

# **On the clarity of the music stave**

Arnold J Wilkins and Isobel Kiff

Department of Psychology

University of Essex

## **Abstract**

The staves of 63 scores from 50 music publishers were measured. The staves had similar height but with lines that varied in thickness by a factor of 4 from 0.1-0.4mm. Evidence from visual psychophysics suggests that when the staff has thick lines, perceptual distortions are likely to affect the clarity of the score adversely. Music students were asked to sight read scores comprising random notes (12 tone) or random notes in the key of G. The scores had staves with lines that were 0.1 or 0.4mm thick (the extremes of current typographic practice). Twice as many errors were made when the staves had thick lines although the scores were read more slowly. Scores in the key of G were read more accurately than the 12 tone scores, but those with thick lines were read with as many errors as 12 tone scores with thin lines. There was a tendency for individuals with high pattern glare scores to read the scores with thick lines relatively slowly. The findings suggest that perceptual distortions can impair sight reading of music manuscript because of the pattern from the lines of the staff; using thinner lines increases sight reading accuracy and speed.

Keywords: visual stress, illusions, staff, clarity, typography, sight reading

Within the Western tradition of music notation the staff began as a single line and then had companion lines added to it. These defined the pitch of the notes, and eventually became standardised for plainsong as a staff of four lines, sometimes with different colours. Standardisation of the five-line staff did not occur until the seventeenth century. Until then, different composers contemporary with one another used a wide variety of staves. In the organ music of Frescobaldi (1583-1643), for example, the staff for the right hand had six lines with eight lines for the left; Sweelinck (1562-1621) used two staves of six lines and Scheidt (1587-1654) used four staves of five lines each. By the middle of the seventeenth century the five line staff had become nearly universal except for plainsong, which to this day continues to use a four line staff; for review, see Apel (1942). In this paper we show that although the five line staff is nearly universal it is not necessarily optimal for sight reading because of the perceptual distortions that the pattern of lines evokes.

In 1832 Sir David Brewster observed:

If the eye looks at (parallel) black lines drawn upon white paper steadily and continuously, the black lines soon lose their straightness and parallelism, and inclose luminous spaces somewhat like the links of a number of parallel chains. When this change takes place, the eye which sees it experiences a good deal of uneasiness... When this dazzling effect takes place the luminous spaces between the lines become coloured, some with yellow and others with green and blue light (Brewster, 1832, p. 170).

There are large differences between people as regards their susceptibility to the above perceptual distortions, which have been termed visual stress (Wilkins, 1995). Anecdotally, some accomplished musicians known to the authors are unable to sight read because of the distortions they see when trying to read music. The following is a written description of the distortions by a chorister:

The bars either side of the one I am looking at appear a mess – the notes oscillate slowly up and down and I cannot say how many notes there are. I cannot see the lines, so I'm not aware of them moving. The lines become a single haze. (Personal communication, 26 October 2013)

It has been proposed that the distortions arise in the visual cortex of the brain (Wilkins et al., 1984). Individuals with migraine are particularly susceptible, perhaps because migraine is associated with a cortical hyperexcitability (Wilkins, 1995; Huang et al., 2011). Only a few individuals will be aware of distortions in a page of music, but if the lines are thicker, and there are more of them, most (but not all) individuals will experience the distortions that Brewster described (Wilkins, 1995). The reader may not be aware of instability in Figure 1A with few thin lines or even 1C with few thick lines, but

most observers find Figure 1D with many thick lines uncomfortable and unstable to view (Wilkins et al., 1984).

Insert Figure 1 about here

The distortions interfere with perception quite generally, so that it becomes difficult to see faint targets superimposed upon a pattern of lines. Chronicle and Wilkins (1996) measured the interference. They optically superimposed faint letters or shapes on patterns of stripes and measured the contrast at which these targets could just be seen, i.e. the threshold contrast, comparing the threshold with different patterns of stripes. They showed that it was more difficult to discern the targets when the stripes had characteristics such that they induced perceptual distortions.

### **Pattern parameters**

In the experiments by Wilkins et al. (1984) and Chronicle and Wilkins (1996) the perceptual distortions and interference with vision were both affected by three important parameters of the pattern: (1) its spatial frequency, that is, the number of pairs of black-white stripes (i.e. cycles of the pattern) occupying one degree subtended at the eye, see Figure 1E; (2) its duty cycle, i.e. the separation of the lines expressed as a proportion of one cycle of the pattern, see Figure 1F; (3) its size, the angle the radius of the pattern subtended at the eye, see Figure 1G.

*Spatial frequency.* The spatial frequency at which the illusions and interference were maximal was about 4 cycles per degree, that is, when four pairs of black and white stripes of the pattern subtended one degree at the eye, or alternatively, when each pair of black and white stripe subtended 15 minutes of arc at the eye, see Figure 2A and 2B.

Insert Figure 2 about here

*Duty cycle.* The duty cycle at which illusions and interference were maximal was 50% (when the black lines and the white spaces between them had equal width). The interference decreased progressively as the duty cycle departed from this value, i.e. as the black lines became thinner or thicker than the white spaces, see Figure 2C and 2D.

*Pattern size.* The size of the pattern was critical: the larger the pattern, the greater the illusions and interference, but the increase was determined by the area of the visual cortex to which the pattern projected. In Figures 2E and 2F, the horizontal axis shows the proportion of the visual cortex stimulated, and beneath it the size of the pattern expressed as the number of degrees the radius of the pattern subtended at the eye of the observer.

The question arises as to whether music scores provide a pattern with parameters appropriate for the induction of perceptual distortions. We therefore surveyed typographic practice in contemporary music publishing to determine the parameters of pattern size, spatial frequency and duty cycle that are typical for the lines of the stave. We surveyed popular songs, but have since found that the measurements apply quite generally.

### **Survey of contemporary publishing practice**

Sheet keyboard music for popular songs published between 1957 and 2008 was sampled in 63 publications: 50 different publishers from the USA and Europe were represented in the sample, and are listed in the Supplementary Materials. A measuring magnifier was used to obtain measurements to the nearest 0.05mm.

Two measurements were taken from each sample: the height of the staves and the thickness of the lines of the staves. The average height of the staves was 6.9mm (SD 0.4mm). Figure 3A shows a histogram of the heights of the staves: as can be seen, they ranged in height from 5.6 to 7.6 mm and the most common height was 7.0mm. More than 75% of the sample had staves measuring between 6.6 and 7.2mm in height (a range of only about 10%). There was therefore considerable consistency from one publisher to another with respect to the height of the staves in published music scores.

Insert Figure 3 about here

By way of contrast, there was little consistency among publishers with respect to the thickness of the lines of the staves. The average thickness of the lines was 0.27mm (SD 0.09). The thickness varied from a minimum of 0.10mm to a maximum of 0.45mm, a range of more than 400%, see Figure 3B. Any tendency for thick lines to be more widely spaced was very weak: the Pearson correlation between thickness and separation was 0.04 and not significant.

The survey of music scores covered the years between 1957 and 2008. Despite the changes in printing technology over this period, there was a little tendency for the thickness of the stave lines to have changed. The Pearson correlation coefficient between line thickness and date of publication was -0.17 and not significant.

The distance from which music is viewed varies somewhat as a result of the constraints provided by the instrument being played. For the present purposes we can estimate the viewing distance as ranging from the near point of (clear) vision (about 0.4m) to the maximum distance at which the score can be read, about 0.6m. The average height of the staves was 6.85mm, which equates to 0.98 degrees at 0.4m and 0.65 degrees at 0.6m. Each stave had 5 cycles so the spatial frequency of the

pattern ranged from about 5 to about 8 cycles per degree. This is close to the spatial frequency for which perceptual distortions are maximally likely, see Figure 2A, and for which interference with perception is maximal, see Figure 2B.

By convention, the duty cycle of a pattern is expressed in terms of the extent of the bright part of the cycle as a proportion of the entire cycle, see Figure 1F. In this context the duty cycle is therefore the distance from the top of one line to the bottom of the line above divided by the distance from the top of one line to the top of the line above. The mean duty cycle was 83.7% and ranged from 73.7% to 94.2%. The lower part of this range is within the range of duty cycles that evoke perceptual distortions, see Figure 2C and 2D. Over this range, the contrast necessary to see a target superimposed on the pattern varied by a factor of 2, see Figure 2D.

The staves on a page of music typically measure about 200mm wide by 300mm high. The radius of the pattern when centrally fixated (>10 degrees) is therefore quite sufficient to induce distortions, see Figure 2A and 2B. Of course the pattern of periodic lines is broken by the spacing between the staves, which reduces the effect of the pattern. On the other hand, each staff on its own subtends about 0.7 degrees, and patterns of this size can in some individuals be sufficient to induce distortions (Wilkins 1995).

Of the three parameters considered thus far, spatial frequency, duty cycle and size, duty cycle is the one that varies most from publisher to publisher. The typical values range from those that are unlikely to induce distortions to those that are likely to do so. The effects of any striped pattern can be greatly reduced by increasing its duty cycle. By increasing the duty cycle of the lines of the staves, making the spaces larger and the lines correspondingly thinner, it may be possible to improve the speed and accuracy of sight reading, even in individuals who are unaware of any distortions. The following experiment put this idea to the test. Students were asked to sight read music printed with staff lines 0.1mm thick and with lines four times as thick (0.4mm). (Both these line thicknesses are within the range of conventional typographic practice, as demonstrated earlier.) The speed and accuracy of the performance were measured.

## **Sight reading from staves with thin and thick lines**

### *Method*

*Participants.* An opportunity sample of 28 students (6 male, 22 female) took part. They were not all music students but all had certification from one of the three UK registered exam boards: Royal Schools of Music, London College of Music and Victoria College of Music. The mean grade in piano was 3.0 and the range 1-7.

*Materials.* A random sequence of notes in the key of G was generated by an algorithm that randomly sampled the notes on a major scale of G. The mean of the distribution was C<sub>4</sub>. The registral range was from C<sub>3</sub> to B<sub>5</sub>. A similar algorithm generated a random sequence from a chromatic scale. The “black keys” were randomly designated as sharp or (the equivalent) flat. The mean of the distribution was F<sub>4</sub>. The registral range was from C<sub>3</sub> to Bb<sub>4</sub>.

Two random sequences of notes were generated for the key of G and two for the chromatic sequence. A music processing package (Sibelius(R) [www.avid.com](http://www.avid.com)) was used to generate a score based on the selected notes. Each score was prepared in music having staves with lines that when printed were 0.1mm wide (“thin”) and in music having staves with lines that when printed were 0.45mm wide (“thick”). Sixty bars, each with 4 crotchets per bar, were laser printed on a single sheet of white A4 paper. Figure 4 shows the first line of some representative scores. The random sequence of notes ensured that participants had to see the notes in order to read them; they could not predict the sequence from a melody.

Insert Figure 4 about here

*Procedure.* The two scores (scale of G and chromatic) and the two versions of each (thick and thin staves) were counterbalanced across participants. Each participant experienced the experimental conditions in one of four possible sequences, 7 participants per sequence. The test order minimised the effects of practice at the task, while balancing the differences in scores across participants. Participants were seated at a keyboard (Yamaha PSR-420) at a distance of about 0.5m from the score. Each participant was allowed 30 seconds to review the score prior to playing it. They were asked to play at a pace that was comfortable, and that allowed them to play accurately. Participants then proceeded to play each score and were recorded whilst doing so. The time taken to play each piece and the number of incorrectly played notes were recorded.

After completion of the sight reading, the individual participants’ susceptibility to perceptual distortions was assessed using the Pattern Glare test (Wilkins & Evans, 2001). This is a simple test

involving the presentation of three gratings, the second of which has a spatial frequency of about 3 cycles/degree and is likely to evoke perceptual distortions. The other gratings have higher and lower spatial frequencies and are less likely to evoke distortions. Participants are simply asked to describe what they see when they look at the three gratings using a checklist: colours, bending of lines, blurring of lines, shimmer/flicker, fading, shadowy shapes, other effects.

Insert Table 1 about here

### *Results*

Table 1 presents the mean number of errors and the time taken to read each score. There were more than twice as many errors in sight reading when the staves had thick lines. An analysis of variance of the errors revealed a significant main effect of the thickness of the lines of the staves,  $F(27)=27.9, p < .001$ , accounting for 30% of the variance. There was a significant main effect of the score,  $F(27)=50.6, p < .001$ , accounting for 22% of the variance, indicating that, unsurprisingly, the chromatic sequence of notes was more difficult to read.

The interaction term was significant,  $F(27)=10.9, p = .003$ , indicating that when the sequence of notes was more difficult to play, the effect of the line thickness was greater. Note that the use of thick staff lines makes music in the key of G major as difficult to read as the chromatic music with thin lines (the difference between these two conditions is non-significant).

A repeated measures analysis of variance of the time taken to read the passages revealed a significant main effect of the thickness of the lines of the staves,  $F(27)=6.72, p = .015$ , which accounted for 14% of the variance, and a significant main effect of the score,  $F(27)=6.54, p = .016$ , accounting for 3%. The interaction term was not significant. Participants took 2.5% longer to read the scores printed with thick lines and were slower sight reading the chromatic material. There was evidently no speed/accuracy trade-off: both speed and accuracy of sight reading were greater with the thinner lines.

*Individual differences in susceptibility.* On the second pattern of the Pattern Glare test the total number of distortions reported averaged 3.1 across individuals and ranged from 0 to 6. For each individual the effect of the line thickness was calculated by subtracting the number of sight-reading errors made with the thin lines from those made with the thick. There was a small positive correlation between this difference score and the number of distortions reported (Pearson  $r=0.36, p = .030$ ) indicating that those individuals who reported more perceptual distortions in a pattern of stripes tended to be those who were most affected by the thickness of the lines of the staves.

## *Discussion*

We have shown that music is sometimes printed with stave lines that are unnecessarily thick. Far from making the stave easier to read, the thicker lines decreased the accuracy and speed of sight reading. Errors doubled when the line thickness increased from 0.1mm to 0.4mm. Music in the key of G major became as difficult to read as chromatic music. There are at least two possible mechanisms for the effect of the stave line thickness and they are probably related one to another.

1. The physical properties of a music score can resemble those of stripes that provoke perceptual distortions, particularly with respect to the duty cycle (thickness/spacing ratio) of the lines of the staves. The lines may provoke distortions that interfere with sight reading. This interpretation is supported by the finding that the individuals who were most susceptible to the effect of line thickness tended to be those who reported more distortions in a pattern of stripes. Anecdotally, the authors know of accomplished musicians who cannot read music. They report that the lines appear unstable. Nevertheless there were many individuals in the present study who did not report distortions but who also read less accurately when the stave lines were thick.

2. Another possible mechanism is the effect that the stave might have in masking the remaining notation. In general, patterns of stripes mask other contours superimposed upon them (Chronicle & Wilkins, 1996). The masking has an effect similar to reducing the contrast by a factor of about two, however, and although this might reduce the visibility slightly, it would not render the notation imperceptible, given its high initial contrast. It might nevertheless contribute to the difficulty of discriminating between the subtle positions and shapes that differentiate the pitch and duration of notes.

The present findings suggest that teachers of music should be aware that perceptual distortions of a musical score are possible. If pupils report perceptual instability (the notes “appear to move around”) then it is possible that they are affected by the thickness of the stave lines. This is particularly likely if the score is photocopied and the stave has a high and atypical contrast, with relatively thick stave lines. It is then possible that coloured filters placed upon the page (i.e. coloured overlays) will reduce the effects of the distortion (Wilkins, 2003). Coloured filters have been shown to reduce perceptual distortion of text and increase reading speed, provided the colour is chosen individually from a range sufficient to permit the selection of a shade appropriate for the individual. Individually selected coloured filters have been shown to reduce cortical hyperexcitability in response to patterns (Huang et al, 2011).

When music was handwritten on printed manuscript paper, the staves were usually printed in grey ink. This has a perceptual effect similar to that of making the lines thinner. The photographic reproduction of music that came later often increased line thickness, depending on the optical quality of the system. Photocopying of music also increased the thickness of lines of the staves because of the contrast enhancement of any blur. It is possible that some of the unnecessarily thick lines occurred in music that originated with these early photographic printing techniques, although staff line thickness has not decreased from 1957 to 2008 despite the improvement in printing techniques over this period. With the advent of laser printing it is now possible to publish staves with lines of any selected thickness, and the high spatial resolution of modern laser printers facilitates the production of staves with thin lines. Publishers should take note that the sight reading of music can be improved by publishing staves with thinner lines. This would benefit all musicians, but perhaps particularly those who experience perceptual instability. The latter would be well advised to seek out editions of compositions that employ a lightly ruled staff.

Sight-reading is an essential skill for most musicians, and in some contexts musical performance is at first sight. Although a knowledge of keyboard topography is recognised as one of the skills required in sight-reading piano, grand staff fluency is also essential (Beauchamp, 1999; Hardy, 1998). The ability to read and recognise groups of notes may be compromised by perceptual distortions that interfere with the perceptual integration of notes on the staff.

## References

Beauchamp, Laura (1999). The 'Building Blocks' of Reading: Suggestions for Developing Sight Reading Skills in Beginning Level College Piano Classes. *Piano Pedagogy Forum* (Columbia, SC: University of South Carolina School of Music) 2(2). <http://www.max.grenkowitz.net/?topic=690>. Downloaded 15 July 2014.

Brewster, D. (1832). On the undulations excited in the retina by the action of luminous points and lines. *The London and Edinburgh Philosophical Magazine and Journal of Science*, Volume 1(3), September; 169-174.

Chronicle E., and Wilkins A.J. (1996). Gratings that induce perceptual distortions mask superimposed targets. *Perception* 25(6) 661 – 668.

Drasdo, N. (1977). The neural representation of visual space. *Nature (Lond)*, 266, 554-556.

Hardy, Dianne (1998). Teaching Sight-Reading at the Piano: Methodology and Significance, *Piano Pedagogy Forum* (Columbia, SC: University of South Carolina School of Music) 1(2). <http://www.max.grenkowitz.net/?topic=710>. Downloaded 15 July 2014.

Huang, J., Zong, X., Wilkins, A., Jenkins, B., Bozoki, A. and Cao, Y. (2011). fMRI evidence that precision ophthalmic tints reduce cortical hyperactivation in migraine. *Cephalalgia* , 31(8), 925-36.

Jainta, S., Jaschinski,W. and Wilkins, A.J. (2010). Periodic letter strokes within a word affect fixation disparity during reading. *Journal of Vision*,10(13), 2.

Kinsler, V., and Carpenter, R.H. (1995). Saccadic eye movements while reading music. *Vision Research*, 35(10), 1447-58.

Penttinen, M., Huovinen, E., and Ylitalo, A-K. (2013). Silent music reading: amateur musicians' visual processing and descriptive skill. *Musicae Scientiae*, 17(2) 198-216.

Wilkins, A.J. (1995). *Visual stress*. Oxford University Press: Oxford.

Wilkins, A.J. (2003). *Reading through colour*. Wiley: Chichester.

Wilkins, A.J., Smith, J., Willison, C.K., Beare, T., Boyd, A., Hardy, G., Mell, L., Peach, C. and Harper, S. (2007). Stripes within words affect reading. *Perception*, 36(12), 1788-1803.

Wilkins, A.J., Nimmo-Smith, M.I., Tait, A., McManus, C., Della Sala, S., Tilley, A., Arnold, K., Barrie, M., Scott, S. (1984). A neurological basis for visual discomfort. *Brain*, 107, 989-1017.

Wilkins, A.J. and Evans, B.J.W. (2001). Pattern Glare Test. iOO Marketing Ltd., London.

Table 1. Mean (and standard deviation) of the number of errors made in sight reading the scores, and the time taken in doing so, shown separately for the two types of score printed with lines 0.1mm and 0.4mm thick respectively.

Stave line thickness (mm)	Score	Number of errors	Duration (s)
0.1	G major	0.9 (1.2)	190 (40)
	Chromatic	1.9 (1.2)	192 (39)
0.4	G major	2.1 (1.4)	194 (44)
	Chromatic	3.8 (1.8)	197 (44)

## Figure legends

Figure 1. A-D Stripes with thick and thin lines. E. Illustration of a pattern with a spatial frequency of 2 cycles per degree; F. Illustration of the concept of duty cycle,  $(w/c) \times 100\%$ ; G. Illustration of the measurement of pattern size: the pattern is fixated centrally, subtends 3 degrees at the eye and has a radius of 1.5 degrees.

Figure 2. A. The mean number of illusions reported (solid curve) and the proportion of participants reporting illusions (broken curves) in two experiments. B. The threshold contrast at which large (lower curve) and small (upper curve) target letters can just be seen on a grating background. Both illusions and threshold contrast are shown as a function of the spatial frequency of the grating. C. The mean number of illusions reported (solid curve) and the percentage of participants reporting illusions (broken curve) shown as a function of the duty cycle of a grating. D. The threshold contrast at which a target symbol (an oblique line) can just be seen on a grating background, shown as a function of the duty cycle of the grating. E. The number of illusions reported when observing patterns. The patterns were discs (solid points), or annuli (open points) of vertical stripes, or sectors of concentric rings (crosses). F. The threshold contrast at which a target letter can just be seen on a grating background, shown as a function of the size of the grating. The upper curve is for small target letters and the lower curve for large. In both E and F, the horizontal axis shows the size of the pattern expressed as a proportion (Q) of the area of the visual cortex to which the pattern projects.  $Q=100[1-\exp(-0.0574A)]$ , where A is the angular radius of a circular region, centrally fixated (Drasdo, 1977). Also shown is the radius of a centrally fixated pattern that projects to the same area of cortex, expressed in terms of the angle subtended at the eye. A, C and E are based on data from Wilkins et al (1984); B, D and F are based on data from Chronicle and Wilkins (1999).

Figure 3 Histograms showing (A) the height of the stave and (B) the thickness of the lines of the stave in 63 musical scores of popular music for keyboard published between 1957 and 2008 by 50 different publishers.

Figure 4. Opening lines of scores printed on staves with lines that are thin (A and B) or thick (C). The notes are random sequences in G major (A and C) or on a chromatic scale (B). Note that the images have been scanned and are not reproduced actual size.

Figure 1

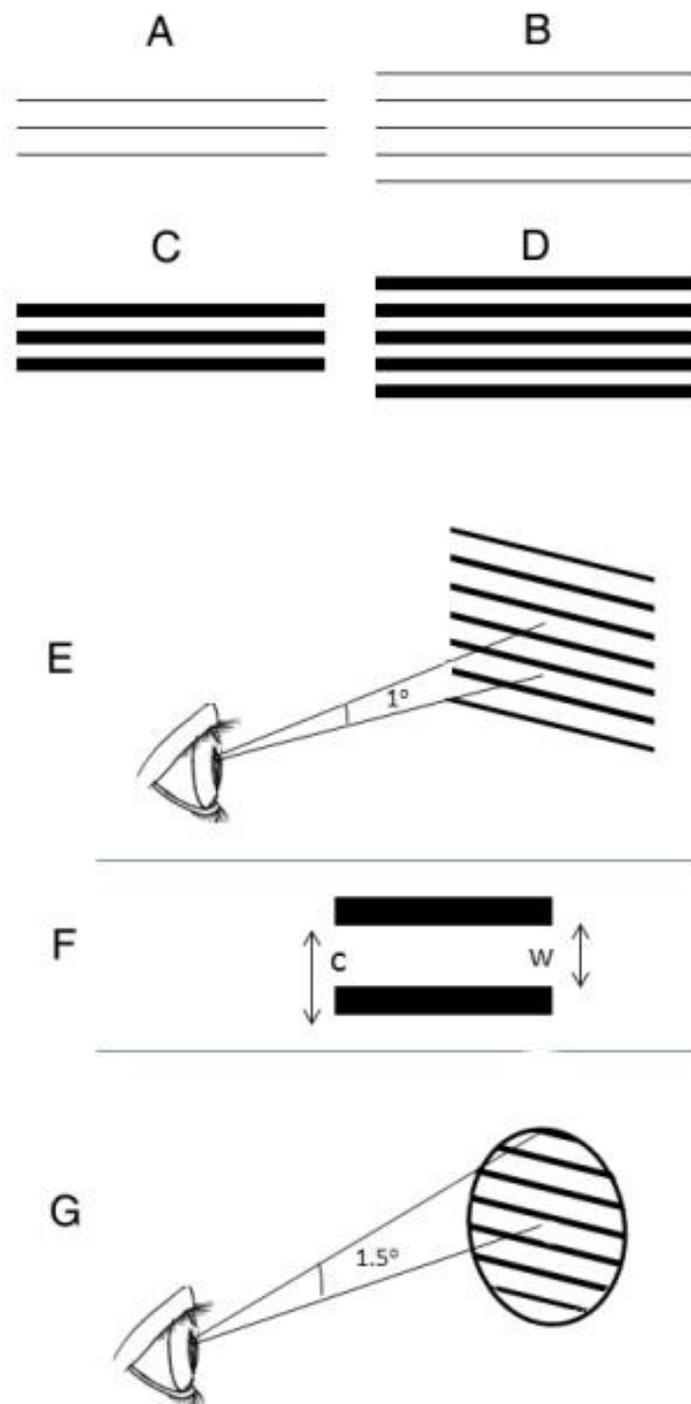


Figure 2

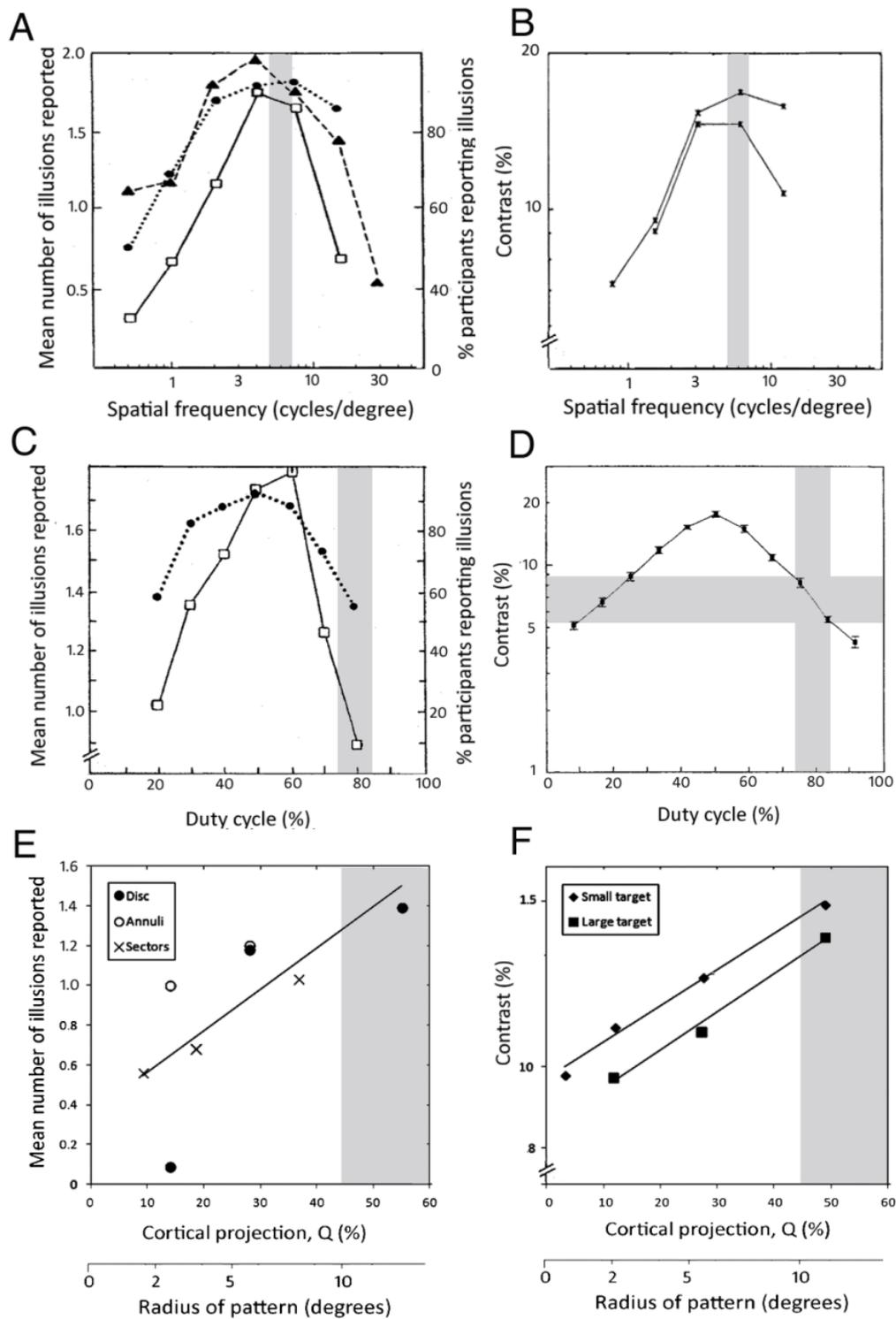


Figure 3

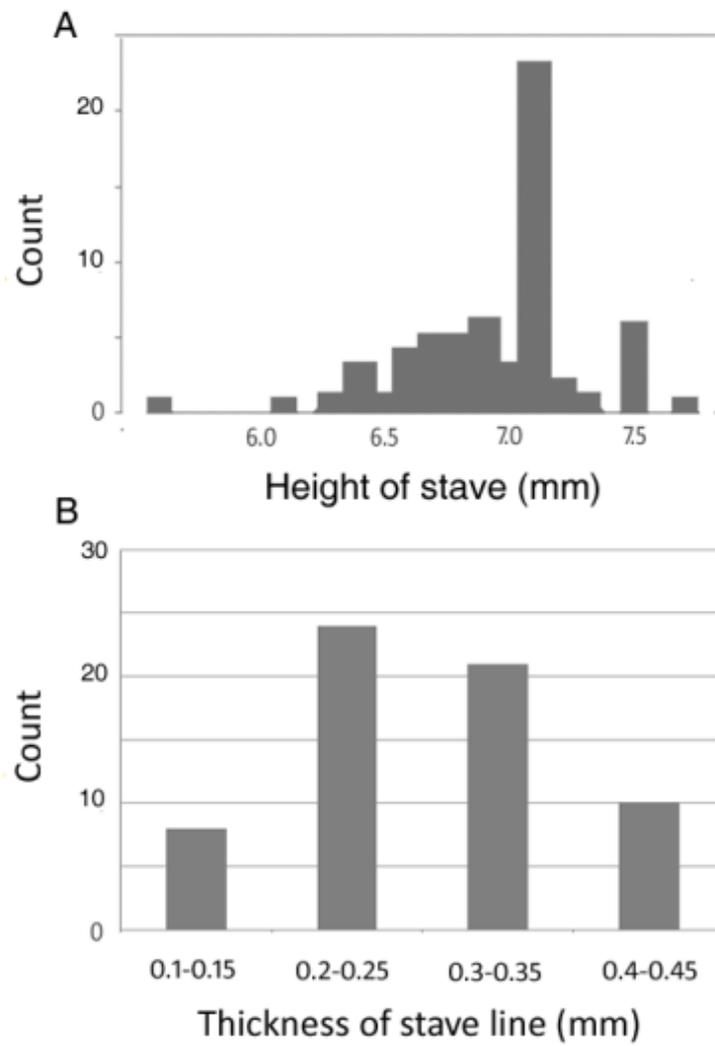


Figure 4

Figure 4 displays three staves of musical notation, labeled A, B, and C, arranged vertically. Each staff begins with a treble clef, a 4/4 time signature, and a key signature of one sharp (F#).  
Staff A: The melody starts on G4, moves to A4, then B4, and continues with a series of eighth and quarter notes, ending on G4.  
Staff B: The melody starts on G4, moves to A4, then B4, and continues with a series of eighth and quarter notes, ending on G4.  
Staff C: The melody starts on G4, moves to A4, then B4, and continues with a series of eighth and quarter notes, ending on G4.