

LOWER PITCH IS LARGER, YET FALLING PITCHES SHRINK: INTERACTION OF PITCH CHANGE AND SIZE CHANGE IN SPEEDED DISCRIMINATION

Zohar Eitan,* Asi Schupak,** Alex Gotler,*** and Lawrence E. Marks****

**School of Music, Tel Aviv University, Israel*

*** Department of Psychology, The Hebrew University, Israel*

****Department of Psychology, The Open University, Israel*

*****The John B. Pierce Laboratory, New Haven, Connecticut, and Yale School of Public Health and Department of Psychology, Yale University, New Haven, Connecticut, USA*

Abstract

Experiments using diverse paradigms, including speeded discrimination, indicate that pitch and visually perceived size interact perceptually, and that higher pitch is congruent with smaller size. While nearly all of these studies used static stimuli, here we use Garner's speeded discrimination paradigm to examine the interaction of dynamic pitch and size: continuous rise/fall in pitch and increase/decrease in object size. Two intriguing results emerged: first, congruence effects without Garner interference, an unusual pattern not consistent with some interpretations of Garner interference; and second, congruence effects where ascending pitch is congruent with growing size, descending pitch with shrinking size. Thus static and dynamic versions of these auditory and visual stimuli interact in opposite ways. Our interpretation of these results focuses on effects of within-trial changes on Garner's tasks and on the association of changes in apparent size with changes in distance.

Beyond its established association with height ("high" vs. "low" tones), auditory pitch has been shown to interact perceptually and associate cognitively with several spatial and spatio-kinetic dimensions, including depth, laterality, and speed (see Eitan, in press). Perhaps most importantly, pitch has been consistently associated with *physical size*, such that larger objects are associated with lower pitch by both adults and young children (Marks et al., 1987; Mondloch & Maurer, 2004). In speeded tasks, where smaller or larger visual stimuli were paired with high or low pitch, participants discriminated both visual stimuli and auditory stimuli faster when pairs were congruent – larger visual stimuli + lower pitches, smaller ones + higher pitches (Gallace & Spence, 2006). A similar congruence effect emerged in an indirect task, when both pitch and visual size were irrelevant to the task, suggesting that the association need not depend on attention or explicit instruction (Evans & Treisman, 2010).

This robust association of higher pitch with smaller size may stem from our life-long experience of correlation between an object's size and the pitch it would produce. In particular, pitch height is correlated across animal species with body size, since larger species tend to produce lower-pitched sounds. This correlation may have important behavioral implications in a variety of species, including humans (Morton, 1977), as discussed later.

Importantly, studies associating pitch and size have mostly used static stimuli (e.g., a stationary auditory pitch or visually-perceived object). Yet, it is far from evident that comparable static and dynamic stimuli (e.g., fixed frequencies vs. pitch glides) would show similar interactions. The perception and mental representations of basic stimulus features, like magnitude or position, vary with dynamic qualities, such as direction of change. For instance, judgments of loudness are more variable in dynamic conditions and are affected

asymmetrically by the direction of change (Neuhoff, 1998). Spatial direction affects the perception of spatial position with both visual and auditory stimuli (Bosbach, Prinz, & Kerzel, 2004; Walker & Ehrenstein, 2000). Direction of change also affects mental representations of spatial position, as memory for a moving object's final position is distorted in ways consistent with its path of motion. Analogies to this "directional momentum" effect (Freyd & Finke, 1984) also arise in auditory pitch (Kelly & Freyd, 1987). Together, such effects strongly suggest that audio-visual interactions may differ with static and dynamic stimuli. Indeed, recent studies examining listeners' spatio-kinetic mappings of dynamic auditory stimuli (e.g., pitch rise or fall, crescendo or diminuendo) revealed effects not previously reported with static stimuli, such as congruence between loudness change and vertical direction (Eitan et al., submitted) and directional asymmetries in mapping auditory and spatio-kinetic features (e.g., pitch fall but not rise associates with motion in the vertical axis: Eitan & Granot, 2006).

Here, we examined cross-modal interaction of *dynamic* pitch and size in two tasks using Garner's (1974) speeded discrimination paradigm. In both tasks, participants received simultaneous auditory and visual stimuli: sinusoids increasing or decreasing in pitch + circles increasing or decreasing in size. In the auditory and visual tasks, participants were asked to rapidly discriminate the direction of pitch change, ignoring the concurrent visual stimuli, and the change in visual size, ignoring the concurrent auditory stimuli. In each task, stimuli were presented in five separate conditions, applying Garner's design. In the *filtering* condition, auditory and visual values varied orthogonally. In two *baseline* conditions, values in one modality remained constant, while values in the other varied (irrelevant increasing in one baseline, decreasing in the other). In two *correlated* conditions, values in the two modalities were correlated, either positively (congruent) or negatively (incongruent), congruence defined as in studies of static pitch and size (described above), where higher pitch = smaller.

Interactions may take two forms. In *Garner interference*, performance is better (faster and/or more accurate) at baseline, compared to filtering. Garner interference indicates failure of selective attention: participants are unable to ignore changes in the irrelevant dimension and selectively attend to the relevant dimension. With *congruence effects*, performance is faster and/or more accurate on congruent trials than incongruent trials.

Method

Participants

64 Tel-Aviv University students, 23 women and 41 men, took part in both experiments (age range 18-39 yrs; mean =23.67, SD=4.00), 32 having musical training (>7 years of formal musical training), other 32 having little or none. Nine additional participants were excluded because of high error rates (more than 6 errors in an 18 trial block; this exclusion rule aimed to keep the mean RTs in all blocks comparable in sample size). All participants had normal or corrected-to-normal visual acuity, as well as normal hearing, as assessed by self-reports. Participants were paid 30 NIS (about 8\$) for a 15-min experimental session.

Stimuli, Apparatus, and Design

Stimuli. To encourage participants to heed the *direction of change* (increase or decrease) in the relevant dimension, rather than the initial *position* (large or small size, high or low pitch), we started both the visual and auditory stimuli at three different values, with equal

numbers of increasing and decreasing stimuli emanating from each starting point. Auditory stimuli consisted of 1-s long 1000 Hz sinusoids, increasing or decreasing in octave from starting points of 157, 440, and 1232 Hz. Level was set at 60 dB(A). Visual stimuli consisted of dark gray circles (RGB: 51:51:51) displayed on a black background, with initial radiuses of 1.6cm, 2.14cm, and 2.68cm, decreasing or increasing in size by 3mm per second. Rates of auditory and visual change were chosen to be as equally discernable as possible, based on RTs in single-dimension pretests performed with a different group of participants. The same set of stimuli was used for both the auditory and visual tasks.

Apparatus. Visual stimuli were displayed on a Lenovo 17-inch color LCD monitor (85 Hz refresh rate, resolution = 1,024 x 768 pixels). Auditory stimuli were played through Sennheiser headphones (HD580 Precision) by an external sound card (Terratect Producer, Phase 24). Sound levels were verified before each session using a Brüel & Kjær precision sound level meter (type 2232). A keyboard attached to a laptop computer recorded responses. Reaction time (RT) was measured using a software timer.

Design. Five experimental conditions (described above) were created for each of the two tasks (visual and auditory). In the auditory task, Baseline 1 combined the 6 auditory stimuli (2 change directions X 3 starting pitch levels) with a circle increasing in size, emanating from a mid-level circle (2.14 cm), located at the center of the screen, while Baseline 2 combined these auditory stimuli with a circle decreasing in size starting from the same mid-level. In the visual task, Baseline 1 combined the 6 visual stimuli (2 directions of motion X 3 starting positions) with a sinusoid increasing from 440 to 880 Hz, while Baseline 2 combined the same visual stimuli with a sinusoid decreasing from 440 to 220 Hz. Each baseline condition comprised 18 trials, each of its 6 auditory-visual combinations being repeated 3 times, randomly ordered. In the positively correlated conditions (both tasks), the 18 congruent auditory-visual combinations (size increase/pitch decrease; size decrease/pitch increase) were randomly ordered into 18 trials, while the negatively correlated conditions used the 18 incongruent auditory-visual combinations (size and pitch increase/size and pitch decrease). In the Filtering conditions (both tasks), all 36 auditory-visual combinations were randomly assigned to 36 trials. To equate the number of trials in all blocks, we divided the 36 Filtering trials into two blocks of 18 trials, such that participants in each task received 6 blocks, equal in the number of trials. Altogether, each task comprised 108 trials.

Procedure

Participants were tested individually in a dimly lit sound-attenuating chamber. They were seated at a viewing distance of approximately 60 cm from the center of the screen, such that each circle subtended a visual angle between $3.2^\circ \times 3.2^\circ$ to $5.36^\circ \times 5.36^\circ$.

Each participant took part in one task, auditory or visual. In each task, half of the participants were musically trained and half were not. Participants completed five conditions (described above) In each task: Filtering (2 blocks), Baseline 1, Baseline 2, Positively Correlated, and Negatively Correlated. A break of 1 minute separated successive blocks. To control for block-order effects, each condition (Filtering, Baseline 1 or 2, and Positive or Negative Correlated), appeared in roughly equal proportions as the first block; the order of the remaining blocks was randomized for each participant. Responses were made by pressing one of two keys (“D”, “K”). Key assignment within and between tasks was counterbalanced across participants, separately for musicians and non-musicians. Stimuli were terminated following 1 s, regardless of participants’ response. Following a 500 ms inter-trial interval, the next stimulus was generated.

Participants were instructed to attend to the relevant dimension and ignore irrelevant variation. They were also asked to discriminate two types of *change* in the relevant dimension, ignoring its initial *position*. To lessen the effect of conventional linguistic labels, change categories were labeled “D” and “K” (the response keys), rather than “increase” and “decrease” or “rise and “fall.” Participants were encouraged to respond as quickly and accurately as possible.

Each experimental session started with practice that introduced the task instructions, target stimuli (pitch glides in the auditory task, circles changing in size in the visual task), and response key assignment. Auditory and visual values in the practice session varied orthogonally. In the visual task, practice ended when a participant completed five consecutive correct responses and at least 18 trials; in the auditory task, 7 consecutive correct responses and at least 24 trials were required. Due to the difficulty of the pitch discrimination task, we repeated the auditory practice session twice – first with a response termination fashion, and second with the constraint of responding in less than one second, as in the experimental session itself. The entire experimental session lasted about 15 min, including instruction, practice, and six experimental blocks.

Results

Analyses focused on congruence effects and Garner interference, evaluated statistically by ANOVA. Musicians responded faster than non-musicians by 65ms on average (608ms and 673ms respectively, $[F(1,60)=9.72, p<.05]$). Musicianship, however, did not interact with any other variable ($F<1$ for all interactions). There was no main effect of task ($F<1$), as participants classified pitch change as quickly as size change (mean RT of 640ms for both).

Congruence effects. Unlike effects found with static stimuli, *incongruent* stimuli were classified faster (629ms) than congruent ones (652ms): significant “negative” effect of congruency $[F(1,60)=38.46, p<.001]$. Congruency also interacted with Task $[F(1,60)=4.55, p<.05]$. While planned comparisons show significant “negative” congruency effects for both the auditory task [626ms for incongruent, 656ms for congruent trials; $F(1,60)=34.72, p<.001]$ and the visual task [633ms for incongruent, 648ms for congruent trials; $F(1,60)=8.28, p<.01]$, the auditory effect is stronger. For the auditory task, planned comparisons show significant negative congruency effects in the Filtering (617ms for incongruent, 646ms for congruent trials; $F(1,60)=18.29, p<.001$), Correlated (622ms for incongruent, and 659ms for congruent trials ; $F(1,60)=20.09, p<.001$) and (unusually) Baseline conditions [638ms for incongruent, 661ms for congruent trials ; $F(1,60)=8.54, p<.01]$. In the visual task, significant effects are shown in the Correlated (624ms for incongruent, and 643ms for congruent trials; $F(1,60)=5.29, p<.05$) and Baseline conditions (635ms for incongruent, and 656ms for congruent trials; $F(1,60)=7.36, p<.01$), but not in Filtering (640ms for incongruent, 643ms for congruent trials; $F<1$).

Garner Interference. No Garner interference was observed. On the contrary, mean RT in the Filtering condition (637ms) was somewhat faster than that in the Baseline condition (648ms), a difference that was marginally significant $[F(1,62)=3.23, p=.08]$.

Errors. Due to space limitations, error results are not reported here in full. The auditory task showed a “negative” Congruence effect, replicating that found for RTs, in non-musicians in the filtering condition (16.37% for congruent trials and 8.33% for incongruent

trials; $F(1,60)=13.83$, $p<.001$). No significant congruence effects were found for musicians, and Garner interference was not observed in either group. Importantly, the error data counter the possibility that the negative congruence effects for RT resulted from error-RT tradeoffs, since no indication for “positive” Congruence effect for errors was found.

Discussion

There are two novel aspects to this study. Methodologically, it employs Garner’s well-established paradigm within a new realm: *dynamic* crossmodal stimuli. Topically, it examines whether an ecologically important crossmodal interaction, previously observed in static stimuli, would also be found in comparable dynamic stimuli.

Results concerning both aspects are intriguing. First, the pattern of results (also replicated in experiments using other dynamic audio-visual stimuli; Eitan et al., submitted) is unusual: significant congruence effects without Garner interference. Second, the congruence effect found for dynamic stimuli is *opposite* to the effect for static stimuli, exhibiting an apparent “paradox”: while higher pitches are *smaller* than lower ones, pitch *grows*, rather than shrinks, as it rises. While any definitive interpretation of these results would be premature, some accounts do suggest themselves.

Congruence effects without Garner interference

While Garner interference may readily occur with no congruence effects, congruence effects with no Garner interference are rare, particularly in cross-modal stimuli (Marks, 2004; but see, e.g., Van Leeuwen & Barker, 1995, Patching & Quinlan, 2002, for exceptions). Some studies (Pomerantz et al., 1989) have even suggested that congruence effects may never occur without Garner interference. Indeed, Garner interference is usually taken to indicate failure of selective attention; if selective attention does not fail, how could properties of the irrelevant dimension – supposedly unattended – significantly affect performance regarding the relevant, attended-to dimension?

Ben-Artzi and Marks (1995) proposed a model that may account for such a pattern of results. The model assumes that abstract information arising from different stimulus dimensions (here, directions of pitch change and size change) accumulates and combines into a single bipolar variable, such as “direction of magnitude change” (“increase” or “decrease” in size or pitch), whose value is continuously compared to two criteria, which govern the responses. Increasing size and increasing pitch height move the variable toward one pole and decreasing size and height move it toward the other pole. The variable will reach the criterion for response more quickly and accurately, therefore, when the stimuli are congruent rather than incongruent. When the positive and negative effects are similar, the average RT in the Filtering and Baseline conditions will be similar, resulting in no significant Garner interference.

The lack of Garner interference may also be specifically related to the use of dynamic stimuli in the present experiment, distinguishing it from previous experiments using Garner’s paradigm. In Garner experiments using static stimuli, irrelevant stimuli remain unchanged throughout the baseline blocks. In contrast, in the present experiment irrelevant stimuli in the baseline blocks, while constant *across* trials, do change *within* trials. These recurrent changes, intrinsic to dynamic stimuli, could attract attention, and conceivably could thus slow down response in the baseline trials, preventing Garner interference.

Importantly, both interpretations suggest situations in which selective attention may fail, yet not produce Garner interference. They thus support earlier work (e.g., Pomerantz, 1983) that questioned the role of Garner interference as a litmus paper for selective attention failure.

Pitch-size congruence reversed in dynamic stimuli

As noted above, the association of pitch and size may carry significant behavioral implications. Morton (1977) observed a widespread tendency among both mammals and birds to use lower-pitched, rough voice in confrontational situations, when the expression of aggression or hostility is called for, and higher-pitched, tone-like voice as an expression of submission or friendliness. Morton suggested that since larger body size is advantageous in a physical conflict, animals in an aggressive stance would try to appear larger both visually (e.g., erecting their feathers, raising their tail) and vocally, by producing lower pitched sounds, associated with larger, stronger animals (see also Ohala, 1994; Scherer, 2003). Correspondingly, an expression of submission would entail appearing small and hence unthreatening, and would thus be related to higher pitch. Associations of lower pitch with dominance and threat, possibly stemming from the pitch/size correspondence, are also effective for humans (Puts, Gaulin, & Verdolini, 2006). Why, then, are these ecologically-important associations completely reversed in dynamic contexts?

The association of apparent size change with change of distance may account for this finding. A continuous change in image size may be associated with approach or withdrawal, since approaching objects produce an apparently growing retinal image, and objects moving away – a shrinking retinal image. Such association may be even stronger for a two-dimensional image on a screen, as used in the present experiment. Pitch fall is associated, via the Doppler Effect, with increasing distance, and though pitch rise is not acoustically related to decreasing distance, a perceptual illusion associating pitch rise with approach has also been reported (Neuhoff & McBeath, 1996). Thus, pitch rise and fall may be related to perceived size change via their shared association with perceived change of distance.

The association of pitch change and size change may also stem from a more abstract analogy, supported by language use. Pitch rise and fall and object growth and shrinkage may both map onto an abstract dimension of magnitude change. Note that Ben-Artzi and Marks' model (1995), described above, indeed suggests such high-level, abstract representation as a basis for our pattern of results, and is thus consistent with the account suggested here.

Regardless of the specific accounts proposed here for the results, one ramification of these results is clear. They strongly suggest that static and dynamic stimuli may differ substantially in their cross-modal mappings, as well as in the sources and processing mechanisms these mappings are based upon. Not less importantly, seasoned tools used reliably and efficiently to examine dimensional interaction in static stimuli, like Garner's paradigm, may reveal unusual patterns of results given dynamic stimuli. By implication, the tools themselves may need to be adapted to the radically different behavior of stimuli in flux.

References

- Ben Artzi, E., & Marks, L. E. (1995a). Congruence effects in classifying auditory stimuli: A review and a model. In B. Scharf and C.A. Possamai (Eds.), *Fechner Day '95*, Cassis, France: International Society for Psychophysics.

- Bosbach, S., Prinz, W., & Kerzel, D. (2005). Is direction position? Position-and direction-based correspondence effects in tasks with moving stimuli. *The Quarterly Journal of Experimental Psychology*, 58a (3), 467-506.
- Eitan, Z. (in press). How pitch and loudness shape musical space and motion: New findings and persisting questions. In A. Cohen, R. Kendall, S. Lipscomb & S. L. Tan (Eds.), *The Psychology of Music in Multimedia*. Oxford: Oxford University Press.
- Eitan, Z., & Granot, R. Y. (2006). How music moves: musical parameters and images of motion. *Music Perception*, 23, 221-247.
- Eitan, Z. Schupak, A., Gotler, A., & Marks, L. E. (submitted). Louder is higher: Loudness change and visual vertical motion interact in speeded classification.
- Evans, K. & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, 10, 1-12.
- Freyd, J. J. & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126-132.
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, 68, 1191-1203.
- Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac, MD: Erlbaum.
- Kelly, M. H., & Freyd, J. J. (1987). Explorations of representational momentum. *Cognitive Psychology*, 19, 369-401.
- Van Leeuwen, C., & Barker, L. (1995). Stroop can occur without Garner interference: Strategic and mandatory influences in multidimensional stimuli. *Perception & Psychophysics*, 57, 379-392.
- Marks, L.E. (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence, and B.E. Stein (Eds.), *Handbook of Multisensory Processes* (pp. 85-106) Cambridge, MA: MIT Press.
- Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, 52, no. 215.
- Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2), 133-136.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855-869.
- Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, 395, 123-124.
- Neuhoff, J.G., & McBeath, M.K. (1996). The Doppler illusion: The influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 970-985.
- Ohala, J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge, UK: Cambridge University Press.
- Patching, G. R., & Quinlan, P. T. (2002). Garner and congruency effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception & Performance*, 28, 755-775.
- Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In B. Shepp & S. Ballesteros (Eds.), *Object Perception: Structure and Process* (pp. 53-89). Hillsdale, NJ: Lawrence Erlbaum Associates.)
- Puts, D., Gaulin, S., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27, 283-296.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227-256.

Walker, B. N., & Ehrenstein, A. (2000). Pitch and pitch change interact in auditory displays.
Journal of Experimental Psychology: Applied, 6, 15-30.