased amplitude is more likely to be perceived as clasp) have been shown to be additive (Yu & Kubovy, 2011, submitted). In vision, Kubovy and van den Berg (2008) demonstrated that the effects of grouping by proximity and grouping by similarity were additive in dot lattices. Yu and Kubovy (2012) demonstrated that the effects of relative area and convexity were additive in figure ground perception. Using another paradigm, Luna and Montoro (2011) examined the interactions among three grouping principles in vision — proximity, similarity and common region. They manipulated the strengths of those grouping cues and asked the participants to rate the perceived grouping. Their results also showed that there was no interaction among those grouping principles.

But this sort of additivity found in multiple Gestalt phenomena is not inevitable. Strother and Kubovy (2012) demonstrated that the manipulations of density, curvature and aspect ratio interacted with each other. Gepshtein and Kubovy (2000) used saptio-temporal dot lattices as stimuli and independently manipulated spatial and temporal cues. Their results showed that temporal cues could affect spatial grouping, indicating non-additivity between spatial and temporal grouping principles.

Although evidence of additive effects among grouping principles have been accumulating, a more thorough survey of how various Gestalt principles work together in multiple sensory modalities is needed for a generalized theory about when and why Gestalt princi-

ples work additively and non-additively. For example, in audition, a natural extension of the current experiments is to explore how the principles of grouping by loudness similarity and grouping by pitch similarity work together.

As we study how multiple principles work together in more Gestalt phenomena, a more important question we need to answer now is why we find counterintuitive additivity in all kinds of Gestalt phenomena. Kubovy and Yu (2012) conjectured that additive conjoint effects are found when the conjoined grouping principles do not give rise to a new emergent property. In all those examples of non-additivity, some new emergent properties seem to appear. The participants may have perceived three dimensional sphere-like objects for the curved dot lattices in the experiments of Strother and Kubovy (2012). This 3-D sphere percept give rise to an emergent property. For the saptio-temporal dot lattices in Gepshtein and Kubovy (2000), the apparent motion might produce a new emergent property.

For future studies, we could design experiments to test this theory that new emergent properties are essential for non-additivity. If we can delicately design stimuli in which several Gestalt cues give rise to new emergent properties, we can examine whether the effects of those cues are non-additive to each other. A potential new emergent property in audition is melody which is similar to the apparent motion (or common fate) in vision.

4.3 Vision and audition

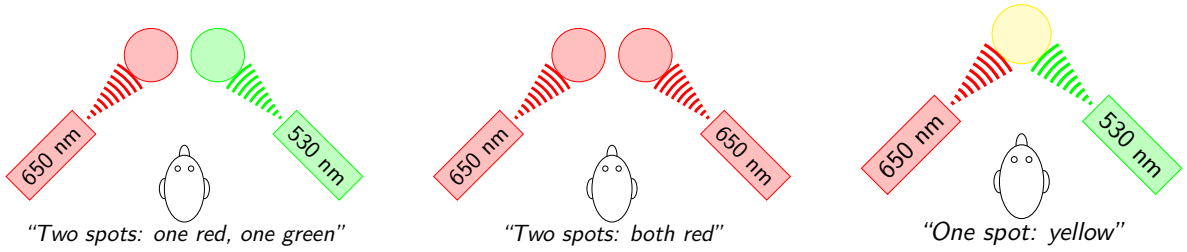
Although we aimed at building a direct analogy between audition and vision in the current dissertation, there is no question that the two sensory modality are different. Kubovy and his colleagues have argued how vision and audition are linked to each other (Kubovy, 1988; Kubovy & Van Valkenburg, 2001; Kubovy & Schutz, 2010) by proposing two dualities between vision and audition.

The first duality deals with the functions of the two modalities. They argued that the primary function of vision is to detect *surfaces* whereas that of audition is to detect *sources*. The secondary function of vision concerns *sources* while the secondary function of audition concerns *surfaces*.

The second duality is the Theory of Indispensable Attributes (TIA) duality. To speak of both visual and auditory objects, Kubovy and Van Valkenburg (2001) offered an operational definition: "A *perceptual object* is that which is susceptible to figure-ground segregation". Based on this definition, they argued that space and time are the two indispensable attributes in vision whereas pitch and time are the two indispensable attributes in audition. To demonstrate this argument, they conducted two thought–experiments, one in vision and one in audition.

The visual thought–experiment runs as follows: we let an observer to look at two colored spots of light (Figure 28a) and ask him or her "how many entities are visible?" We assume the answer is two. Then we collapse over wavelength and show the observer two spots with the same color (Figure 28b), the observer will still say, two. Hence wavelength is not an indispensable attribute for visual figure-ground segregation. But if we collapse over space (Figure 28c), the observer will respond (because of color metamerism), one. Hence spatial location is an indispensable attribute for visual figure-ground segregation. We can replace space with time in this thought–experiment and obtain the same results.

Similarly, the auditory thought–experiment runs as follows: we ask an observer to listen to two sounds (of different pitch) over two loudspeakers (Figure 29a) and ask him or her "how many entities are audible?" We assume the answer is two. If we collapse the display over space (Figure 29b), the observer will still say, two. Hence spatial location is not an indispensable attribute for auditory figure-ground segregation. But if we collapse over frequency (Figure 29c), the observer will respond (because of auditory localization

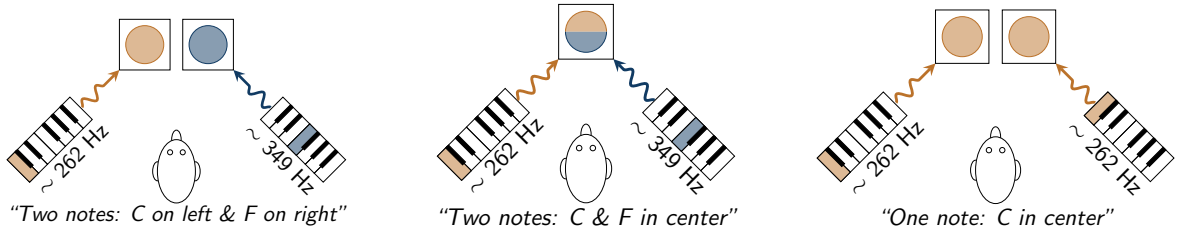


(a) Start.

(b) Collapse over wavelength: not indispensable.

(c) Collapse over space: indispensable.

Figure 28: *Theory of Indispensable Attributes*: The visual thought–experiment.



(a) Start.

(b) Collapse over space: not indispensable.

(c) Collapse over frequency: indispensable.

Figure 29: *Theory of Indispensable Attributes*: The auditory thought–experiment.

mechanisms), one. Hence frequency is an indispensable attribute for auditory figure-ground segregation. Again, we can replace pitch with time in this thought–experiment and obtain the same results.

In visual grouping experiments, participants group objects in space (e.g. group dots into rows or columns) by using various Gestalt grouping principles. However, since space is not an indispensable attribute in audition, we can only ask participants to group auditory objects in pitch or in time. In the current experiments, we asked participants to group notes in time by using temporal proximity, loudness similarity and pitch similarity cues. We did not ask participants to group concurrently played sounds in pitch space because concurrent segregation is very hard in audition and sometimes requires participants to have formal training.

Despite of the difference in grouping dimension between audition and vision, we have found consistent results between the two modalities. In our previous studies (Yu & Kubovy, 2011, submitted), we have studied grouping principles specific to grouping in time such as the run principle and the accent principle. The run principle enables participants to identify a long sequence of consecutive notes and choose the first note of this sequence as the clasp. The accent principle enables participants to use the increased amplitude as a cue to identify the clasp. If visual experiments of grouping in time are conducted, we should be able to find the effects of similar principles.

In Experiment 3, we found that when grouping by temporal proximity and grouping by pitch similarity were applied to the same stimuli, participants ignored temporal proximity and depended only on pitch similarity to group the stimuli. If we increase b/a of the ANs to be very large, the participants may be able to perceive the proximity clasps in the experiments. But we used the same b/a levels in Experiment 2 and Experiment 3. And participants did sometimes use temporal proximity cues when it was applied to the ANs with loudness similarity in Experiment 2, whereas they did not use it when applied to the ANs with pitch similarity in Experiment 3. In Experiment 3, even when $b/a = 1.14$, a level that an average participant would perceive proximity clasps 80% if no other grouping cue is available, participants did not depend on temporal proximity to group the ANs. In Experiment 3a, the pitch differences were very small and some participants may not even be able to discriminate such differences. But they still ignored the temporal proximity cue in the experiment. Therefore, the results we found in Experiment 3 was not just due to our stimulus manipulation. There is something special for grouping by pitch similarity or the combination of grouping by pitch similarity and grouping by temporal proximity.

The results may suggest that pitch is a more important indispensable attribute than

time in audition. Time is very important in audition, and may be more important than time in vision (e.g. Bertelson & Aschersleben, 2003; Aschersleben & Bertelson, 2003). But the complexity of pitch (and timbre) information (including harmonic spread over frequency, and envelopes modulated in time) may let it play a more essential role in fulfilling the primary function of our auditory system: detecting and segregating the sources.

Another interesting but mysterious finding of our experiments is that although most participants followed our manipulation and used the principles of grouping by loudness similarity and grouping by pitch similarity, a small number of participants grouped the ANs in a very different way. It may reflect that our auditory system is more flexible than our visual system. We conjecture that formal musical training or simply exposure to certain type of music may be the reason that those participants behave differently.

To better understand the linkage between audition and vision, it will be very interesting and very important to build a more direct analogy between vision and audition. Because time is the shared indispensable attributes in both vision and audition, we can explore whether the grouping principles in audition such as run and gap principles also work in vision and how they work together in vision. Of course, if we can recruit participants to complete a well-designed auditory experiments of concurrent grouping in pitch, it will be a more direct analogy to those visual experiments of grouping in space.

4.4 Conclusion

In three experiments, this dissertation explored the separate and conjoint effects of grouping by temporal proximity, grouping by loudness similarity and grouping by pitch similarity. The separate effects of all three principles were found to be lawful, as their counterparts do in vision. When conjointly applied to the same stimuli, grouping by temporal

proximity and grouping by loudness similarity worked additively as grouping by spatial proximity and grouping by luminance similarity worked additively in vision. However, participants depended only on grouping by pitch similarity when it was applied to the stimuli together with grouping by temporal proximity. This may be because pitch plays a more important role than time in the auditory system. The quantitative psychophysical approach used in the current dissertation could be applied to more Gestalt phenomena and lead to a better understanding of how our perceptual systems use multiple cues in perceptual organization.

References

Akaike, H. (1974, November). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *AC-19*(6), 716–723.

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. New York, NY, USA: Springer.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension2. evidence from sensorimotor synchronization. *International Journal of Psychophysiology*, *50*, 157–163.

Bates, D., Maechler, M., & Bolker, B. (2011). `lme4`: Linear mixed-effects models using S4 classes [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999375-42)

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics*, *1*, 300–302.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension1. evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*, 147–155.

Boker, S. M., & Kubovy, M. (1998). The perception of segmentation in sequences: Local information provides the building blocks for global structure. In D. A. Rosenbaum & C. E. Collyer (Eds.), *Timing of behavior: Neural, computational, and psychological perspectives* (pp. 109–123). Cambridge, MA, USA: MIT Press.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389.

Ellis, W. D. (Ed.). (1938). *A source book of Gestalt psychology*. New York, NY, USA: Harcourt, Brace and Company.

García-Pérez, M. A., & Alcalá-Quintana, R. (2005). Sampling plans for fitting the psychometric function. *The Spanish Journal of Psychology*, *8*, 256–289.

Gepshtein, S., & Kubovy, M. (2000, July). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(14), 8186–8191.

Kubovy, M. (1988, May). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception & Performance*, *14*(2), 318–320.

Kubovy, M. (1994). The perceptual organization of dot lattices. *Psychonomic Bulletin & Review*, *1*(2), 182–190.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*(1), 71–98. doi: 10.1006/cogp.1997.0673

Kubovy, M., & Schutz, M. (2010). Audio-visual objects. *Review of Philosophy and Psychology*, *1*, 41–61. doi: 10.1007/s13164-009-0004-5

Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115*, 131–154.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*(1–2), 97–126. doi: 10.1016/S0010-0277(00)00155-4

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science*, *6*(4), 225–234.

Kubovy, M., & Yu, M. (2012). Multistability, cross-modal binding and the additivity of conjoined grouping principles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 954–964. Retrieved from http://rstb.royalsocietypublishing.org/content/367/1591/954.abstract doi: 10

.1098/rstb.2011.0365

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception*, *11*(2), 115-128.

Luna, D., & Montoro, P. R. (2011). Interactions between intrinsic principles of similarity and proximity and extrinsic principle of common region in visual perception. *Perception*, *40*, 1467–1477.

Marr, D. (1982). *Vision*. New York, NY, USA: W.H. Freeman.

Martin, J. G. (1972). Rhythmic (heirarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79*, 487–509.

Oyama, T. (1961). Perceptual grouping as a function of proximity. *Perceptual & Motor Skills*, *13*, 305–306.

Oyama, T., Simizu, M., & Tozawa, J. (1999). Effects of similarity on apparent motion and perceptual grouping. *Perception*, *28*, 739–748.

Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, *5*, 253–259.

Peterson, M. A., & Lampignano, D. W. (2003). Implicit memory for novel figure-ground displays includes a history of cross-border competition. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(4), 808–822.

Preusser, D., Garner, W. R., & Gottwald, R. L. (1970). Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, *83*(2), 151-170.

Rock, I., & Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology*, *67*, 531–538.

Royer, F. L., & Garner, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, *1*, 41-47.

Royer, F. L., & Garner, W. R. (1970). Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics*, *7*, 115-120.

Ruskey, F. (2011). Information on necklaces, lyndon words, de bruijn sequences. In *The (Combinatorial) Object Server* (May 23, 2011 ed.). Retrieved from http://www.theory.csc.uvic.ca/~cos/inf/neck/NecklaceInfo.html (Retrieved on February 24, 2012 from http://www.theory.csc.uvic.ca/~cos/inf/neck/NecklaceInfo.html)

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C. J., & Pressnitzer, D. (2012). Multistabilitiy in perception: binding sensory modalities, an overview. *Philosophic Transactions of the Royal Society, B*, *367*, 869–905. (Introduction to a theme issue)

Strother, L., & Kubovy, M. (2012). Structural salience and the nonaccidentality of a gestalt. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 827-832.

Sugiura, N. (1978). Further analysis of the data by akaikes information criterion and the finite corrections. *Communications in Statistics*, *A7*, 13–26.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Investigations of the principles of Gestalt, II]. *Psychologische Forschung*, *4*, 301–350. (Translated extract in Ellis, 1938, pp. 71–88.)

Yu, M., & Kubovy, M. (2011). The additivity of organizational principles in the segmentation of auditory necklaces. Poster presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA.

Yu, M., & Kubovy, M. (2012). Quantifying the organizational principles in figure-ground segregation. Poster presented at the 53rd Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Yu, M., & Kubovy, M. (submitted). The whole is equal to the sum of its parts: An auditory remix. *Journal of Experimental Psychology: Human Perception and Performance*.

Appendix

Histograms of estimated parameters in the simulation study

Table 12: List of sampling methods in Appendix figures

	Method
Method 01	1-2, Δ-=0.03
Method 02	1-2, Δ-=0.06
Method 03	1-3, Δ-=0.03
Method 04	1-3, Δ-=0.06
Method 05	k=2, Δ-=0.03
Method 06	k=2, Δ-=0.06
Method 07	k=3, Δ-=0.03
Method 08	k=3, Δ-=0.06
Method 09	1-2/1-3, Δ-=0.03
Method 10	1-2/1-3, Δ-=0.06
Method 11	k=2/k=3, Δ-=0.03
Method 12	k=2/k=3, Δ-=0.06

Figure 30: Histograms of estimated bs in Simulation 1 from runs of 100 trials. Each histogram includes 1000 simulated runs.

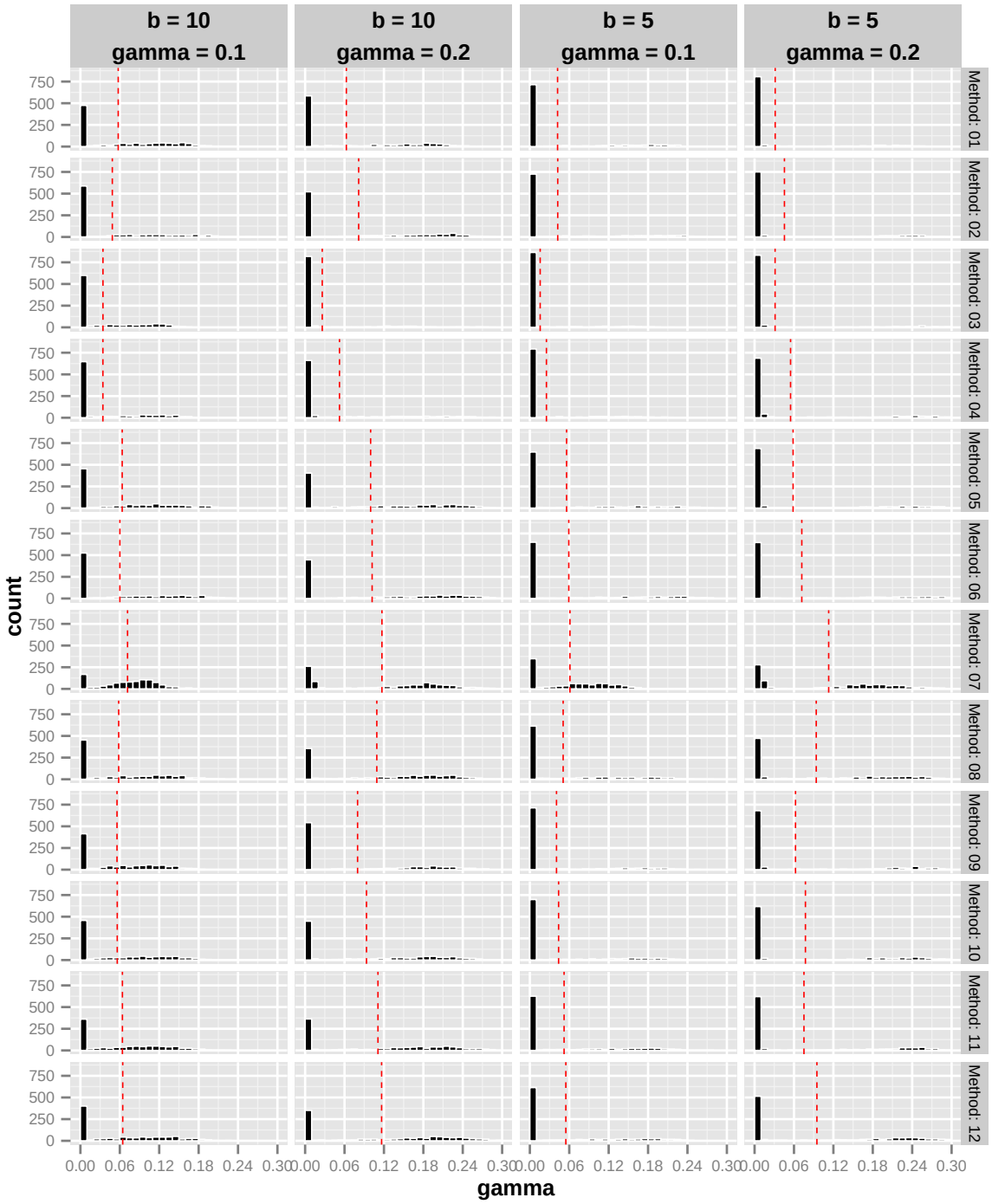


Figure 31: Histograms of estimated γs in Simulation 1 from runs of 100 trials. Each histogram includes 1000 simulated runs.

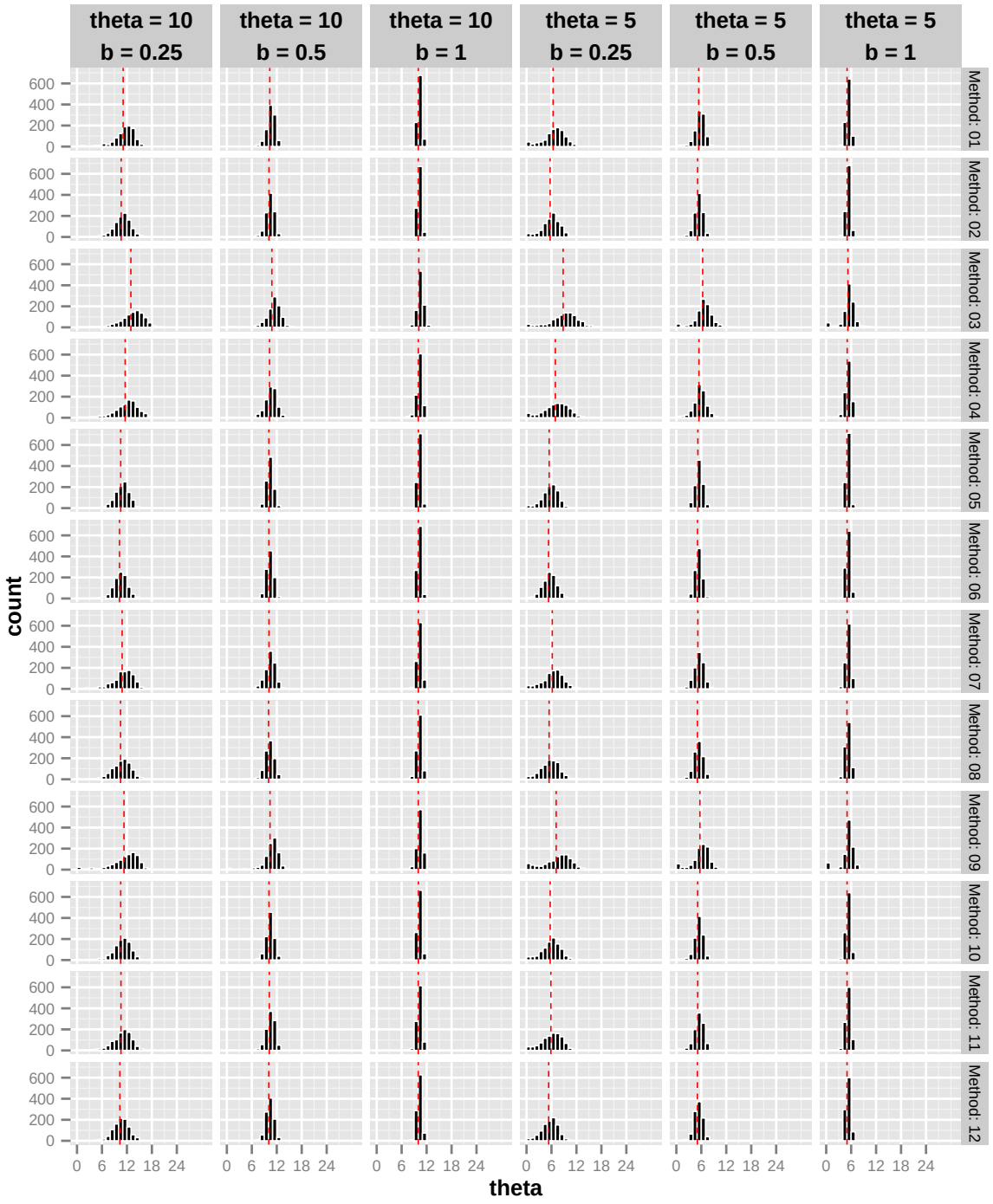


Figure 32: Histograms of estimated θs in Simulation 2 from runs of 100 trials. Each histogram includes 1000 simulated runs.

Figure 33: Histograms of estimated bs in Simulation 2 from runs of 100 trials. Each histogram includes 1000 simulated runs.