Testing the Perception of Auditory Popout in Relation to Timbre

John (Jack) Tokarz

260456166

Submitted for PSYC 494

April, 2015

Author Note

John Tokarz, Department of Psychology, McGill University

Research conducted under supervision by Dr. Stephen McAdams, supervisor, Schulich School of Music, associate to Department of Psychology, McGill University

Correspondence regarding this paper should be sent to:

John Tokarz at john.tokarz@mail.mcgill.ca

Abstract

Using a feature-integrated model of vision, researchers have demonstrated the existence of a popout phenomenon, which automatically attracts the eye to an object which is different in one feature from an array of identical objects. Using an experimental design based on these studies, but translated to investigate auditory instead of visual stimuli, this paper primarily aimed to prove the existence of an auditory popout effect. The features manipulated in this study were three dimensions of timbre (attack time, spectral centroid and the attenuation of even harmonics). Additionally, this study investigated the existence of direction salience in timbre, through its influence on the popout effect. The study found substantial evidence for a popout effect existing in the auditory domain, but weak evidence for direction salience in any of the three timbre dimensions.

Introduction

Defining popout

In a crowded street you hear many sounds: the words of multiple pedestrians' conversations, the hum of a car's motor and the scrape of a shovel clearing a driveway. Although all of these sounds are heard simultaneously, they are understood to be separate. In addition, each of these individual sounds can be broken down; they have distinct features that mesh to create the acoustic pattern.

Similarly, vision can be interpreted as a modular sense, simultaneously perceiving and interpreting many features separately, such as color, orientation, size, shape, and more, and then combining them to create our visual field (Treisman, 1988). Despite having this array of dimensions to vigilantly monitor, humans' ability to recognize an abnormality in any one dimension remains sharp.

When one feature of an object differs to a nontrivial degree from that of other homogeneous objects in the visual field, people automatically register this object as different from the others. This is called "the popout phenomenon" (Berti & Schröger, 2001). This phenomenon occurs spontaneously and immediately in at least 5 dimensions of vision: shape and color (Treisman & Gelade, 1980), and orientation, size, and location (Ahissar & Hochstein, 2005), drawing one's attention to this individual item, whether it is sought or a distractor.

Testing popout

Experiments have been conducted to understand the cause and determinants of this visual effect in one or more dimensions of vision (Ahissar & Hochstein, 1996; Maljkovic & Nakayama, 1996; Treisman, 1998; Treisman & Gelade, 1980). In these experiments, participants were presented with a visual array of objects, and asked to press a button to indicate whether or not one object "pops out" from the others. Response time and accuracy were recorded along changes in the salient feature of the popout object. Universally, the effect has been found to exist, each dimension of vision having an ideal salience value which minimizes reaction times, and this value is quite small. For example, for stimulus orientation at any angle difference greater than 15 degrees from distractors gave a nonsignificant decrease in response time relative to a 15 degree difference (Ahissar & Hochstein, 1996), meaning 15 degrees is an ideal salience value because any difference greater than this does not enhance the popout effect.

To test the spontaneous expression of this effect, Treisman (1988) conducted a series of experiments based on the design detailed above, but with modifications to the stimuli, distractors or instructions. In one experiment, she separated participants into two groups, one of which were told which feature would be salient, the other was not. For example, "blue" would be written on the screen prior to presentation of a screen with one blue object among many white distractors for the informed group. She did not find a significant difference in reaction time between the informed and uninformed groups. A difference between these groups was found, however, in subsequent studies involving conditions with homogeneity or heterogeneity of an irrelevant feature in the distractor objects. For example, the popout object may be blue and vertical, whereas all distractors are white but randomly chosen to be vertical or horizontal. This heterogeneity of distractors increased the unaware participant group's response time by an average of 90 ms—a significant increase. A further study found that distractors that were heterogeneous in the relevant feature—for example, the popout was a single blue object, whereas the distractors were either red, green or white, with multiple instances of each color—gave a uniform increase in response latency by an average of 250 ms relative to homogeneous distractors. This difference existed independently of informed or informed group membership (Triesman, 1988). This finding indicates that although prior knowledge of the visual dimension relevant to popout perception may give a slight advantage in certain circumstances, it generally does not aid the perceiver to a large extent, meaning that the popout effect does not rely on executive knowledge, but occurs automatically as an integrated aspect of visual perception.

Direction Salience in Popout

The popout effect may not be equivalent for a change in one feature in both directions, as indicated by a study that intended to eliminate the learning effect inherent in earlier popout study models (Ahissar & Hochstein, 1996). A learning effect occurred in earlier, repetitive popout experiments, which decreased response time after the same stimuli were presented multiple times (Ahissar & Hochstein, 1996; Maljkovic & Nakayama, 1996; Triesman, 1988). This learning effect also occurred for similar features of popout objects: completing trials with a popout object smaller than the distractors decreased reaction time for subsequent trials with a larger popout object (Ahissar & Hochstein, 1996). However, this specific effect did not occur in the inverse condition: trials with a larger popout object did not decrease reaction time in subsequent trials with a smaller popout object (Ahissar & Hochstein, 1996), indicating that certain features can have asymmetric popout effects for inverse value changes.

Auditory Popout?

Research on the popout phenomenon focuses solely on its role in visual perception, but its model could fit auditory perception as well and theoretically produce a similar effect. The popout phenomenon is generally accepted to be derived from the novelty seeking behavior of sensation; when something is different, it is given more importance, so it attracts attention. This attention allocation is known to occur in both visual and auditory perception (Berti & Schröger, 2001). However, the popout effect cannot automatically be equated between the visual and auditory domains, because the existing visual research relies on parallel processing of an entire field of objects, whereas an auditory popout object would have to occur serially in an array of distractors.

Studies have assigned popout qualities to auditory stimuli, then measured reaction time in a serial perception task (Berti & Schröger, 2001, Parmentier, 2014); but these studies did not label the effect as “auditory popout”, they instead called the outlier an “oddball sound” or a “deviant stimulus”. These names were chosen because the stimuli were used to distract the participant from a task that required attention. Berti & Schröger (2001) investigated if a tone of surprising pitch or timing could distract participants and increase their reaction time for identifying the length of a visual stimulus. Ignoring semantic disagreements, the definition of these terms matches that of a popout stimulus, and their research shares common results with popout studies. A change in a salient feature of one auditory or visual stimulus grabs the participant’s attention and increases their reaction time (Berti & Schröger, 2001, Parmentier, 2014).

Research on deviant stimuli typically uses a hybrid of visual and auditory stimuli to increase or distract participant’s attention to relevant stimuli for feature perception. For example, in one study participants were asked to indicate whether a visual bar was short or long while tones were played in a pattern that either did or did not include a deviant tone. Or, to investigate whether the effect occurs in both directions, there were asked to indicate whether a tone was short or long while bars of different sizes were presented on a screen (Parmentier, 2014). A study comparing the distracting effects of tones that could deviate in either their onset time or pitch found that an unexpected change in onset time caused a greater distraction (increase in response time) than a pitch change (Parmentier, 2014). As in popout studies, these results revealed that a stimulus with a salient feature automatically attracted participant attention, even an auditory stimulus.

Integrating Timbre

Even though the array of objects to perceive must be presented serially, the determinants of a visual object that causes the popout phenomenon can be replicated in an auditory object due to its many features. Auditory stimuli consist of several features that must simultaneously be processed and then integrated to be perceived. Past studies of tones' ability to attract attention automatically investigated pitch and duration, but auditory perception also includes the dimensions of timbre. Three dimensions of timbre which vary independently in perception are attack time (ATT), spectral centroid (SCG), and the spectrum fine structure: the selective attenuation of even harmonics relative to odd (EHA). A perceived distance in any one of these dimensions is unaffected by the values of the other two, pitch, loudness, or perceived duration (Caclin, McAdams, Smith & Winsburg, 2005). Therefore a change in one of these dimensions in a series of tones equated for all other factors matches the criteria for a stimulus emulating the popout phenomenon.

Direction Salience in Timbre

Within sound, the salience of a specific feature changes with the feature's value, making that feature attract attention more or less in the overall sound space. For example, high values in pitch tend to have high salience, so musicians typically give melodies to instruments with higher natural frequencies, making them easier for the listener to follow (Chon, 2013). This asymmetry of direction salience has been found in the auditory features of loudness and pitch: louder and higher values, respectively, are more salient.

Within the dimensions of timbre, direction salience research remains inconclusive. In an auditory stream recognition task (Chon, 2013) based on a pre-existing model (Bey &McAdams, 2003), two instruments played the same pitch in alternation, and participants were asked to tap along with one (Experiment 1) or a pair of sounds was played and listeners identified which “grabbed their attention better” (Experiment 2). These experiments gave conflicting results, which did not match those of Bey & McAdams (2003). Context provided a large influence on instrument salience. In Experiment 1, listeners tended to tap based on different dimensions of timbre than those used in difference perception tests (ATT, SCG, EHA) (Caclin et al., 2005). EHA was the only overlapping timbre dimension between the perception test and this salience test. As the attenuation of the even harmonics decreased, making the sound less hollow, it became more salient. In Experiment 2, certain timbres were found to be more salient when presented first, whereas others were more salient when presented last, for an unknown reason (Chon, 2013, Chapter 2). These inconclusive results indicate that there is a direction saliency within timbre, but its relevant dimensions and their respective directions are uncertain.

For this study the “down” direction refers to the trials is which the popout tone’s salient feature has a lower value than the baseline, and “up” direction refers to the trials in which the popout’s salient feature value is greater than the baselines’s. Specifically for ATT, the up direction trials assign the popout a comparatively longer attack— perceived as a softer attack, and its down direction trials assign a relatively shorter attack— perceived as a harder attack. For EHA, the up trials feature a popout with greater even harmonic attenuation than the baseline— perceived as a hollower sound, whereas the down trials’ popout features relatively less even harmonic attenuation— perceived as a less hollow sound. For SCG, the popout tone in the up direction trials has a higher centroid than the baseline tones— giving it a brighter sound, whereas the popout in the down direction trials has a lower centroid— perceived as a darker or mellower sound.

Outline

The current study intended to explicitly prove the existence of the auditory popout phenomenon through an experiment re-interpreted from those originally used by Treisman & Gelade (1980), but applied to a serial presentation of auditory stimuli. This experiment explored the application of this phenomenon within each of the three aforementioned dimensions of timbre: ATT, SCG and EHA. It investigated four size differences between the popout and distractor tone's values along the salient dimension. It also hoped to determine how sensitive individuals' perception is to changes in these dimensions, by determining whether there is a certain size difference for which all larger size differences do not yield significantly improved reactions (similar to how Ahissar & Hochstein, 1996, found that differences in orientation of more than 15 degrees failed to significantly decrease reaction time).

Additionally, the experiment tested direction salience of the popout effect within each dimension, with trials featuring a low baseline and high popout in the relevant dimension as well as a baseline with the same high value and popout with the same low. This aimed to clarify whether these three dimensions have direction salience, under the hypothesis that a more salient timbre direction would grab the listeners' attention more, reducing average reaction time more than trials of the inverse direction. Finally, the experiment investigated the effect of heterogeneous distractors in an irrelevant feature by including trials in which the pitch of each tone is randomly selected, but only one tone varies in the popout-relevant timbral dimension.

The serial presentation of stimuli should not alter the popout effect, although the distractors are not processed in parallel; a basic working memory can be used to compare a tone to its predecessor, revealing any obvious changes in value of any dimension. Due to the serial presentation of stimuli, reaction time is not measured relative to the beginning of array presentation as in visual studies, but from the onset of the popout tone: the moment the participant should feasibly realize a difference.

I hypothesized that for each dimension, and in both directions, as the size difference between popout and baseline value increases the reaction time (RT) should decrease and the number of correct responses (NC) should increase. It was difficult to hypothesize the results of direction salience for the three dimensions due to inconclusive previous research. However, I predicted that less attenuation would have greater salience than more in the EHA trials, as Chon (2013) found. I predicted ATT and SCG would have a direction salience, but I could not know in which direction. Finally, I hypothesized that in the trials with pitch randomization, RT would be universally larger and NC would be universally smaller for all dimensions and pitch conditions.

Methods

Participants

The data were collected from 30 participants ranging from 18 to 48 years old, with a mean age of 24 (20 females and 10 males). This information was collected in a survey completed anonymously prior to the experiment. Additionally, all participants signed consent forms prior to participation and passed a pure-tone audiometric test using a MAICO MA 39 (MAICO Diagnostic GmbH, Berlin, Germany) audiometer at octave-spaced frequencies from 125 Hz to 8 kHz (ISO 389–8, 2004; Martin & Champlin, 2000) and were required to have thresholds at or below 20 dB HL to proceed. The participants volunteered to participate in exchange for $10 CAD. All participants signed an informed consent form, and the study was certified by the McGill Research Ethics Board II (certificate 67-0905).

Stimuli

Stimuli were constructed based on previous psychophysical research that identified three perceptually salient timbre dimensions (Caclin et al., 2005). Using multidimensional scaling, these authors showed that the logarithm of attack time (ATT), spectral center of gravity (SCG), and spectrum fine structure (EHA; selective attenuation of even harmonics relative to odd) represent three orthogonal dimensions of timbre space. These dimensions presented an ideal method of creating timbral stimuli in a systematic and nuanced fashion, while building on timbre research in typical populations.

The stimulus for each trial was a 14-tone sequence with an interonset interval of 450 ms synthesized using Pd (http://puredata.info/, accessed 23 March 2015) on a Macintosh Pro workstation (Apple Computer, Cupertino, CA) at 44.1 kHz sampling rate and 16-bit amplitude resolution. Synthetic tones were used to ensure complete control of stimulus parameters. Depending on the condition, the pitch of the individual tones was either set to C4 or chosen randomly from the set of eight quarter tones starting with C4 (fundamental frequency $f_0 \in \{261.62, 269.29, 277.18, 285.30, 293.66, 302.27, 311.12, 320.24\}$ Hz), except for the popout tone, whose pitch was restricted to the six middle frequencies (269.29, 277.18, 285.30, 293.66, 302.27 or 311.12 Hz).

Timbres varied as a function of three parameters: attack time (ATT), spectral center of gravity (SCG) and even-harmonic attenuation (EHA). Each tone was a steady-state signal multiplied by a time-varying amplitude envelope. The steady state signals were harmonic tone complexes, s(t), with 40 harmonics:

$$s(t) = L_c A_{scale} \sum_{k=1}^{40} A_k \cos(2\pi k f_0 t),$$

where L_c was a loudness compensation factor, and A_k was the amplitude of harmonic k:

$$A_k = h_k / k^{\alpha},$$

A_{scale} was a scaling factor for constant signal energy, $h_k = 1$ when k was odd, $h_k = h_{even}$ when k was even, and α was chosen to satisfy the equation:

$$SCG = \frac{\sum_k k A_k}{\sum_k A_k}$$

which controlled the SCG by changing the slope of a linear spectral envelope of log amplitude as a function of linear frequency. SCG values represented the amplitude-weighted average frequency of all harmonics in a given tone, expressed in harmonic rank. h_{even} was calculated from the Even Harmonic Attenuation, $h_{even} = 10^{EHA/20}$. A_{scale} was calculated to make the signal's energy before loudness compensation the same for all sounds:

$$A_{scale} = \frac{1}{\sqrt{\sum_{k=1}^{40} {A_k}^2}}$$

The perceived loudness was influenced by the spectral center of gravity. Therefore, a loudness matching procedure was used to calculate the loudness compensation L_c, giving the formulæ:

$$L_{db} = -6.35187 \log_e SCG + 4.75832$$

$$L_c = 10^{L_{db}/20}$$

The time-varying amplitude envelope consisted of a linear attack of duration *AT*, a sustain portion at constant amplitude (*SUS*), and an exponential decay at a rate of –60 dB every 50 ms. *SUS* was adjusted as a function of *AT* to keep the perceived sound duration constant:

$$SUS = 210 - 0.8 * AT \ (in\ ms).$$

The perceived start time of the individual tones was influenced by the attack time (a tone with slower attack was perceived to start later than a tone with faster attack). In order to keep the sequences perceptually isochronous, the tones were presented with a small deviation from onset isochrony. To calculate this value, the effect of attack time on perceived tone start time was measured in a pre-experiment, giving a formula for the deviation from isochrony of:

$$23.23 - 23.23\ log_{10}AT.$$

For a given sequence, all tones had the same timbre, which we will call the background timbre, except for one tone, called the popout tone. The two timbres were chosen to be perceptually equidistant from a baseline timbre by modifying one of the three timbral parameters, depending on the condition. The values of each popout tone's salient feature is detailed by condition in Table 1. The parameters for the baseline timbre were:

SCG = 3.24, EHA = –9.0 dB, and AT = 59.2 ms.

Table 1.

Popout and Baseline Values by Timbre Dimension and Condition

	Condition	Low_4	Low_3	Low_2	Low_1	High_1	High_2	High_3	High_4
Block									
Monotone									
SCG		2.50	2.67	2.85	3.04	3.46	3.68	3.94	4.20
EHA		-3.6	-4.5	-5.4	-6.3	-11.7	-12.6	-13.5	-14.4
ATT		31.8	35.7	39.2	43.1	81.2	89.2	98.1	110.2
Variable									
SCG		1.90	2.17	2.48	2.84	3.70	4.23	4.83	5.53
EHA		0.0	-1.4	-2.8	-4.2	-13.8	15.2	-16.6	-18.0
ATT		26.9	30.2	33.8	37.5	93.3	103.7	115.9	130.2

Note. Values of each timbre dimension. Condition refers to: whether the popout's salient feature value is greater or less than the baseline's, used for Direction (Low, High), and the Size of the difference between the popout and baseline (1-4). So in every series the popout and baseline notes have the same size value for the salient feature, but one is low and the other is high. Block refers to whether the tones have monotone or variable pitch and the variable timbre dimension (SCG in harmonic rank, EHA in dB, ATT in ms). For example, the monotone SCG block series for Direction up, Size 3 will have 13 notes with the "Low_3" feature value, and one with the "High_3" feature value. Their baseline values when not relevant to the current block were SCG = 3.24, EHA = –9.0, and AT = 59.2.

Experimental Design

The experiment consisted of six blocks of 64 series of 14 tones. Each series consisted of 13 tones with identical timbre (the baseline) and one tone with a higher or lower value of the salient timbre (the popout). The blocks were separated into two groups of three, a group for each dimension of timbre (ATT, SCG, EHM) with either monotone or variable pitch (M or V) making six blocks in total (ATM, ATV, EHM, EHV, SCM, SCV). The tones were played through Sennheiser HD280 Pro headphones (Sennheiser Electronic GmbH, Wedemark, Germany). Sound levels were kept between 55- 61 dBA SPL, as measured by a B&K 2250 sound level meter with an artificial ear coupler. For each participant the ordering of monotone and random pitch groups of blocks was randomly determined, as well as the dimensions' ordering within each group. The four size distance values between popout and baseline values were swapped, with the baseline as either the lower or upper value, to test direction salience, resulting in eight total conditions per block. The eight conditions were repeated eight times, making 64 total series for each of the six blocks. Within each block the conditions were presented in a random order. These random orderings prevented the possibility of a learning effect to change reactions over the course of the experiment (Ahissar & Hochstein, 1996).

In each block, reaction time and number of correct responses were recorded. RT measured the speed at which participants registered the popout tone as distinctly different from the others, measured in ms as the time from the popout tone's onset to the spacebar press. NC measured participant accuracy in popout identification. If the participant pressed the spacebar within 1,800 ms (the duration of four successive tones in the sequence), the trial passed the threshold for being "correct". NC represented how many trials per block were correct.

Procedure

Participants completed the entirety of the experiment inside an Industrial Acoustics model 120-act3 double-walled sound-isolation booth (IAC Acoustics, Bronx, NY). After successfully passing the audiometric exam and completing the background information survey, participants read through the instructions, which explained how to perform the reaction time task; pressing the 's' key when they wanted the series to begin, and the spacebar key the moment they heard the popout tone. The program gave no indication of correctness and played the entire series regardless of when the participant hit the spacebar. Every participant completed all 384 series at their own pace; they could wait as long as they desired before beginning each trial. After reading the instructions, the participant and the experimenter discussed the procedure to ensure they understood the task.

They then listened to six tones, in three groups of two, the lowest and highest examples of each timbre dimension, to familiarize themselves with the relevant timbre dimensions. These tones could be repeated as many times as desired by the participant to understand the change in sound quality between contrasting tones. After familiarization, they began the practice trials: two series of the highest Size difference (4) for the relevant block in each Direction. If they did not press the spacebar or seemed confused about any aspect of the experiment, the experimenter offered assistance. Afterwards the experimenter left the booth, and the participant proceeded through the six blocks without further assistance. In between each block the participant completed the practice trials with the new dimension of timbre for the upcoming block. After completion of all six blocks, they were given a debriefing form informing them of the details of the experiment, they signed a receipt and were given their money.

Results

For each of the six blocks, we performed a repeated measures ANOVA to investigate the effect of Size (small - 1, medium low - 2, medium large - 3 and large - 4), Direction (up and down), and Direction × Size for the average response time (RT) and number correct (NC). Sphericity violations were determined using Mauchly's test. The Greenhouse-Geisser ε was used to correct the degrees of freedom for values of $\varepsilon < .75$. For effects with a Greenhouse-Geisser $\varepsilon > .75$, we used the Huynh-Feldt ε. Corrected degrees of freedom and *p*-value are reported. Afterwards, we performed a normality test on the residuals of an analysis of variance, using the $\log_{10}$(RT) for RT results. For the blocks with significant normality violation, we performed corresponding nonparametric tests to confirm the finding. If they had a significant effect for Direction (2 levels), we performed a Wilcoxon Signed Rank test. If they had a significant effect for Size (4 levels), we performed a Friedman ANOVA. For the block with a significant effect for Direction × Size, we performed a Wilcoxon Signed Rank test comparing the two directions for each size difference.

For the average response time, several participants' data could not be used because they did not react quickly enough for any of the 8 repetitions of one or more trials, meaning they had no data for one or more of the conditions. The largest number of participants dropped was from the block with attack time and variable pitch, with an N of 22 out of 30. In the monotone attack time, monotone spectral centroid and variable spectral centroid blocks, N= 25. The rest of the blocks' analyses used all 30 participants' responses. All analyses of the number correct used all 30 participants' responses.

Table 2.

Analysis Results

				N	df	F	P	ε	η^2_p	nonpara. P
AT	Mono	RT		25						
			Direction		1,24	1.22	0.28	-	0.048	-
			Size		2.06, 49.34	9.19	<.001	0.685	0.277	0.002[B]
			Dir*Size		2.83, 67.93	<1	-	-	-	-
		NC		30						
			Direction		1, 29	<1	-	-	-	-
			Size		2.46, 71.42	28.45	<.001	0.821	0.495	<.001[B]
			Dir*Size		3, 87	1.61	0.19	-	0.052	-
	Var	RT		22						
			Direction		1, 21	<1	-	-	-	-
			Size		2, 63	5.15	0.003	-	0.197	-
			Dir*Size		2.71, 56.95	<1	-	-	-	-
		NC		30						
			Direction		1, 29	<1	-	-	-	-
			Size		3, 87	14.84	<.001	-	0.338	<.001[B]
			Dir*Size		3, 87	1.82	0.15	-	0.059	-
EH	Mono	RT		30						
			Direction		1, 29	2.36	0.14	-	0.075	-
			Size		3, 87	16.14	<.001	-	0.358	<.001[B]
			Dir*Size		3, 87	3.33	0.023	-	0.103	varied[C]
		NC		30						
			Direction		1, 29	5.96	0.021	-	0.171	0.015[A]
			Size		3, 87	4.46	0.006	-	0.133	0.002[B]
			Dir*Size		2.21, 63.96	1.58	0.21	0.735	0.052	-
	Var	RT		30						
			Direction		1, 29	<1	-	-	-	-
			Size		3, 87	5.51	0.002	-	0.16	<.001[B]
			Dir*Size		1.81, 52.56	<1	-	-	-	-
		NC		30						
			Direction		1, 29	<1	-	-	-	-
			Size		3, 87	6.24	0.001	-	0.177	.003[B]
			Dir*Size		3, 87	<1	-	-	-	-
SC	Mono	RT		25						
			Direction		1, 24	<1	-	-	-	-
			Size		1.28, 30.81	58.15	<.001	0.428	0.708	<.001
			Dir*Size		1.45, 34.87	<1	-	-	-	-
		NC		30						
			Direction		1, 29	10.43	0.003	-	0.265	.003[A]
			Size		1.67, 48.53	110.11	<.001	0.558	0.792	<.001[B]
			Dir*Size		1.95, 56.59	2.62	0.083	0.651	0.083	-
	Var	RT		25						
			Direction		1, 24	1.47	0.237	-	0.058	-
			Size		1.53, 36.69	26.5	<.001	0.51	0.525	<.001[B]
			Dir*Size		1.90, 45.53	<1	-	0.632	-	-
		NC		30						
			Direction		1, 29	29.48	<.001	-	0.504	<.001[A]
			Size		1.80, 52.07	131.99	<.001	0.598	0.82	<.001[B]
			Dir*Size		2.21, 64.12	1.46	0.24	0.737	0.048	-

Note. The N, df, *F*, *p*, ε , η_p^2, and nonparametric test *p*-values for the reaction time and number correct of the monotone and variable pitch blocks of spectral centroid, attack time and attenuation of even harmonics conditions. In the nonparametric *p*-value column (nonpara. *p*), the test performed is indicated by A for Wilcoxon Signed Rank, B for Friedman, or C for Wilcoxon Signed Rank for each Size.

Attack Time

For attack time, an increase in Size had a significant decrease in RT and increase in NC for the monotone and variable pitch blocks. The details of these results can be seen in Table 2. Although the NC had four breaches of normality in ATM and three in ATV, both retained significance in their respective Friedman ANOVAs. The mean values of ATM can be found in Figure 1 and those of ATV in Figure 2. In all attack-time-related analyses, the effects of Direction and Direction × Size were not significant, suggesting that there is no salience effect related to attack time. Although the data for ATM and ATV cannot be directly compared due to their different values of attack time, examination of Figure 3 indicates that they do not strongly differ in their relative RT. One can see that the mean values are near each other and the 95% confidence intervals overlap in the area where the change in attack time for both pitch conditions overlap. To the contrary, examination of Figure 4 indicates that ATV has a lower response accuracy (non-overlapping confidence intervals) than ATM for the stimuli with a similar change in attack time, suggesting that pitch variability affects accuracy, but not response time.

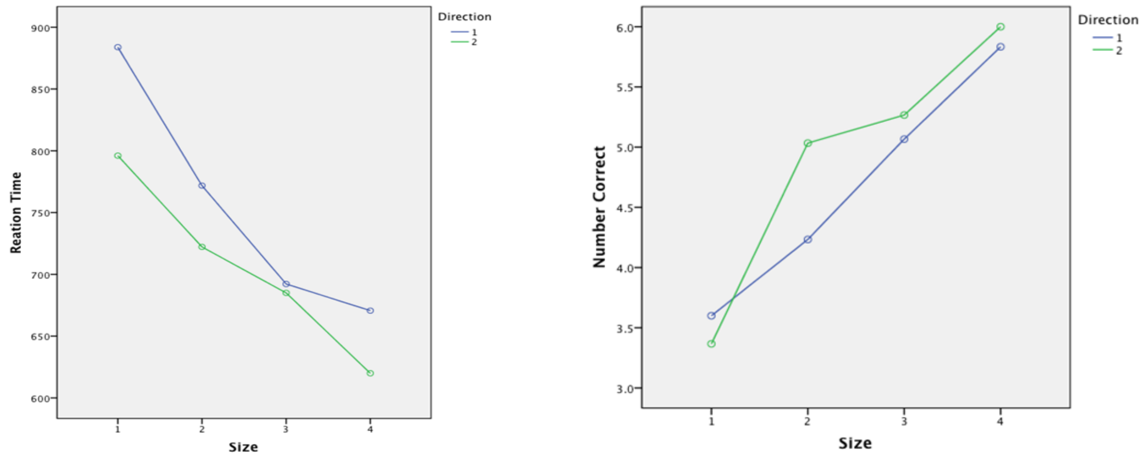

Figure 1. Average reaction time (left) and number correct (right) for the attack time condition with monotone pitch.

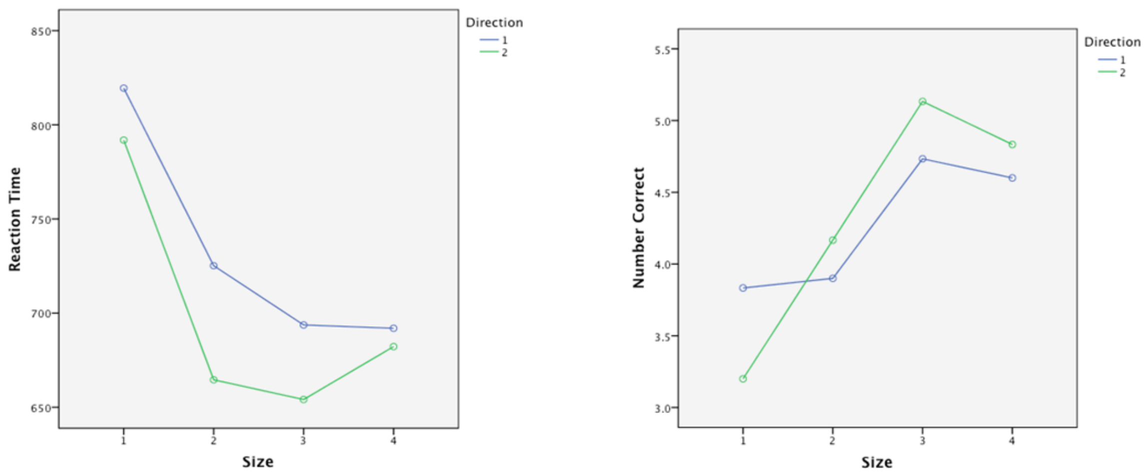

Figure 2. Average reaction time (left) and number correct (right) for the attack time condition with variable pitch.

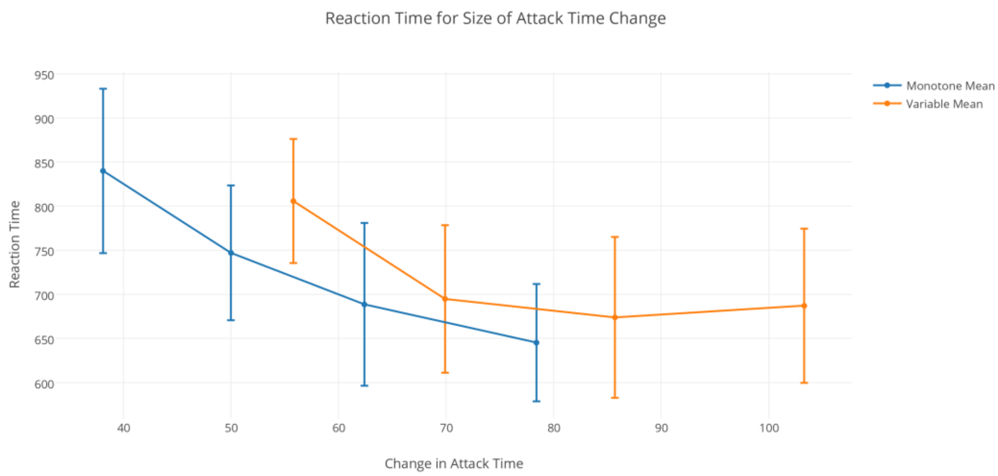

Figure 3. Average reaction time with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the attack time condition.

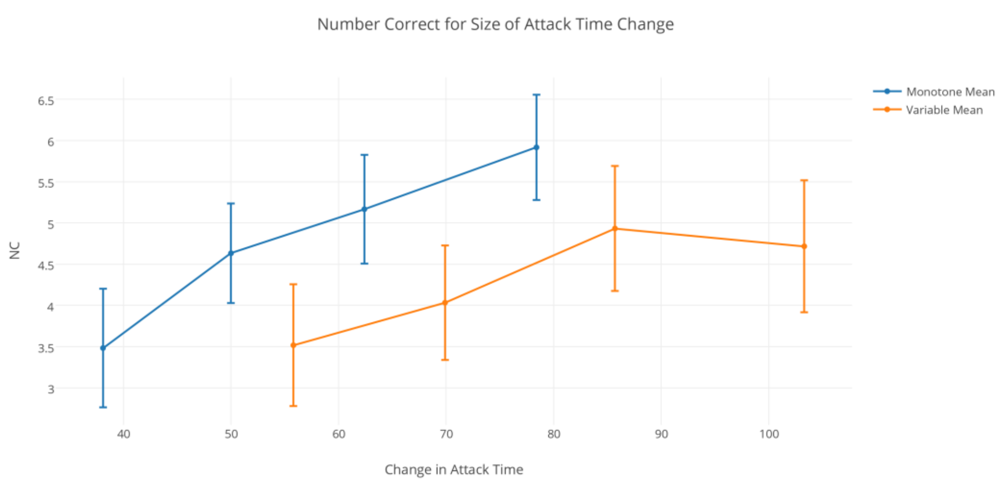

Figure 4. Average number correct with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the attack time condition.

Attenuation of Even Harmonics

An increase in Size difference of the attenuation of even harmonics caused a significant decrease in RT and a significant increase in NC for both pitch conditions (see Fig. 5 for EHM and Fig. 6 for EHV). Additionally, in the EHM condition changing from more attenuation to less attenuation elicited significantly lower values of NC than an equivalent change of less to more attenuation. For the Direction × Size interaction in EHM, there was a nonmonotonic effect on RT (see Figure 5, left panel) for going from more to less attenuation. For this interaction we used Wilcoxon Signed Rank tests on Direction for each Size. Using a Bonferroni-corrected $\alpha = .0125$ for the four sizes of timbre change, the results are: Size 1: $p = .010$, Size 2: $p = .62$, Size 3: $p = .83$, and Size 4: $p = .015$. This indicated that the significance of the interaction came principally from Size 1, as it had the only significant difference, and partially from Size 4 which had a marginally significant difference. The significant Size and Direction effects were confirmed with their respective nonparametric tests. Direction did not cause a significant change in RT, nor was there a significant effect of the Direction × Size interaction on NC. The change from more to less and less to more attenuation elicited the same RT, despite its effect on NC. Even though the change from more to less attenuation did have an effect on the interaction between Direction and Size for RT, it did not affect NC. Examination of Figures 7 and 8 indicate asymmetric relationships between the pitch constancy conditions for RT and NC, respectively. Figure 7 implies that EHV has a universally greater RT value than EHM for similar changes in attenuation. Figure 8, however, shows universally lower NC values for similar changes in EHA in the variable and monotone pitch conditions.

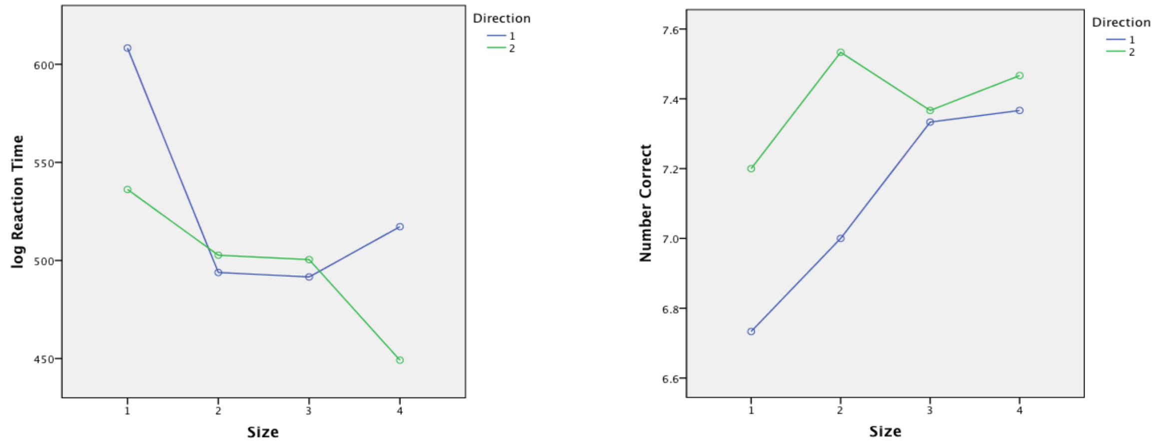

Figure 5. Average reaction time (left) and number correct (right) for the even-harmonic attenuation condition with monotone pitch.

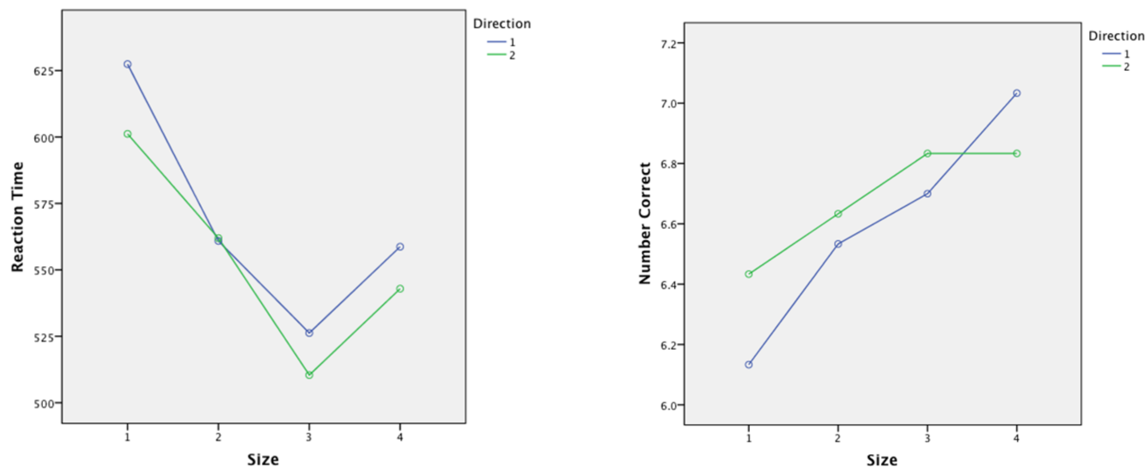

Figure 6. Average reaction time (left) and number correct (right) for the even-harmonic attenuation condition with variable pitch.

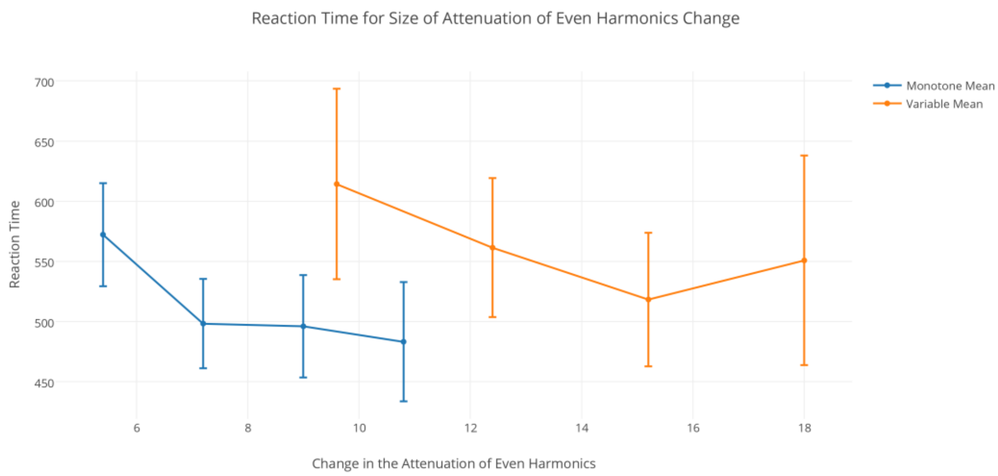

Figure 7. Average reaction time with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the attenuation of even harmonics condition.

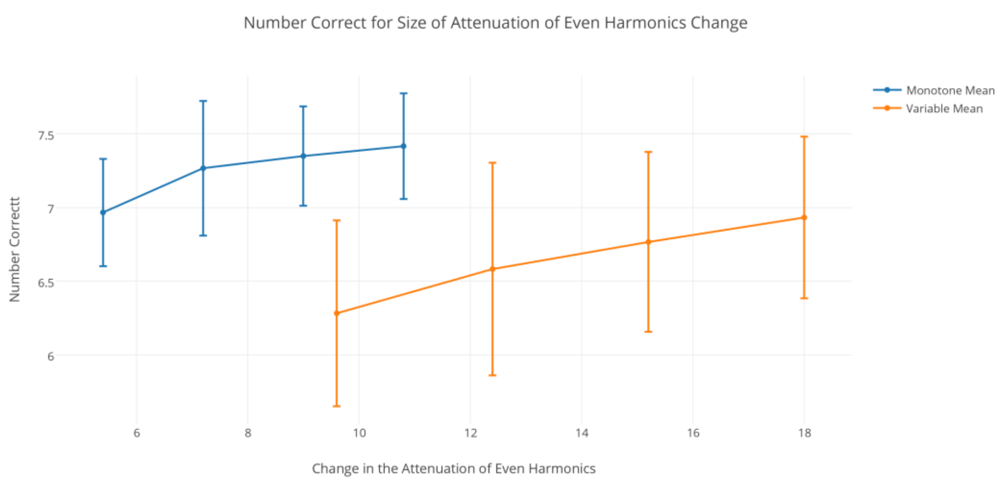

Figure 8. Average number correct with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the attenuation of even harmonics condition.

Spectral Center of Gravity

Following the same pattern as the other two timbre dimensions, an increase in Size difference of the SCG caused a significant decrease in RT and a significant increase in NC for both pitch conditions (see Fig. 9 for SCM and Fig. 10 for SCV). The significant effect of Direction reveals that an increase in SCG elicited significantly higher values of NC than did a decrease in SCG for both SCM and SCV conditions. Or, the relatively brighter popout tones were more accurately detected than the darker popout tones. All of these effects were confirmed with their respective nonparametric tests. Figures 11 and 12 indicate a similar relationship between the RT and NC of the monotone and variable pitch conditions in SCG to those in EHA. Examination of Figure 11 shows that SCV has a slower reaction times overall than SCM for similar values of change in SCG. Figure 12, however, shows globally lower accuracy in the variable pitch condition compared to the monotone condition for similar changes in SCG.

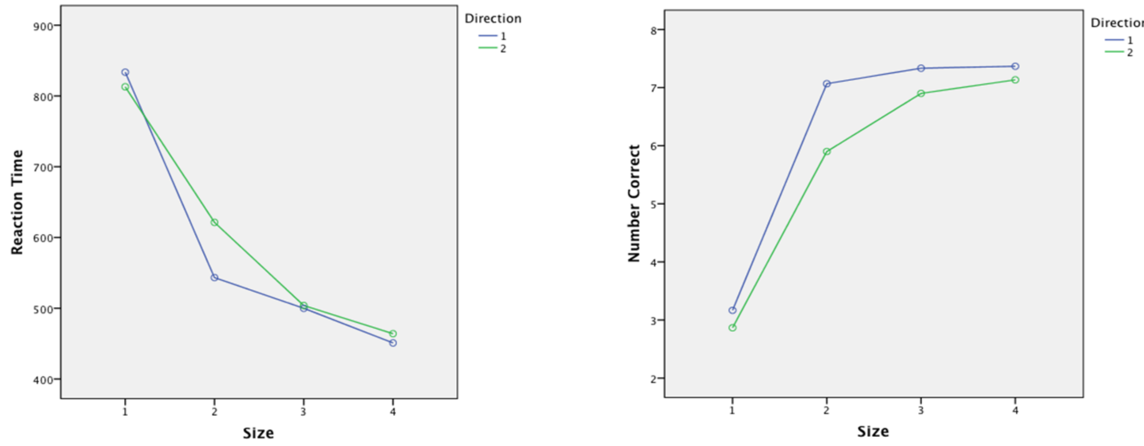

Figure 9. Average reaction time (left) and number correct (right) for the spectral centroid condition with monotone pitch.

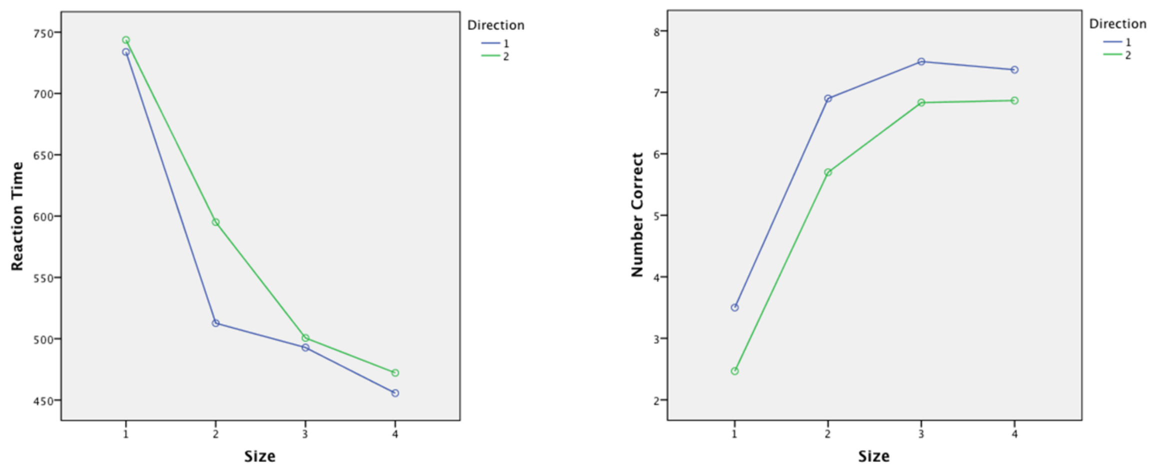

Figure 10. Average reaction time (left) and number correct (right) for the spectral centroid condition with variable pitch.

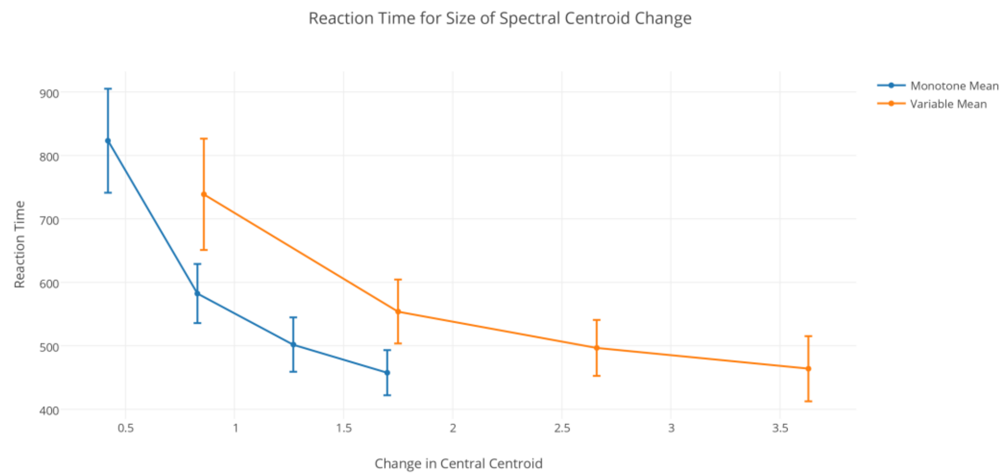

Figure 11. Average number correct with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the spectral centroid condition.

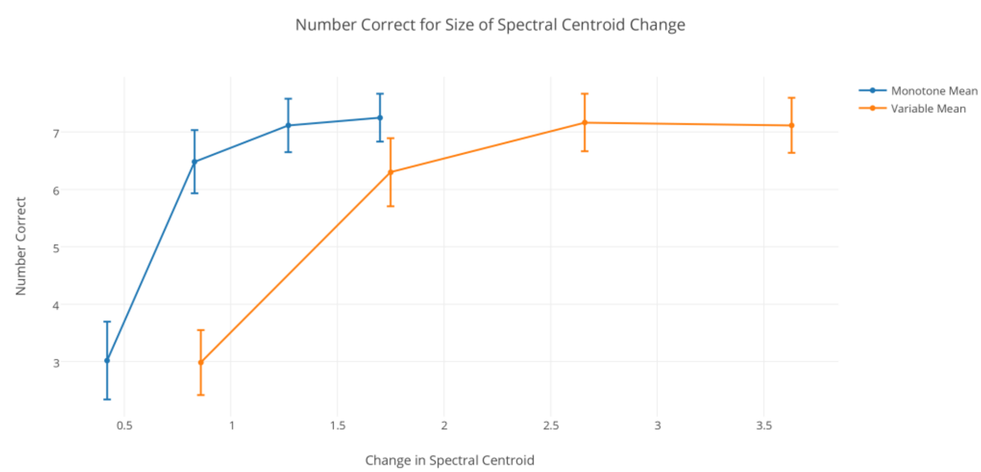

Figure 12. Average number correct with 95% confidence interval for the four size differences in the monotone (blue) and variable (orange) pitch sequences of the spectral centroid condition.

Discussion

This experiment supports the existence of the popout effect as an auditory phenomenon relative to the timbre dimensions attack time (ATT), spectral center of gravity (SCG), and attenuation of even harmonics (EHA). Given instructions to indicate when a tone sounded different from the others, participants routinely selected the popout tone among a series of identical tones, even among the series of tones with random pitch variation. As predicted, when the size of timbre difference increased, the speed and accuracy of responses improved (a decrease in RT and an increase in NC) in all blocks. This means that a larger difference caused the popout to stand out more from the others, bringing it more immediately to attention for the listener, or at least making it easier to detect. This result matches previous findings that auditory stimuli that vary from their surroundings demand the listener's attention (Berti & Schröger, 2001, Parmentier, 2014).

The results concerning direction of change, which would be interpreted as indicating timbral salience, were once again conflicting, giving only weak evidence that would suggest the existence of timbre salience for EHM, SCM and SCV. When significant, these effects were not consistent within all results of a block. We expected that RT and NC results would be correlated, as found in Size, but in Direction and Direction × Size this correlation did not exist for EHA or SCG. A significant effect in NC without RT— as seen in Direction for EHM, SCM and SCV, means that whether participants responded correctly or not did not affect their speed; the saliency caused more popout tones to be heard, but the reaction to these tones on the cusp of detection elicited the same speed of reaction as the others. The significance of Size differences across Direction in RT but not NC in EHM, showed that the popout tones were not easier to detect by Direction for each Size, but the speed of detection did change by Direction for each Size.

Although these effects do not follow any obvious pattern, they all retained significance even when tested with the nonparametric Wilcoxon Signed Rank test. There is evidence for a stronger salience of more attenuation than less attenuation for EHA, and a stronger salience of a higher centroid than a lower one for SCG. These results are weak, however, because only two out of four tests for each gave significant results.

Attack time did not give any indication of a salience for longer or shorter attack. This lack of significant results may not be due to the timbre dimension ATT itself, but the particular stimuli used in this experiment. Participants routinely reported that ATM or ATV were the hardest to accurately complete, and that they rarely heard a difference between the popout and baseline in these trials. An increase in the difference between the longer and shorter ATT values could compensate for this difficulty. However, this would require the tone duration to increase. Tone duration was set at 260 ms to keep trials short and engaging. Because the original model of popout perception is of an array presented simultaneously, we intended to present the series quickly. We did not want to stray from the original popout model more than was necessary. For this tone duration selection, the attack time could not be set any higher than it was (130.2 ms for Size 4 in ATV) without the structure of the tone changing, which would create the possibility of popout detection by an auditory feature other than ATT (due to an inconsistent tone onset isochrony). Therefore, making the ATT blocks easier would not be as simple as for the other timbre dimensions. Alternatively, perhaps changes in this timbre dimension simply do not stand out to the same degree as changes in the other dimensions, a certain explanation cannot be made from these data.

Our final hypothesis, that RT would be universally higher and NC would be universally lower for the variable pitch condition relative to the monotone pitch condition in the same dimension was confirmed. For all results except the RT of ATT, the variable pitch condition had higher RT and lower NC than the corresponding monotone condition (see Figure 4—NC of ATT, Figure 7—RT of EHA, Figure 8—NC of EHA, Figure 11—RT of SCG and Figure 12—NC of SCG).

Another possible issue in this study is that participants were not reacting instinctually to the stimuli but actively listening and attentively waiting for a difference. The definition of popout stipulates it must be an automatic response, not something that arises from concentration on the stimuli. In the experiment we instructed participants to listen for a different note and indicate when they heard it, which inherently involves concentration. However, visual popout studies also instructed participants to actively observe a scene and indicate when a different stimulus was present (Ahissar & Hochstein, 1996; Treisman & Gelade, 1980; Triesman, 1988). In our experiment we also trained participants in the tones they would hear and detection of differences within timbre dimensions, which could prime an unnatural response. Visual popout experiments would also indicate the salient feature before stimuli presentation (Ahissar & Hochstein, 1996). Although certain elements of our experiment have possible contention with popout effect model qualifications, past studies investigating, or even introducing (Triesman & Gelade, 1980), the popout effect share these design aspects.

Further Research

To further investigate perceptual load and the effect of heterogeneity of distractors in auditory processing and specifically the popout effect, an interesting follow-up study would be to perform a similar study but with a second heterogeneity of distractors factor included to further distract participants. In this study two dimensions of timbre remained constant across all trials of a block while the third dimension varied in its popout tone. A follow-up study could have one of these non-salient dimensions of timbre change its value randomly for all notes in the series, like pitch did in the variable pitch condition. This study would investigate the possibility of an undiscovered link between the perception of two of these dimensions, whether the determinants of RT in popout follow the same variance pattern in audition as in vision (Triesman, 1988) relative to distractor type, and whether auditory perceptual load findings apply to timbre, specifically in the popout phenomenon (Murphy, Fraenkel & Dalton, 2013). This study found that a distractor of pitch variation reduced performance nearly unanimously, so it would be interesting to see if a timbral distractor has a similar effect.

In addition, further experimentation on the directional salience of these three timbre dimensions, using this experimental model or another, would be useful to better interpret this data and that of Chon (2013) to better understand direction salience in timbre.

Conclusions

Through 30 participants' detection of one tone within a series of 14 as different from the others, due to a change in value along one timbre dimension, this experiment gives strong evidence for the existence of an auditory popout effect, similar to that found in vision (Triesman & Gelade, 1980), which automatically attracts listener attention to a tone with a novel feature that differentiates it from its surroundings. This popout effect increases in strength as the value difference between the popout and baseline tones of the series increases. This experiment also gives weak evidence of a negative directional salience for the attenuation of even harmonics relative to odd, and a positive directional salience for the spectral center of gravity.

References

Ahissar, M., & Hochstein, S, (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research*, *36*(21), 3487-3500.

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 267-279.

Berti, S., & Schröger, E. (2001). A comparison of auditory and visual distraction effects: behavioral and event-related indices. *Cognitive Brain Research*, *10*(3), 265-273.

Caclin, A., McAdams, S., Smith, B., & Winsburg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Acoustical Society of America*, *118*(1), 471-482.

Chon, S. H. (2013). *Timbre saliency, the attention-capturing quality of timbre* (Doctoral dissertation, McGill University).

ISO 389–8. (2004). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones* (Tech. Rep.). Geneva, Switzerland: International Organization for Standardization.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, *58*(7), 977-991.

Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64–66.

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition*, *129*(2), 345-355.

Parmentier, F. B. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: a review. *Psychological Research*, *78*(3), 321-338.

Treisman, A., Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.

Treisman, A. (1988) Features and objects: The fourteenth Bartlett memorial lecture, *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *40*(2), 201-237.