Phonesthetic "Pointiness" in the Auditory and Visual Modalities

Shelly Wynecoop

1998 Trial Research, Department of Psychology, Cognition and Communication

Professor H. Nusbaum
Professor D. McNeill
Professor T. Regier

Abstract

Twenty-eight subjects were asked to provide ratings of "fit" for comparisons between consonant and vowel pairs and two-dimensional shapes. The analyses of this fit rating data provide strong evidence for the contested existence of consistent across-subject phonesthesia. Particular hypotheses were experimentally examined which concerned the perceived "roundness" or "pointiness" of a set of representative consonant sounds. These hypotheses were found to describe significant differences in the data. As an improvement over the precedent of subjective labeling, computationally derived measures of shape were used. The shape generation software originated for this experiment provided the complex visual stimuli necessary for the investigation of sound-shape synesthesia -- stimuli which were up to this point difficult or impossible to create. The measures of shape employed here are shown to be effective predictors of subjects' fit rating choices, and have helped to make the phenomenon of phonesthesia methodologically tractable. Additional attention to the auditory stimuli at the level of the distinctive feature provides some insight into the phenomenon in question, and allows for more precise descriptions of sound-shape association than have been attempted in the past.

Introduction

Throughout this paper, the term *phonesthesia* will be used to refer to a widely occurring, i.e., non-clinical, form of synesthetic phonetic symbolism (sometimes also called phonosymbolism). The term *phonetic symbolism* is often used to refer to a group or individual's intuitive associations between sub-morphemic linguistic sounds or arrangements of these sounds and domains as diverse as emotions, the physical properties of objects, or social categories. In using the term phonesthesia here, however, we are referring only to a small subset of the aforementioned possible associations. We will devote our attention entirely to the perceived similarity (or dissimilarity) between sub-morphemic linguistic sounds and two-dimensional visual forms.

The inquiry into phonetic symbolism is know to have begun before the time of Aristotle, when the Greek philosopher Democritus undertook speculations on the nature of this form of cross-modal association. (Stratton, 1917) In this century, the psychological community's interest in phonetic symbolism became revitalized when Köhler (1929) asked us to "match the nonsense words *takete* and *maluma* with the two patterns shown (See Figure 1), and observed that "most people answer without hesitation." (p. 224)

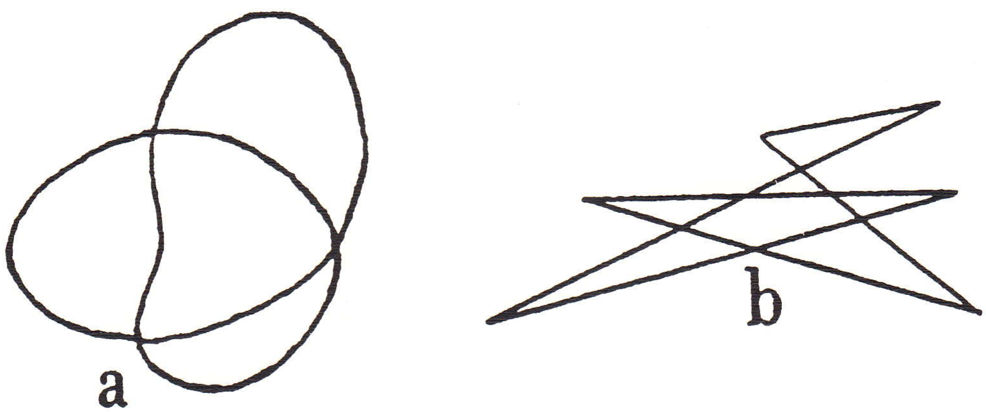


Figure 1. The figures *maluma* and *takete* as presented by Köhler in 1929. If you had to name one of the shapes *maluma* and the other *takete*, which shape would receive which name?

The presentation of *maluma* and *takete* provided the seeds for a favorite hypothesis among the many new researchers in phonetic symbolism who sought to answer not only Köhler's "what" question, but also his implied "why?" This hypothesis took the rough form: /m/, /l/, and /b/ (Köhler changed *baluma* to *maluma* for the second edition of Gestalt Psychology) are perceived as "round" sounds, and sounds like /p/, /t/, and /k/ are perceived to be more "pointy." (e.g. Wescott, 1971; Oyama and Haga, 1963; also Uznadze, 1924)

Modern researchers in cross-modal association may have had early interests in sound-to-shape mappings, but by far the greatest proportion of studies done on cross-modal similarity to this date have involved the connection between (auditory) loudness and (visual) brightness. (Marks, 1987). The reason for this emphasis in the allocation of research can be traced to the ease of measuring the two domains of the association in question. Loudness and brightness can rely upon psychophysical methods for describing their unidimensional variation, putting mathematical comparisons easily within reach. Linguistic sounds are not so readily mathematized, and in phonesthetic research they have generally been approached as indissectable atoms, or gestalts, an approach which provides no reliable basis for the comparison or differentiation of experimental sound stimuli. Although attempts have been made to classify sounds according to the properties of their formants and Fourier spectra, a description of speech sounds at the level of the distinctive feature offers easily graspable advantages for sound comparisons for the purpose of phonesthetic inquiry. (Jakobson, 1987). However, Jakobson's critique of research into phonetic symbolism, and his urging for analyses on the level of the distinctive feature have yet to be recognized by researchers in field.

The implementation of useful, consistent analyses for two-dimensional visual stimuli in phonesthetic research has been even less encouraging. Investigations into sound-to-shape associations have typically relied upon subjective accounts of the visual stimuli such as in Lundholm, 1921: "...a few long and low waves, ... a few high waves of medium length, ... numerous small waves of varying shape, ... a few obtuse angles, ... a few approximate right angles, ... numerous acute angles..., p. 45) and have hardly improved over the decades, for example, Daniel and Toglia's "toward and away variations" on shape skewness (p.471). These problems concerning the level and type of analyses used by phonesthesia researcher have been the primary obstacles to obtaining dependable results. An additional problem consists in the lack of standardization for auditory experimental stimuli, which, without exception have been either spoken to the subject by another individual (e.g.: Svartdal and Iverson (1989)), or presented in ambiguous (and visually confounding) written forms (e.g.: Cohen and Izawa (1976). The experiment discussed here attempts to overcome all of these difficulties, and provide a positive template for further research of this nature.

A minor corrective addition to previous experimental attempts was the introduction of a voice synthesizer for the presentation of the auditory stimuli. However, the primary tool for making this experiment possible was the creation of software capable of generating and analyzing visual stimuli. The introduction of multiply-definable visual materials to the study of sound-shape phonesthesia allows not only for a more exacting mapping from linguistic sounds to pertinent details of the visual stimuli, but provides the opportunity to systematically discover which details of the visual stimuli are actually important for any given phonesthetic association. At present, the software allows for the creation of shapes conforming any specifiable level along at least 38 morphological[1] axes; for this experiment we have examined four of those factors.

Although the development of a robust methodology for the study of phonesthesia was one of our goals, the motivations for this study neither began, nor ended at methodological acumen. The attempt to employ methods leading to clearer and more systematic associative descriptions was, for us, entirely practical. It is what one can *do* with a good model that is interesting, not necessarily the existence of a model per se -- the model being sought here could provide the opportunity for a number of creative computational interactions. Sound is a modality that is underutilized in computational interaction, particularly on the level of input. By taking advantage of widely existent psychological associations sound could be more seamlessly with the visual modality of today's computational multimedia. The applications that are of the greatest interest to us are creative, but other uses come readily to mind, such as naming and marketing protocols (having a phonesthetically "correct" labels for shapes or ideas improves people's memory for them, for instance, see Cohen and Izawa (1976), Denes, G. and Spinaci, M. P. (1981)). This experiment represents a first step toward the goal of a full model for a particular type of sound to shape association. Further research will be devoted to the greater elaboration of that model.

[1] The term "morphological" and its cognates will be used only in the description of shape from this point on, and should not be understood here as a linguistic designation.

Materials

The complex visual stimuli necessary to carry out this experiment required the engineering of computer software capable of generating and analyzing parametrically-defined, two-dimensional visual forms along particular factors of interest. The software begins by generating a number of radial spokes comprising the "skeleton" of the shape. These spokes vary randomly within limits on their number, length, and regularity of orientation. At the distal ends of these spokes is an angle of randomly chosen size and vector strength specifications for each leg of that angle. The angle and vector specifications are the same for each spoke within any particular shape, and determine the placement and strength of cubic Bezier (spline) control points. See Figure 2.

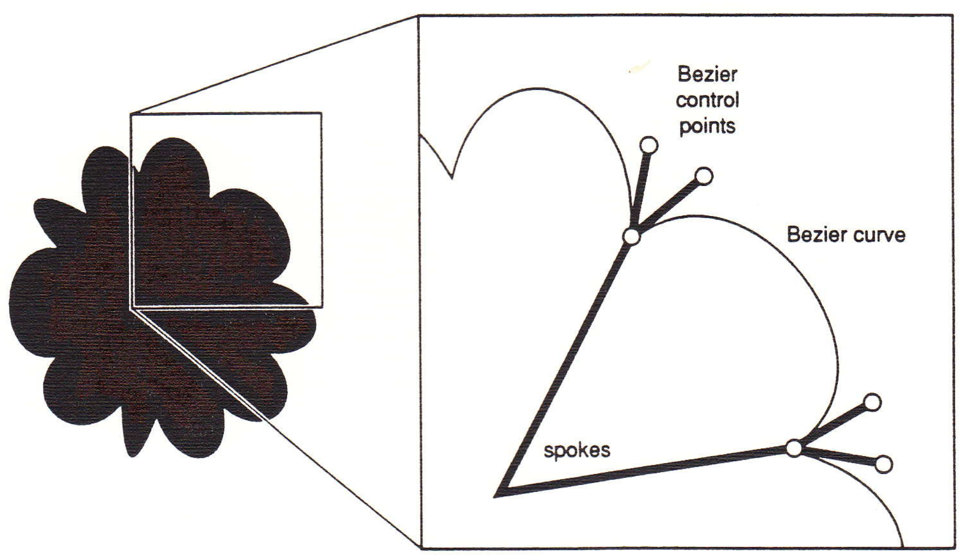


Figure 2. A representation of the factors involved in the generation of the shapes used in this experiment. A number of spokes comprising the "skeleton" of the shape are generated which vary randomly within limits on their number, length, and regularity of orientation. At the ends of each of these spokes are angle size, and vector strength specifications which determine the placement and strength of Bezier spline control points. The control points provide the information necessary to create curves which connect each spoke to the next, forming the perimeter of the shape. Shapes whose perimeters intersect themselves are discarded. Shapes can be generated to conform to 38 distinct, computationally-derived morphological characteristics.

The control points provide the information needed to create Bezier splines (computationally generated curves) which connect each spoke with the next to form the perimeter of the shape. Shapes whose perimeters are self-intersecting are discarded, due to the fact that the resultant "loops" create emergent perceptual features that are difficult to account for computationally. The (ideal) mathematical curves that constitute the perimeter splines are densely sampled and thereby represented in the software's data structures by large numbers of extremely short, straight line segments. The perimeter of the shape is then filled in black, and scaled so that the largest diameter of any shape extends no more, or less than 270 pixels (approximately three inches). The number of different, valid (i.e. non-self-intersecting) shapes that can be created using these procedures is theoretically unlimited, and each shape can be generated to conform to, or be later analyzed along 38 distinct computationally tractable morphological axes. For images of the shapes created for this experiment, see Figures 4 and 5.

Method

Participants. Twenty eight students from the University of Chicago (11 women and 17 men) participated in this experiment for ten dollars each.

Stimuli. The stimuli used in this experiment were of two types: 1) consonant-vowel and vowel-consonant sounds, and 2) two-dimensional shapes. The sounds chosen for experimentation were the consonants m, l, b, p, t, and k, paired (in consonant-vowel and vowel-consonant orders) with each of the following four vowels: /ɪ/ as in hit, /ɛ/ as in bet, /a/ as in pop, and /ʌ/ as in cup. More than one vowel was chosen primarily to help keep the participants' attention over the large number of trials, but vowels were additionally screened to minimizing the occurrence of consonant and vowel pairs homophonous with extant English words. For these reasons the four vowels listed above were used in the construction of the sound stimuli. The full set of possible consonant-vowel/vowel consonant combinations (minus sounds homophonous with English words, for example, "up," and "el") were used in the experiment.

The shapes used in this experiment were generated for their presence at the extreme ends of particular measures of interest. This approach in the selection of experimental shapes was taken in order to allow for the initial substantiation of the phenomenon of phonesthesia in the clearest morphological cases, before moving on to more subtle examples in later experimentation.

The verbal characterizations of "pointiness," or "angularity," which have been used in past literature on the topic of phonesthesia to describe both shapes and sounds, were captured by at least three measures of shape available to us. We characterized the shapes according to three metrics: 1) circularity, 2) overall curvature, and 3) the shape's area to perimeter ratio. A fourth measure, the orientation of a shape's principle axis was included as a control variable, and was not expected to be a significant predictor of sound-shape associations in this experiment. An explanation of each of these measures follows.

The first measure that seemed to get at the idea of visual pointiness was a shape's "circularity". The circularity value for a given shape was determined by calculating the standard deviation of the distances from the shape's center of mass (centroid) to sampled points on the shape's perimeter.[2] See Figure 3. The degrees of circularity of interest to us in this experiment involve shapes at the highest and lowest ends of the circularity spectrum. Shapes with low circularity values are the most circular, as the deviations from the average distance of the edge points from the center of mass are minimized. Rather than using the terms "low" and "high" circularity to refer to the more opaque circularity value itself, to avoid confusion, shapes with low circularity values will be called "high circularity shapes" in accordance with the shapes' appearance. Likewise, shapes that receive high circularity values will be called low circularity shapes (following morphological, not mathematical

---

[2] If d_i represents the distance of any given point on a shape's edge from that shape's center of mass, and d-bar is the average distance from the center of mass of all the d_i perimeter samples taken, then the circularity value is calculated using the following equation: $(\Sigma(d_i - \text{d-bar})^2)^{1/2}$.

standards of description). Regardless of the labels applied to various segments along the circularity continuum, a shape's presence at one end as opposed to the other was entirely mathematically defined.

The second measure of relevance in our investigation into "pointiness," is a characterization of a shape's curvature. Conventionally, curvature is defined to be the reciprocal of the radius of curvature, where the radius of curvature is the length of the radius of the circle which conforms to the curve segment. This standard equation results in curvature assessments for single, monotonic segments only, and has a value output range from zero to infinity. Seeking a curvature metric which would be responsive to salient perceptual differences between shapes as a whole (non-monotonic curvature), and which would also produce more statistically-manageable values, we instead developed a measure computed as the third moment of the angles between the perimeter segments.[3] See Figure 3. This metric -- which we have, for the sake of understanding, named "curvature" -- is more suitable for the analysis of the discrete (sampled) data being collected here, and also provides a valenced output limited to the range 0 to 360. The third moment was chosen over the standard deviation to retain valence information, which was useful for computationally distinguishing "star" shapes (with negative curvature values) from "flower" shapes (which have positive curvature values) -- a salient perceptual difference. In addition to the positive and negative extremes of the curvature measure, the central region -- those shapes that return curvature values that are close to zero -- described a third, perceptually distinct area for investigation. For this reason, three levels of the curvature measure were represented in the experimental shapes: those with positive, neutral, and negative curvature.

Curvature and circularity were chosen as axes of criteria for shape selection due to the high sensitivity of each of the measures to salient perceptual detail, while, additionally

---

[3] If a_i represents the angle created by any given pair of perimeter segments sampled on a shape's edge, and a-bar is the average angle created by such perimeter samples, then the curvature value is calculated using the following equation: $(\Sigma(a_i - \text{a-bar})^3)^{1/3}$.

maintaining mutually independent variation as descriptors of what appears to be the same broad perceptual phenomenon: that is, the quality termed "pointiness".

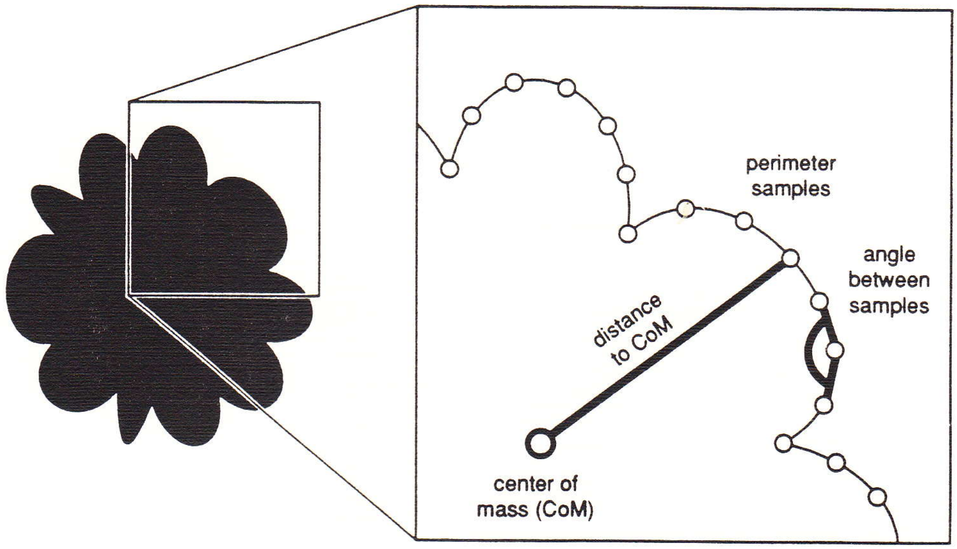


Figure 3. A representation of the factors involved in the calculations of the circularity and curvature measures. For circularity, the distance from sample points along the perimeter to a shape's center of mass is found, and the standard deviation of the distance of these points to the center of mass computed. For curvature, the third moment of the angle between boundary segments is calculated to provide a measure of a shapes overall angular character.

The third measure of shape is the ratio of the area to the perimeter. This measure was chosen for its descriptive power and computational ease. Although the area/perimeter measure does not vary independently of the circularity or curvature measures, it was possible to avoid the selection of shapes recommended by more than one factor: the shapes used in the experiment had either low or high circularity; or negative, neutral, or positive curvature; or low or high area to perimeter ratios. To maintain the independence of the measures of shape to the greatest extent possible, shapes chosen to fall in one of these categories were specifically chosen not to fall in any of the other categories.

Five shapes were generated for each of the aforementioned levels of circularity and curvature, and two were generated for each level of the area to perimeter measure, for a total of 29 shapes. In addition to these shapes, eight more "average" shapes were added, which did not fall into any of the above categories. From these additional eight, four were chosen which had the most extreme values on a measure of "orientedness" (a measure computed from a shape's principal components, i.e., from the eigenvectors of a shape's tensor matrix), and each of these four was manipulated to "point" vertically, horizontally, and at 45 degrees, resulting in an additional set of 12, for a total of 45 experimental shapes. See Figures 4 and 5.

Hypotheses and design. This last set of 12 shapes allowed for the testing of the orientation factor as a predictor of fit rating, a factor hypothesized to have no significant effect on the mapping of shapes to sounds such as m, l, b, p, t, and k. On the other hand, we hypothesized that the circularity, curvature, and area/perimeter measures should be significant predictors of the fit ratings subjects choose relative to the sound presented. More particular hypotheses were, first, that low circularity shapes, when compared with sounds containing p, t, or k would result in high fit ratings, but when compared to sounds containing m, l, or b would result in low fit ratings, and vice versa for high circularity shapes. Second, that negative curvature shapes, when compared to sounds containing p, t, or k would result in high fit ratings; low ratings when compared to sounds containing m, l, or b, and vice versa for shapes with neutral curvature. For shapes with positive curvature, the hypothesis was uncertain. Would subjects pay as much attention to the internal points of the "flower" shape as they appear to attend to the external points of a "star" shape when evaluating the fit to sound? Our hypothesis was that the subjects would take less note of these internal angles than the external ones, thus, that there should be a reversal in fit ratings from negative to positive curvature shapes, but that the angles may still have an effect, making positive curvature shapes "pointier" than shapes with curvatures near zero. For the area/perimeter measure, we hypothesized that the low area/perimeter shapes would be rated to have a high fit to p, t, and k sounds, low fit ratings for sounds containing m, l, and b, and vice versa for shapes with a high area/perimeter value.

The consonant and vowel pairs and the shapes were presented to the subject randomly without replacement until all the stimuli of one of the types was entirely used, at which point all tokens of the sound or shape type were again made available to be chosen randomly for presentation.

The experiment consisted of 4 blocks of 100 trials apiece. The only difference between blocks was in the sound data: each block consisted entirely of either consonant-vowel or vowel-consonant auditory stimuli. The order of blocks for each subject was randomly assigned.

Apparatus. The experimental software a self-contained application developed in Microsoft Director (this was not the same software which was used to generate the shape stimuli). The experiment was conducted using 120MHz Power Macintosh computers with 8" x 10.5" screens set to 832 x 624 pixels, 75 Hz. At this setting the experimental shape stimuli subtended a maximum[4] visual angle of approximately 9.5°. The auditory stimuli were presented by the computer through headphones, and were generated by Apple's Speech-to-Text voice synthesizer in a plain, baritone voice ("Fred").

---

[4] The maximum value is given as shapes vary in their extents along various axes, but do not exceed 3 inches along any dimension.

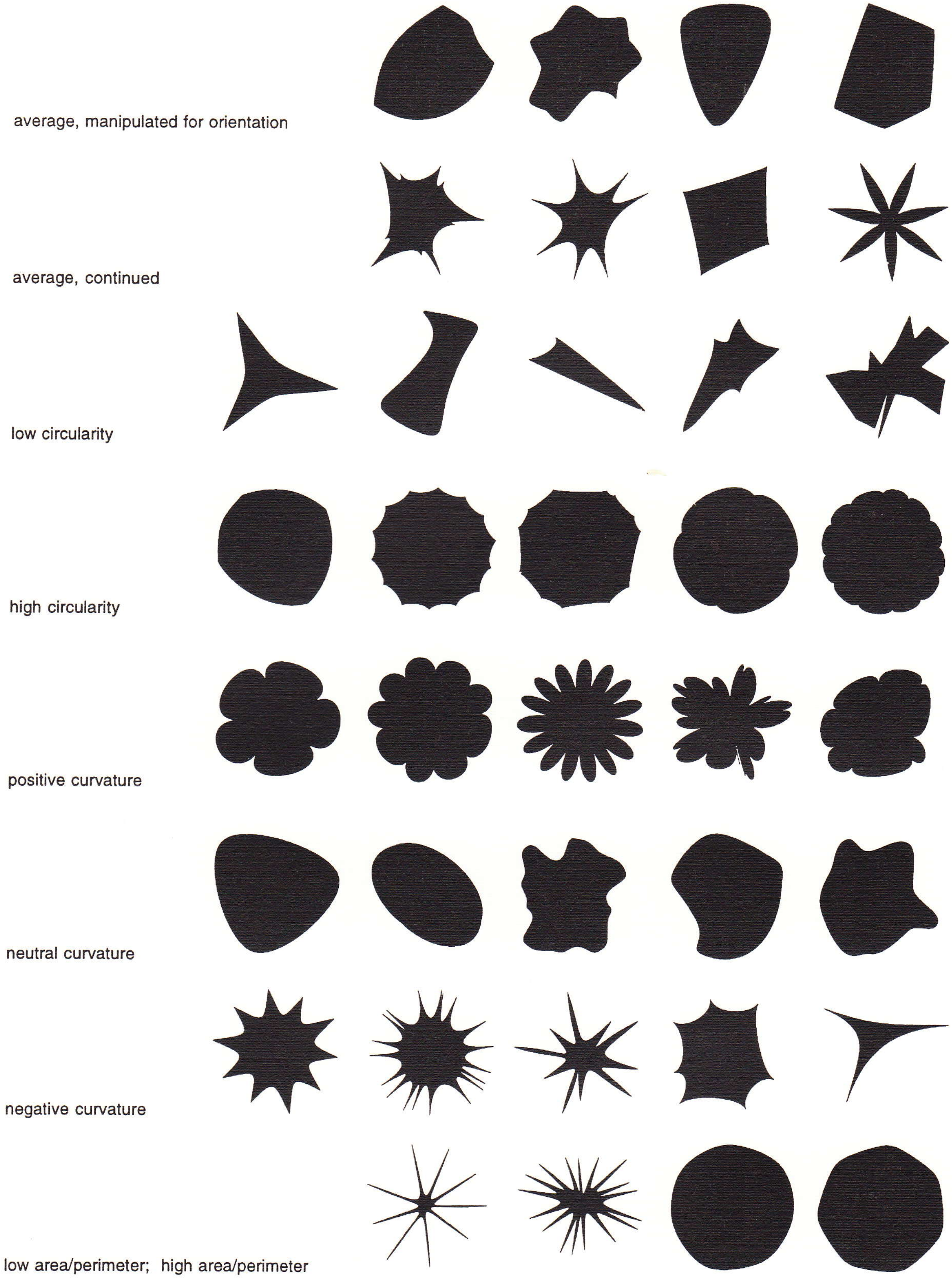


Figure 3. Partial company of shapes used in the experiment. The first row of shapes was manipulated to create a set designed to test subjects' sensitivity to orientation in rating fit to auditory stimuli containing m, l, b, p, t, or k. The four, as pictured here, were not used in their original orientations. See Figure 4. Average shapes were generated to fall outside the selection criteria for the curvature, circularity, and area to perimeter ratio levels.

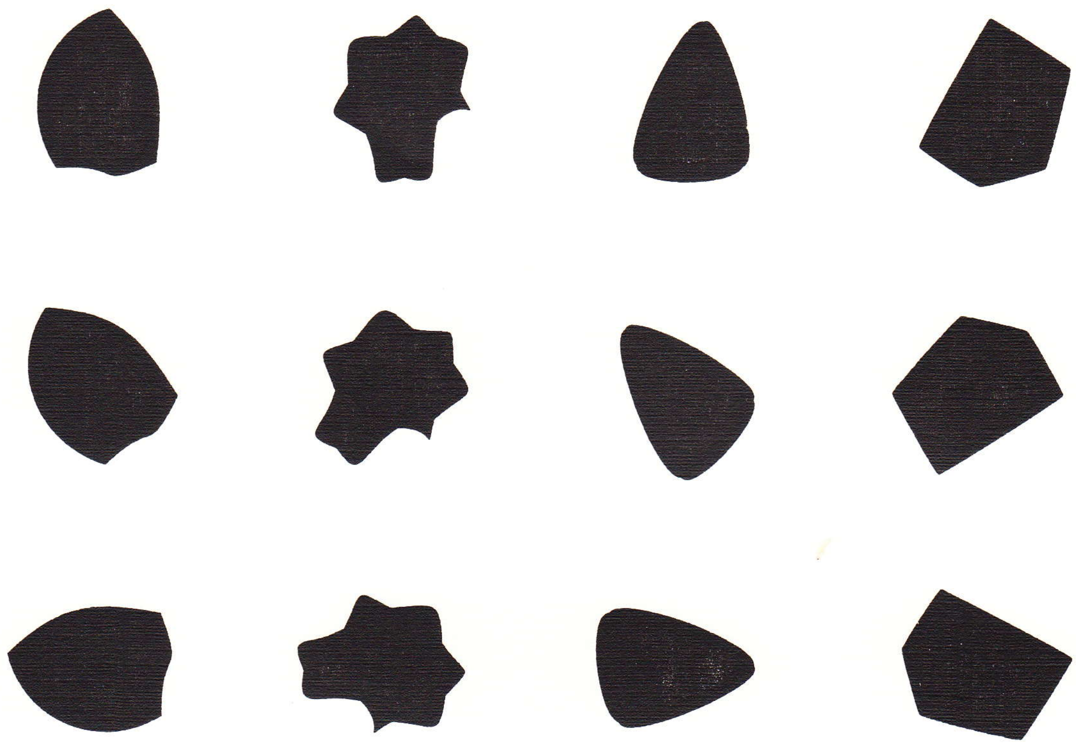


Figure 5. The remaining 12 experimental shapes. This set results from the manipulation of the orientations of the first four average shapes, which these twelve then replaced.

Procedure. See Figure 6. In each of 400 trials, two shapes were presented on either side of an on-screen button, which the participant clicked with the mouse. When this button was pressed the subject heard a consonant-vowel, or vowel-consonant pair generated by the voice synthesizer from the set of auditory stimuli previously detailed. Beneath each of the two shapes presented there was a rating slider, which was also mouse-driven. The rating sliders were numbered on a seven-point scale from -3 to 3, and participants manipulated the sliders to generate numerical responses corresponding to the nature of the perceived fit between the sound heard, and the shape presented above that slider. No fit rating could take place until the sound button was pushed, and the sound

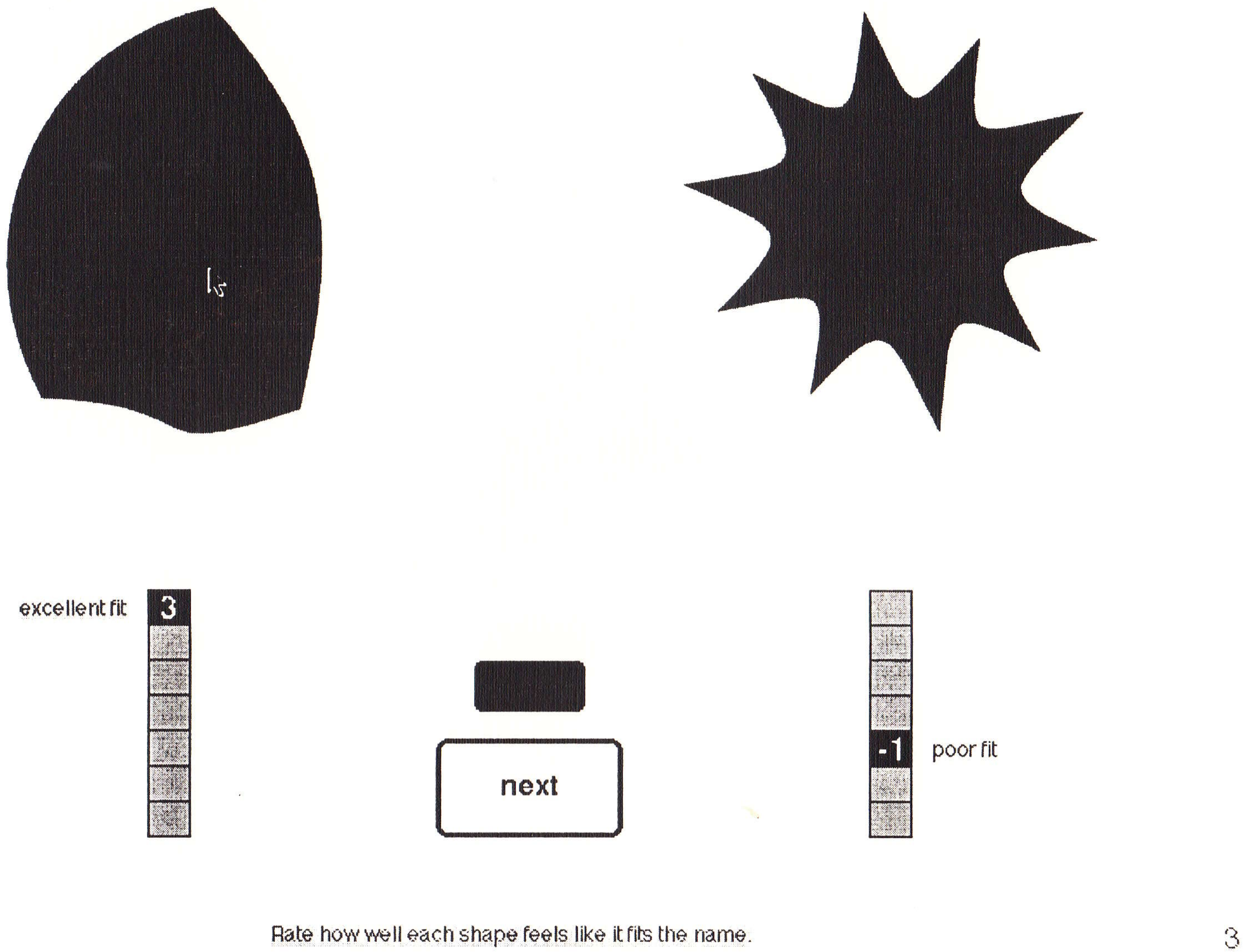


Figure 6. A representative trial of the experimental setting. The black rectangle in the center is the mouse-driven button for the auditory stimuli. No rating can take place until this button has been pressed with the mouse, and the sound heard. Each of the rating sliders beneath the shapes lights up, along with its verbal tag when chosen. Once both shapes have been given non-identical ratings, the [next] box appears, which allows the subject to move to the next trial. The trial number is shown in the lower right corner so subjects can track their progress.

presented. The subjects could listen to the sound as many times as they liked while choosing ratings for the sound-shape fits. Subjects could change their ratings until they were satisfied with them. When a rating was chosen, the corresponding number and verbal tag lit up. Verbal tags for the rating sliders were as follows: -3: terrible fit; -2: bad fit; -1:poor fit; 0: no fit; 1: slight fit; 2: good fit; 3: excellent fit. Once both shapes had been given non-identical ratings for their fit to the sound presented, a button labeled "next" would appear, which, when clicked upon, allowed the subject to move to the next trial.

Participants were tested two at a time on separate computers. They were given a brief tutorial on using the mouse-driven interface to manipulate the sound button and rating sliders over two trials which were not included in the experimental results. Once subjects were comfortable with the form of the interface, their only instruction concerning the content of their responses was that they should rate how well they felt each shape on screen "fit" the sound they heard for that trial. The experiment lasted approximately 45 minutes.

Results

The data sets {mlb} versus {ptk}.[5] Using logistic regression, the data were analyzed within each shape classification set (negative curvature, neutral curvature, positive curvature, low circularity, high circularity, low area/perimeter, high area/perimeter) for systematic rating differences between trials involving comparisons with auditory stimuli containing an m, l, or b, versus trials using stimuli containing a p, t, or k sound. The rating given by the participants to each shape-sound comparison was the dependent variable. The particular rating given (from -3 to 3) was analyzed relative to its occurrence with sounds containing m, l, or b, versus those containing p, t, or k. There were significant systematic differences within all levels of all shape factors ($p < .0001$ for all seven analyses) between fit ratings given to {mlb} auditory stimuli versus {ptk} stimuli

[5] The notation {mlb}, {m}, and {mlb}-vowel, vowel-{mlb} (for example) will be used to denote respectively: all auditory stimuli containing an m, l, or b sound; all auditory stimuli containing an m; all

when compared to shapes within each level. Shapes with negative curvature, low circularity, or low area to perimeter ratios were consistently judged to be a better fit to consonant and vowel dyads that contained the sounds p, t, or k, and conversely, shapes with any of these same characteristics were judged to be worse fits to auditory stimuli containing m, l, or b sounds. Experimental shapes exhibiting any of the remaining factors (neutral or positive curvature, high circularity, or a high area to perimeter ratio) were judged to fit sounds containing m, l, or b significantly better ($p < .0001$) than stimuli containing p, t, or k sounds.

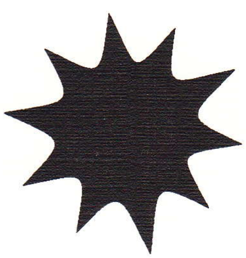
The rate of change as given by logisitic regression gives us information on how extreme the {mlb} vs.{ptk} rating divergence is as you move from the -3 to 3 (terrible to excellent) fit rating. For example, in the case of the negative curvature shapes, see Table 1, we can see that the odds of receiving any particular fit rating for an {mlb} to negative curvature shape comparison decreases by an average of about 45% (relative to comparisons involving {ptk}-containing sounds) for each increment from -3 to 3 along the rating scale. Receiving a particular rating for such a comparison ({mlb} to negative curvature) becomes multiplicatively less likely as we move up the fit scale, by a factor of .45. Simply put, {mlb}-containing sounds are significantly more likely to be rated a worse fit to negative curvature shapes than {ptk}-containing sounds. See Figures 7 - 13.

---

auditory stimuli of the consonant-vowel form, where the consonant is either an m, an l, or a b; all auditory stimuli of the vowel-consonant form, where the consonant is either an m, an l, or a b.

Shape parameter and level	Weighted average fit rating {mlb}	{ptk}	Rate of change (logistic regression)	n	p-value
negative curvature	-0.66	0.98	-45.12%	2367	<.0001
neutral curvature	0.84	-0.60	78.14%	2196	<.0001
positive curvature	0.60	-0.55	56.73%	2270	<.0001
low circularity	-0.46	0.79	-36.91%	2298	<.0001
high circularity	0.30	-0.17	19.38%	1833	<.0001
low area/perimeter ratio	-1.15	1.09	-50.08%	866	<.0001
high area/perimeter ratio	0.69	-0.80	64.31%	888	<.0001

Table 1. The results of the analyses of differences in fit ratings for auditory stimuli containing the sounds p, t, or k, versus sounds containing the consonants m, l, or b. The rate of change figures for this table are interpreted with respect to the set {mlb}: for example, for shapes of negative curvature, ratings of fit given to comparisons with sounds containing an m, l, or b are increasingly less likely as we move from -3 (terrible fit) to 3 (excellent fit) along the rating scale by a factor of about .45 relative to the occurrence of ratings for {ptk}-containing auditory stimuli when compared to the same shapes. Weighted averages are included as a heuristic for comparison, but did not figure significantly in the analyses, which provide a more elaborated measure of comparison.



sample shape

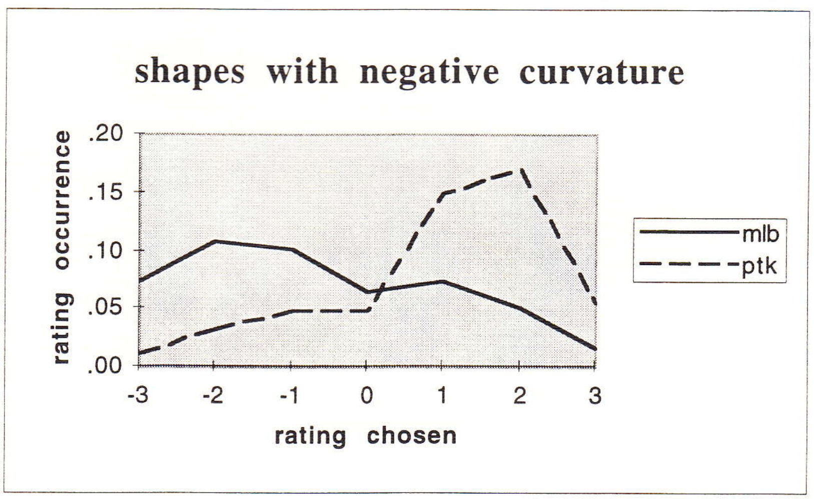
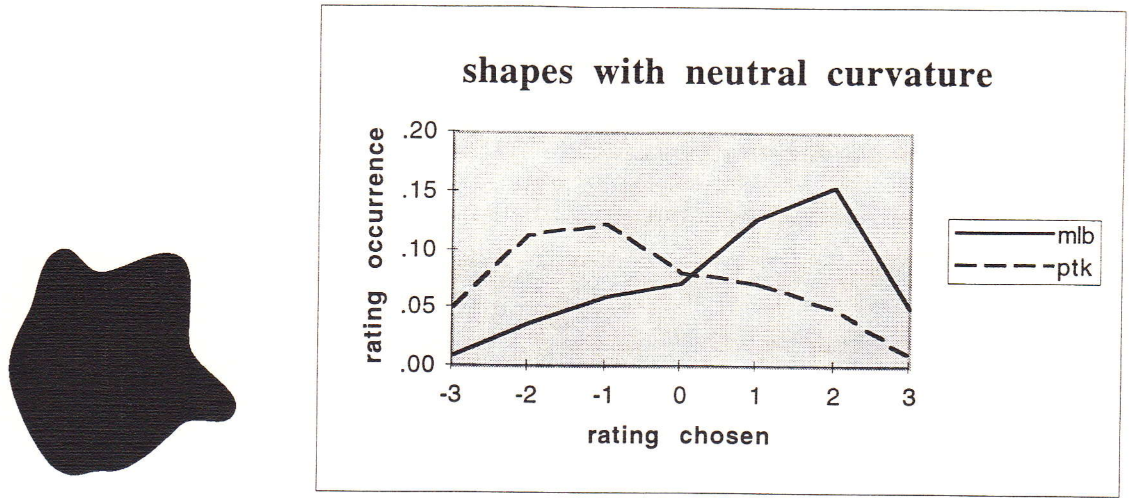
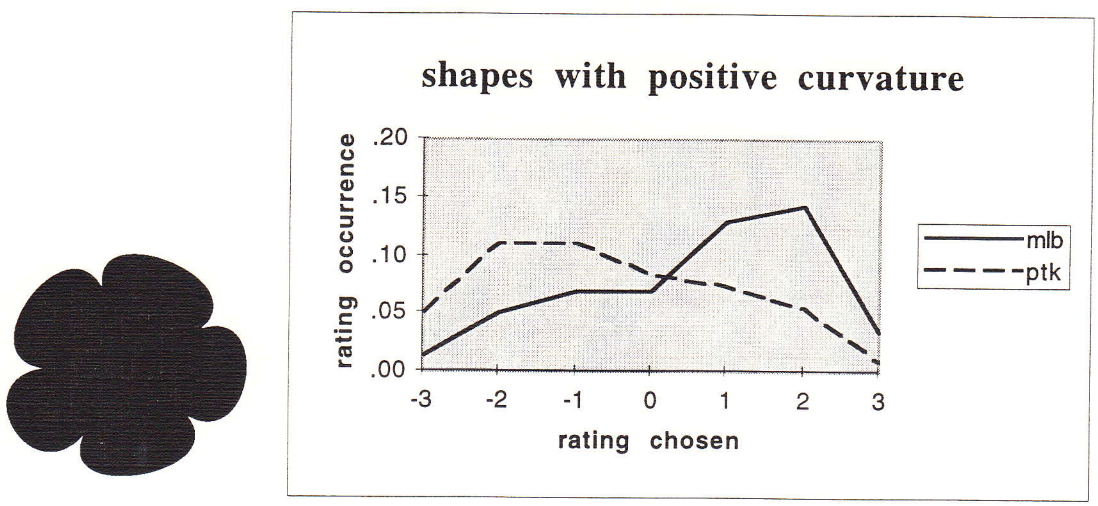


Figure 7. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes of negative curvature, versus {ptk}-containing sounds when compared with the same shapes. For this logistic regression analysis the odds of receiving a particular {mlb} to negative curvature comparison fit rating decreases by an average of 45% for each increment in the rating scale as one moves from a -3 to a 3 rating (from a "terrible" to an "excellent" fit), $n = 2367$, $p < .0001$, that is, the sounds represented by m, l, and b are considered to be a bad fit to shapes with negative curvature, relative to their auditory counterparts, which contained a p, t, or k..



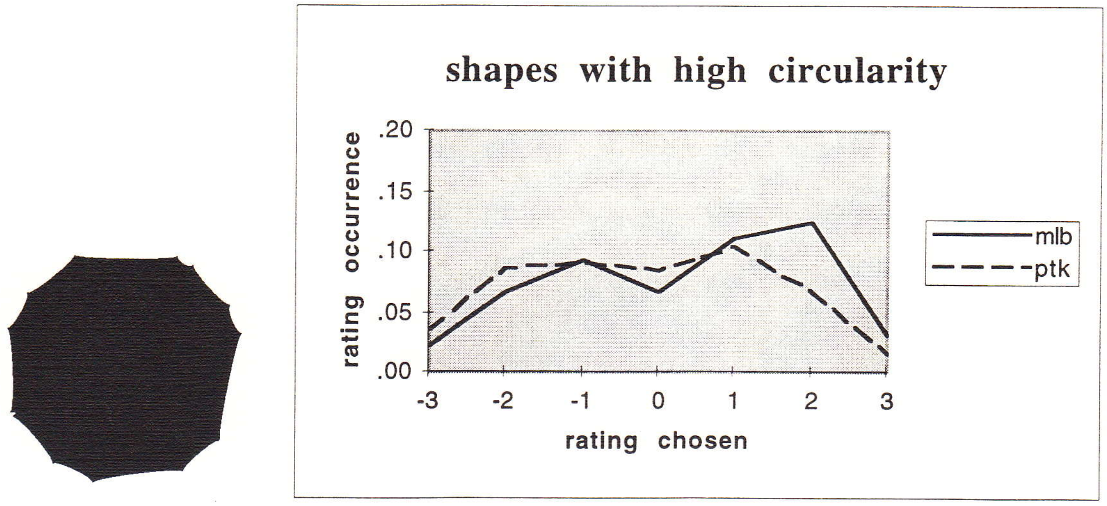
sample shape

Figure 8. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes of neutral curvature, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to neutral curvature comparison fit rating increases by an average of 78% for each increment in the rating scale as one moves from a -3 to a 3, n = 2367, p < .0001. The appearance of a consistent but non-significant fall-off in rating occurrence at 0 (no fit) throughout all of the following analyses may indicate that subjects were less willing to select a rating which would suggest a complete lack of comparability between sound and shape. The unusual nature of the comparisons subjects were asked to make did not result in responses indicating the possible, basic incomprehensibility of the comparison



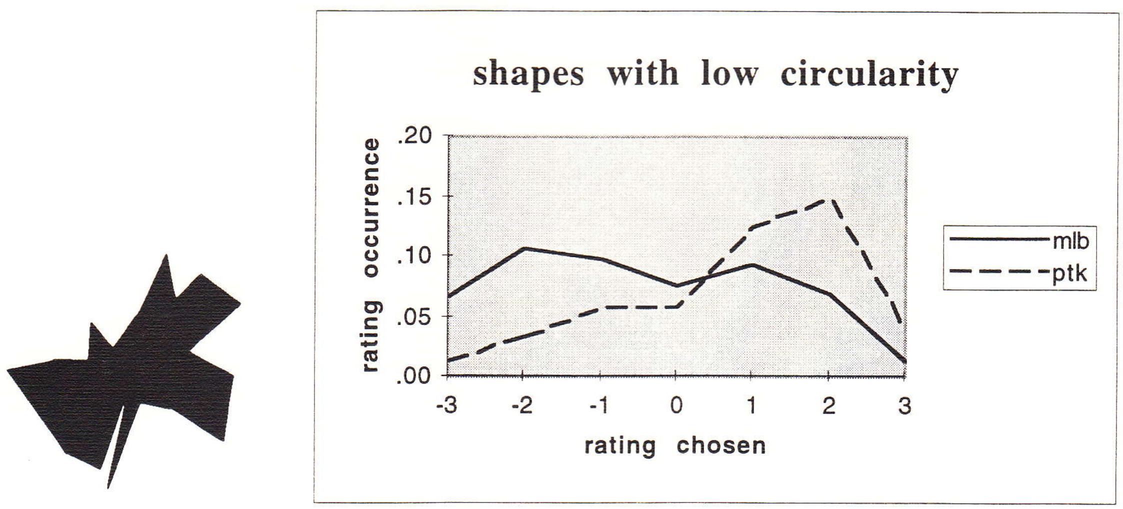
sample shape

Figure 9. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes of positive curvature, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to positive curvature comparison fit rating increases by an average of 57% for each increment in the rating scale as one moves from a -3 to a 3, n = 2270, p < .0001.



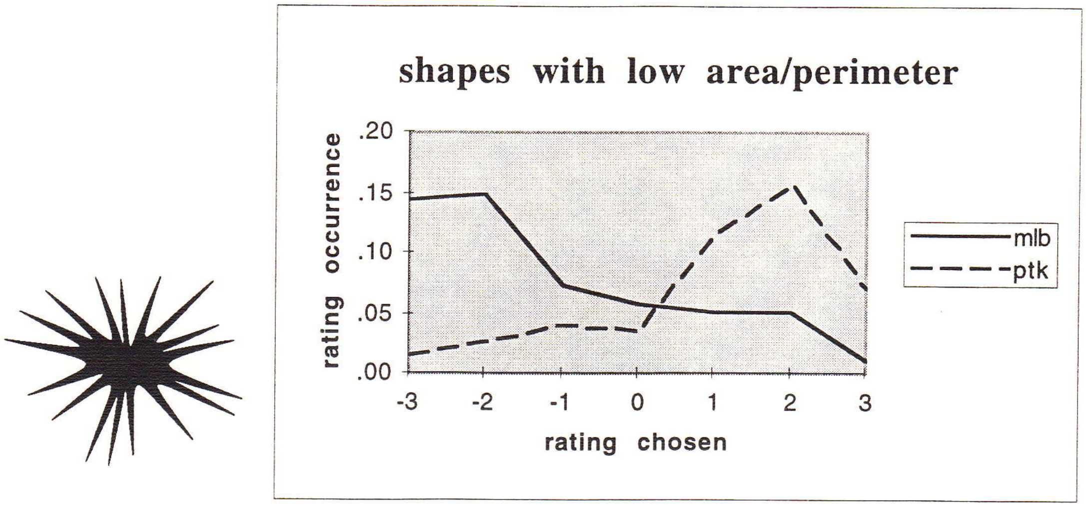
sample shape

Figure 10. This graph shows the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes of high circularity, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to high circularity comparison fit rating increases by an average of 19% for each increment in the rating scale as one moves from a -3 to a 3, n = 1833, p < .0001.



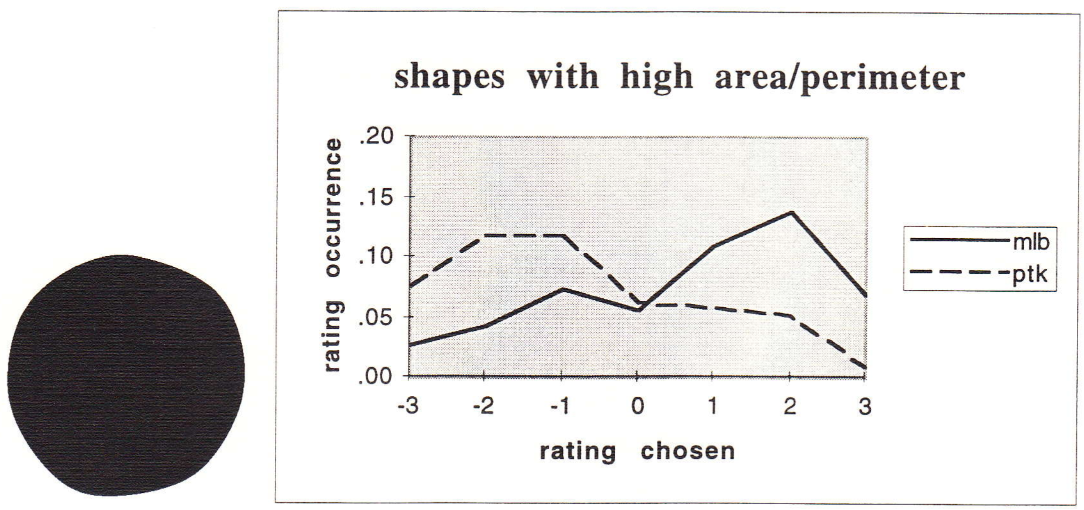
sample shape

Figure 11. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes of low circularity, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to low circularity comparison fit rating decreases by an average of 37% for each increment in the rating scale as one moves from a -3 to a 3, n = 2298, p < .0001.



sample shape

Figure 12. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes with a low area to perimeter ratio, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to low area/perimeter comparison fit rating decreases by an average of 50% for each increment in the rating scale as one moves from a -3 to a 3, $n = 866$, $p < .0001$



sample shape

Figure 13. This graph shows an interaction between the fit ratings given by subjects for comparisons of {mlb}-containing auditory stimuli with shapes with high area to perimeter ratio values, versus {ptk}-containing sounds when compared with the same shapes. The odds of receiving a particular {mlb} to high area/perimeter comparison fit rating increases by an average of 64% for each increment in the rating scale as one moves from a -3 to a 3, $n = 888$, $p < .0001$.

Orientation analyses between {mlb} and {ptk}. Additional analyses were done between the shape classification sets designed to test the significance of changes in orientation on the rating behavior of the participants. Three orientation shape sets were created. Each of the three sets was composed of shapes in either a vertical orientation, horizontal orientation, or a 45 degree orientation respectively. All factors other than those involved in the direction of the shapes' orientation were held constant between sets. Using the data from trials containing m, l, or b sounds (and likewise for the data involving comparisons which contained p, t, or k sounds), the ratings given by subjects were compared over all changes in orientation. These analyses would tell us if there were any systematic differences in the fit rating given to comparisons involving one of the subsets of sounds (i.e., {mlb} or {ptk}) over changes in a shape's orientation. All possible changes in orientation were compared. As hypothesized, orientation was not a significant factor influencing the ratings chosen for any of the particular sound-shape mappings investigated in this paper. Logistic regression, $p > .07$ for all analyses (for four of the six analyses, $p > .25$).

Consonant-vowel versus vowel-consonant auditory stimuli. A test to examine whether fit ratings were systematically effected by the temporal order of the vowel and consonant in the auditory stimuli was conducted. These analyses will answer the question: do subjects find that vowel-consonant sounds appear to fit shapes that are more or less circular, curved, or have different area to perimeter ratios relative to consonant-vowel sounds? Analyses revealed no difference in the ratings given to {mlb}-vowel versus vowel-{mlb} sounds within any level of three the shape factors: logisitic regression, p-values all > .23. Sounds of the form {mlb}-vowel were never judged to fit the shape characteristics we are examining differently than sounds of the vowel-{mlb} form.

For auditory stimuli that use one of the consonants within the {ptk} set, however, the outcome was markedly different. Using logistic regression, the data were analyzed within each shape classification set (negative curvature, neutral curvature, positive curvature, low circularity, high circularity, low area/perimeter, high area/perimeter) for systematic rating differences between trials involving comparisons with auditory stimuli of the form vowel-{ptk}, versus trials using stimuli with the {ptk}-vowel order. The subjects' ratings for each shape-sound comparison was the dependent variable. Analyses within five of the seven shape levels showed significant differences between comparison ratings to {ptk}-vowel sounds versus vowel-{ptk} sounds (all $p < .05$), and the two remaining non-significant differences nonetheless exhibit differences consistent with the directions found in the other analyses: The direction of all differences is consistent with the general statement that vowel-{ptk} sounds were judged to be "pointier" than {ptk}-vowel sounds. Shapes with negative curvature, low circularity (not significant), or low area to perimeter ratios were consistently judged to be a better fit to vowel-{ptk} auditory dyads, and conversely, shapes with any of these same characteristics were judged to be worse fits to auditory stimuli of the form {ptk}-vowel. Experimental shapes representing any of the remaining factors (neutral or positive curvature, high circularity, or a high area to perimeter ratio (not significant)) were judged to fit sounds of the form {ptk}-vowel better than stimuli of the vowel-{ptk} configuration. See Table 2 and Figures 14, 15 and 16.

As with the analyses of the differences in rating responses between {mlb} or {ptk} to shape comparisons, the rate of change in these analyses, as provided by logisitc regression, gives us information on how extreme the {ptk}-vowel versus vowel-{ptk} rating divergence is along the -3 to 3 fit rating scale. For shapes with negative curvature, for instance, we can see that the odds of a subject responding with any particular fit rating for a {ptk}-vowel to negative curvature shape comparison decreases by an average of about 13% (relative to comparisons involving vowel-{ptk} sounds) for each increment from -3 to 3 along the rating scale. Stated briefly, {ptk}-vowel sounds are more likely to be rated a worse fit to negative curvature shapes than vowel-{ptk} sounds.

Shape parameter and level	Weighted average fit rating {ptk}-vowel	vowel-{ptk}	Rate of change (logistic regression)	n	p-value
negative curvature	0.80	1.10	-13.32%	1199	<.0001
neutral curvature	-0.40	-0.80	18.17%	1084	<.0001
positive curvature	-0.44	-0.66	9.03%	1115	<.05
low circularity	0.72	0.86	---------	887	ns
high circularity	0.03	-0.32	12.02%	1096	<.01
low area/perimeter ratio	0.87	1.33	-17.65%	400	<.01
high area/perimeter ratio	-0.70	-0.90	--------	454	ns

Table 2. The results of the analyses of differences in fit ratings for auditory stimuli containing the sounds p, t, or k, in the consonant-vowel order ({ptk}-vowel) versus the vowel-consonant order (vowel-{ptk}). The rate of change figures for this table are interpreted with respect to the set {ptk}-vowel: for example, for shapes of negative curvature, ratings of fit given to comparisons with sounds {ptk}-vowel form are increasingly less likely as we move from -3 (terrible fit) to 3 (excellent fit) along the rating scale by a factor of about .13 relative to the occurrence of ratings for auditory stimuli of the form vowel-{ptk} when compared to the same shapes. The differences seen in the non-significant items suggest directions consistent with the remaining significant analyses. Weighted averages are included as a heuristic for comparison, but did not figure significantly in the analyses, which provides more elaborated measures for comparison.

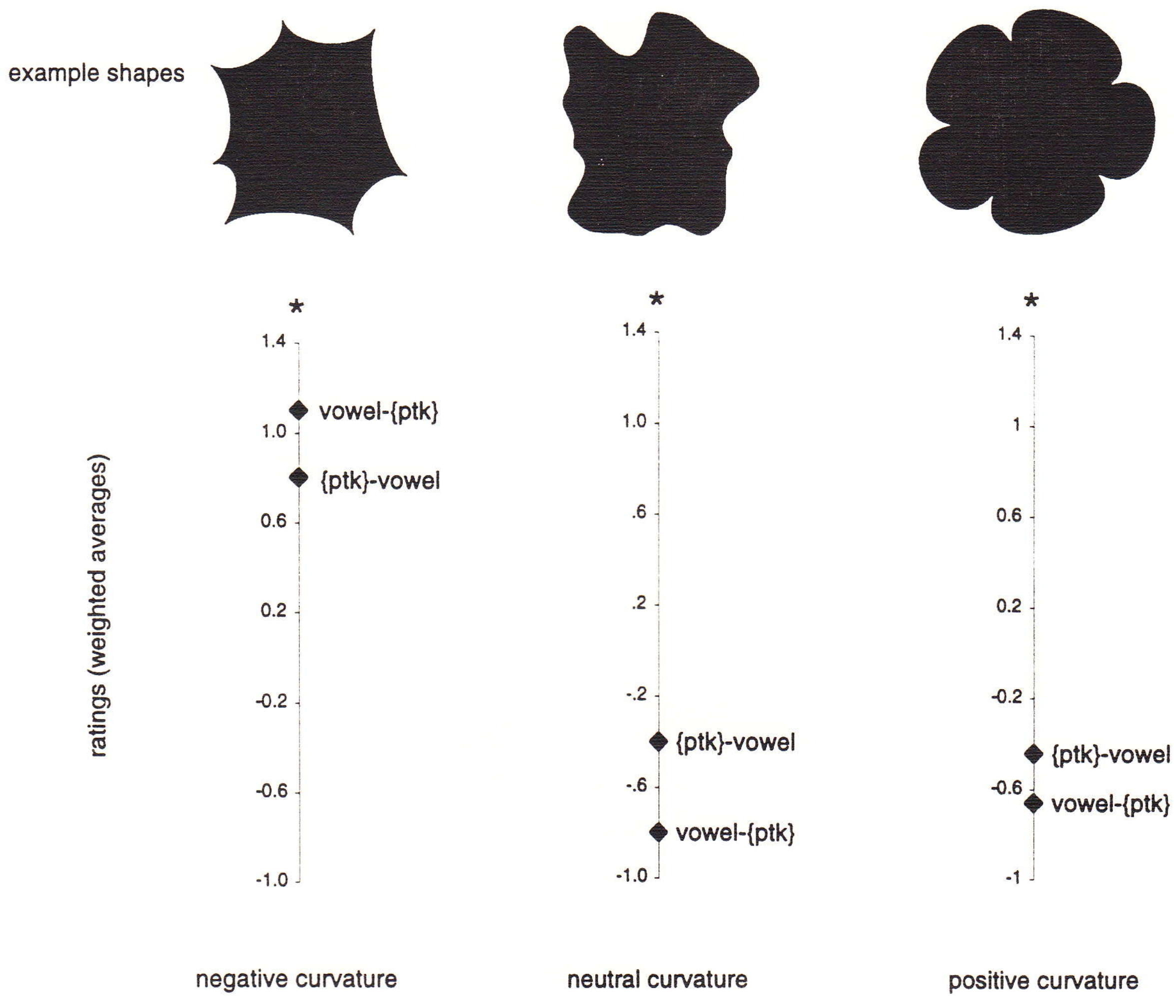


Figure 14. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with negative, neutral, and positive curvature values for auditory stimuli which include the consonants p, t, or k. A higher rating indicates that subjects found a sound of that form to be a better fit to shapes like the one indicated above the scale. Significant differences are marked with a star.

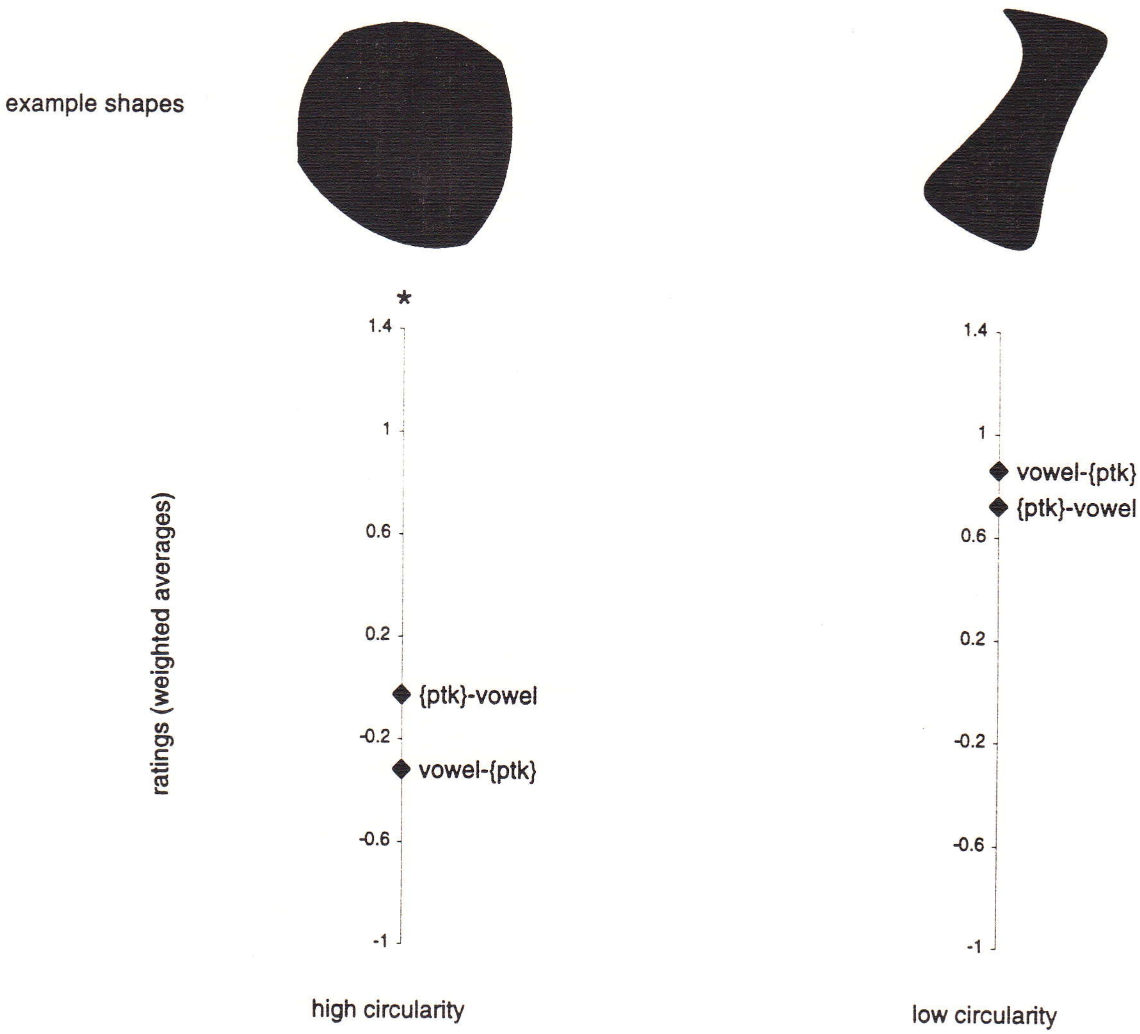


Figure 15. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with low, and high circularity values for auditory stimuli which include the consonants p, t, or k. A higher rating indicates that subjects found a sound of that form to be a better fit to shapes like the one indicated above the scale. Significant differences are marked with a star.

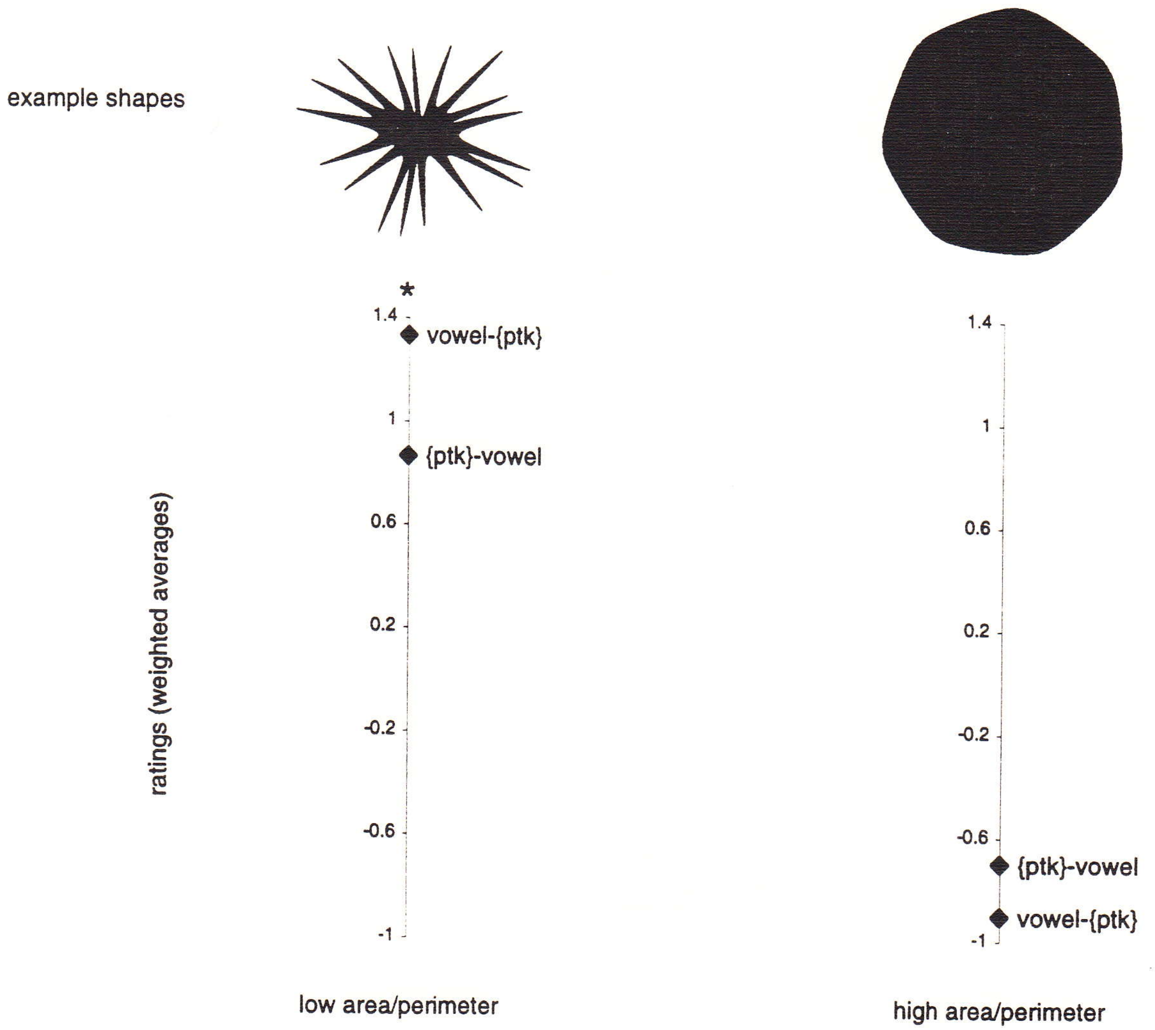


Figure 16. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with low, and high area/perimeter ratios for auditory stimuli which include the consonants p, t, or k. A higher rating indicates that subjects found a sound of that form to be a better fit to shapes like the one indicated above the scale. Significant differences are marked with a star.

Relationships within the {mlb} and {ptk} sets. Additional analyses were conducted to investigate the occurrence of consistent differences in subject response between the members of each of the {mlb} and {ptk} sets. The data were again analyzed within each shape classification set, this time for systematic rating differences in trials with auditory stimuli containing m, versus those containing l, versus those containing b, and likewise for the members of the {ptk} set. The participants' fit ratings, again, comprised the dependent variable. Within the {mlb} set, only one systematic difference arose: the b sound was rated significantly different from both m and l for all but one shape classification set -- high circularity. See Table 3.

Shape parameter and level	Weighted average fit rating {b}	{ml}	Rate of change (logistic regression)	n	p-value
negative curvature	-0.35	-0.83	-15.92%	1543	<.0001
neutral curvature	0.54	1.02	24.40%	1485	<.0001
positive curvature	0.39	0.70	15.98%	1567	<.005
low circularity	-0.20	-0.60	-15.50%	1578	<.005
high circularity	0.32	0.40	----------	1618	ns
low area/perimeter ratio	-0.49	-1.54	-43.03%	634	<.0001
high area/perimeter ratio	0.30	0.89	22.52%	615	<.05

Table 3. The results of the analyses of differences in fit ratings for auditory stimuli containing the consonant b versus sounds containing the consonants m, or l for each shape classification set. The rate of change figures for this table are interpreted with respect to the set {ml}: for example, for shapes of negative curvature, ratings of fit given to comparisons with sounds containing an m or an l are increasingly less likely as we move from -3 (terrible fit) to 3 (excellent fit) along the rating scale by a factor of about .16 relative to the occurrence of ratings for b-containing auditory stimuli when compared to the same shapes. The data from the one non-significant analysis suggest directional consistency with the outcomes of the remaining analyses. Weighted averages are included as a heuristic for comparison.

There were no such systematic differences to be found between members of {ptk} in comparison ratings over the shape classification sets. There was one significant, although not systematic difference between {k} and {tp}for shapes of negative curvature where {k} was found to be different from both {t} ($p < .05$) and {p} ($p < .0001$). See Figure 17.

This difference was not supported by analyses of the data for any of the other shape factor levels, but does appear to speak for a trend throughout the remainder of the data, as {k} consistently appears to fall at the outmost extreme in the {ptk} sets' fit ratings for all shape analyses. Regardless, according to analyses, the {ptk} set remains more cohesive than its {mlb} counterpart. See Figures 17, 18 and 19.

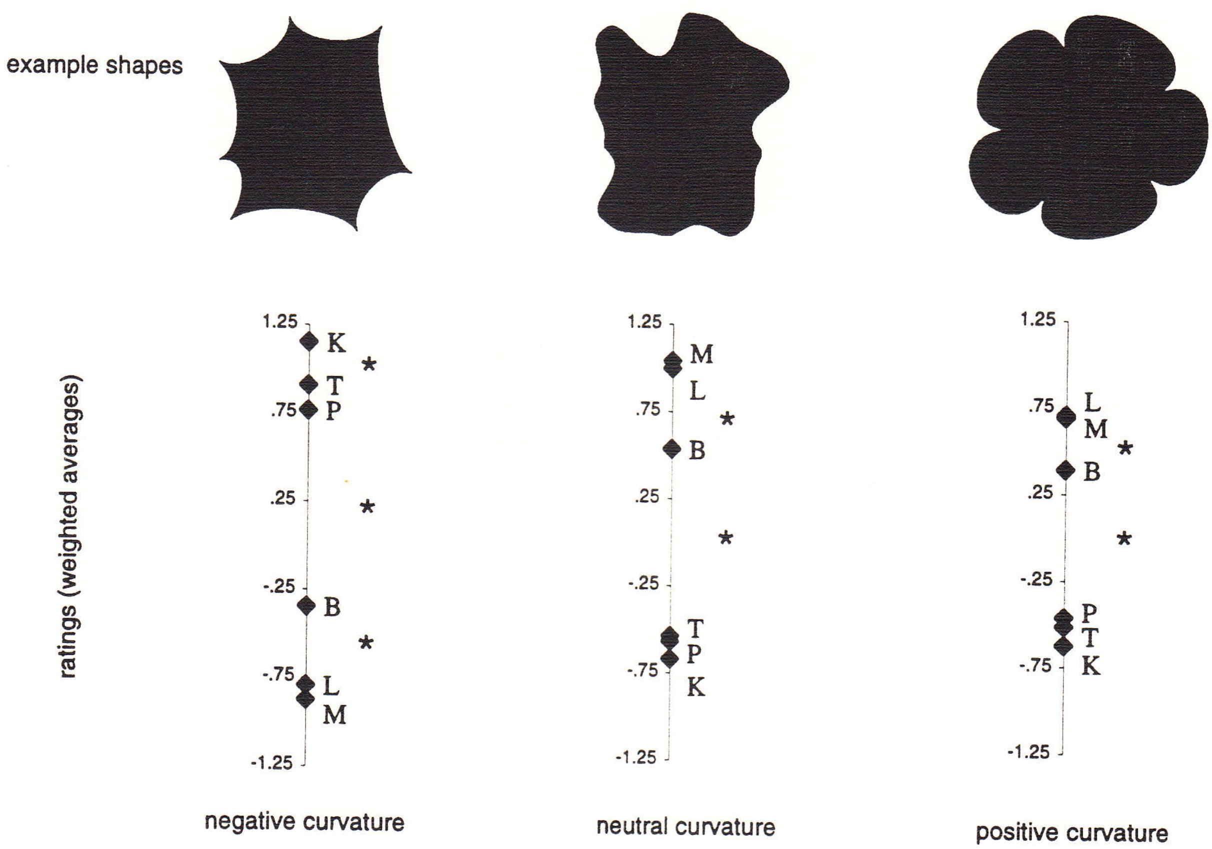


Figure 17. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with negative, neutral, and positive curvature values for all auditory stimuli. A higher rating indicates that subjects found that consonant to be a better fit to shapes like the one indicated above the scale. Significant differences, via logistic regression, are marked with a star.

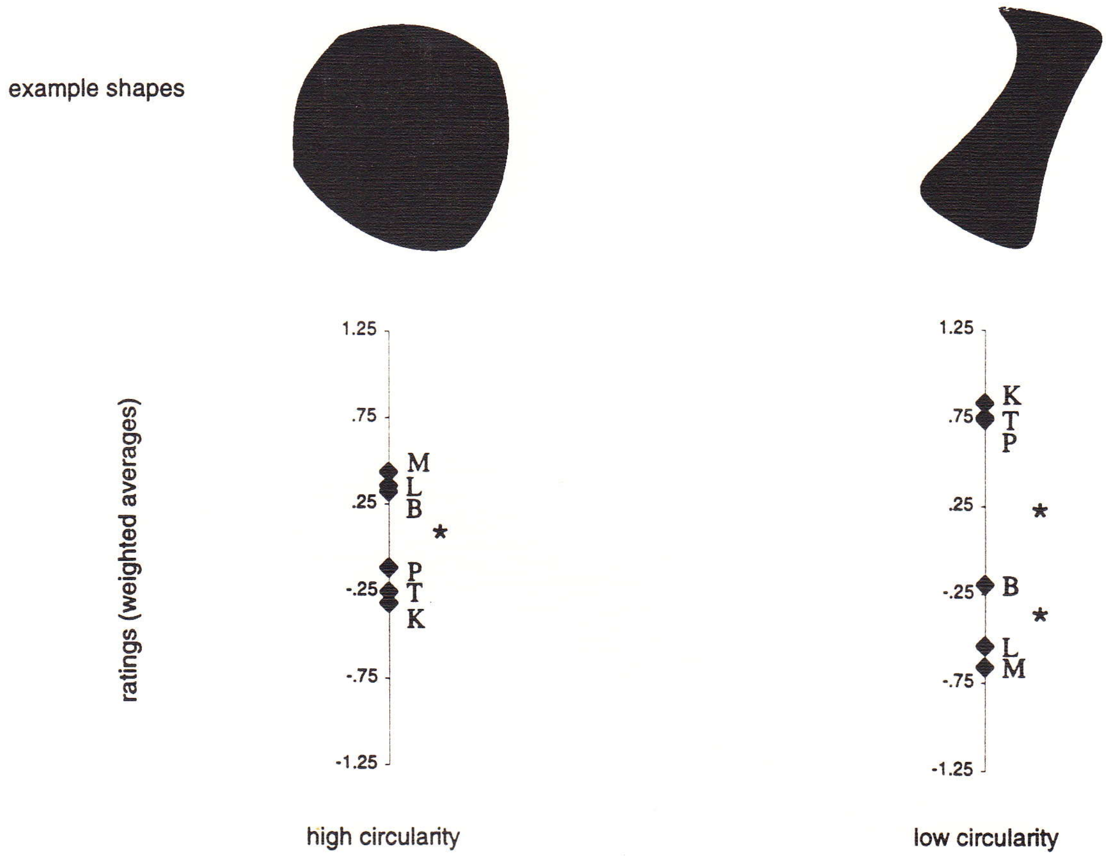


Figure 18. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with low, and high circularity for all auditory stimuli. A higher rating indicates that subjects found that consonant to be a better fit to shapes like the one indicated above the scale. Significant differences, via logistic regression, are marked with a star.

Discussion

This experiment was undertaken to test a number of hypotheses: first, the often contested claim (Firth, 1964) that phonesthetic mapping is reliable and consistent across

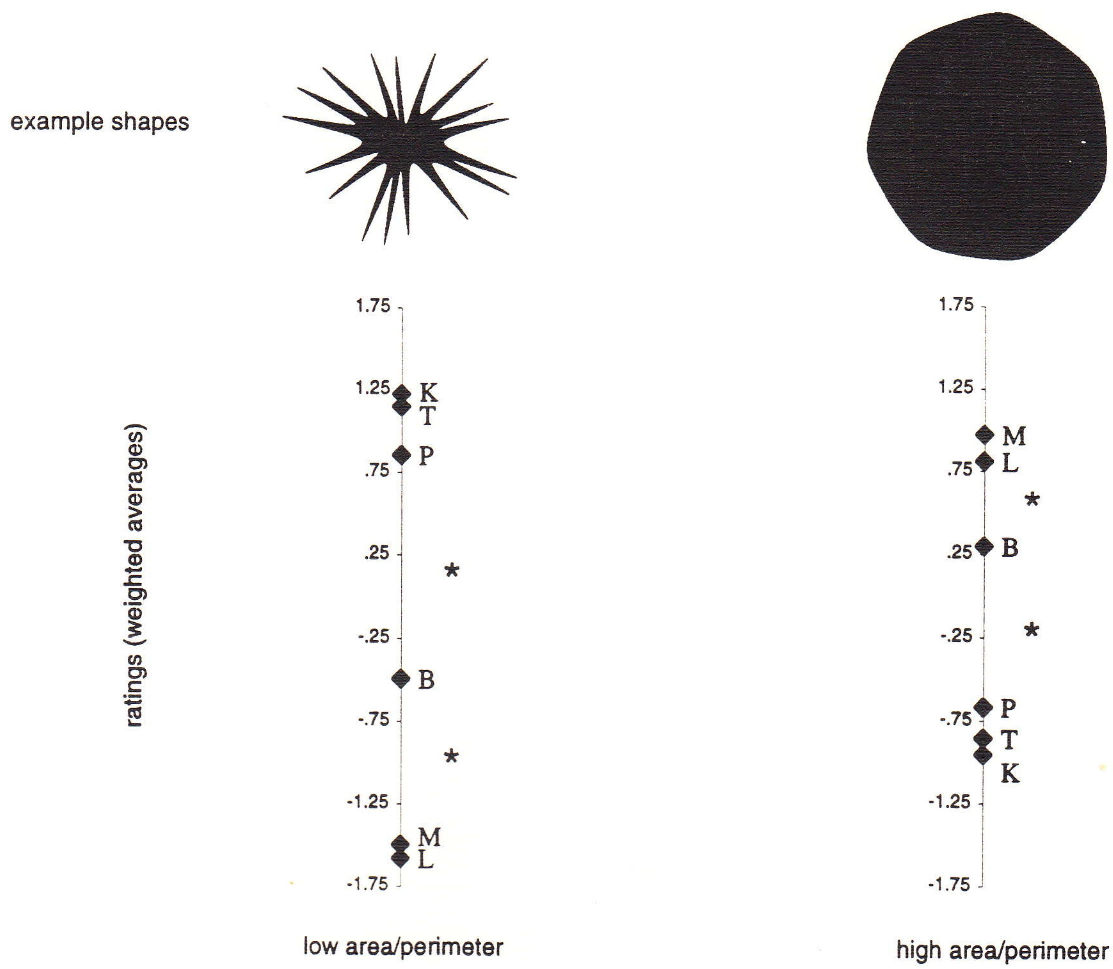


Figure 19. The weighted average of fit ratings (-3 to 3) subjects gave to shapes with low, and high area to perimeter ratios for all auditory stimuli. A higher rating indicates that subjects found that consonant to be a better fit to shapes like the one indicated above the scale. Significant differences, via logistic regression, are marked with a star.

subjects; second, that particular mappings of sound to shape exist within the general perceptual domain described as "pointiness"; and third, that such sound to shape mappings, if they exist, can be demonstrated experimentally with the tools and procedures selected.

In addition to the hypothesis testing goals stated above, this experiment was used for the evaluation of a number of tools created to get a closer and more controlled look at phonesthetic claims involving shape and sound. Among the tools originated for the purpose of investigating this phenomenon were: shape generation software, which produces parametrically definable shapes; computationally tractable algorithms for capturing perceptually salient distinctions between shapes; and an experimental platform using a voice synthesizer to present auditory stimuli. The introduction of novel tools into this domain of inquiry gives rise to a few questions on the level of implementation, especially concerning a) the usefulness of the measures of shape which the new software has made possible, and b) the effectiveness of the experimental platform in its role of facilitating the comprehensibility of the multi-modal comparison participants were asked to make. Questions concerning the usefulness of the tools mentioned above, however, are best addressed as a corollary to the success we have had in meeting our hypothesis testing goals.

The mlb vs. ptk sound sets. The long hypothesized phonesthetic distinctiveness of the sound sets {mlb} and {ptk} was confirmed by the analyses conducted on the data from comparisons involving parametrically defined shapes over three different morphological measures. The sounds of the {mlb} set were consistently judged to be better fits to shapes which were more circular, or had either neutral or positive curvature, or high area to perimeter values. All of these results support the broad claim that the sounds m, l, and b are perceived to be "rounder" or "softer" than the sounds of the {ptk} set. In complementary fashion, the sounds of the {ptk} set were judged to be a better fit to the remaining shape factor levels.

Our analyses additionally revealed that shapes of neutral curvature were consistently judged to be less pointy than shapes of positive curvature (flower-like shapes). This result appears to support the idea that subjects do in fact attend to the internal or concave angles of a shape (as found in shapes of positive curvature) when making sound-shape comparisons of this type, but such internal angles appear to be less visually salient than external or convex angles of similar size (as in shapes of negative curvature).

Consonant-vowel versus vowel-consonant orderings. The analyses examining the possible effect of the use of two different consonant and vowel orders upon participants' fit ratings revealed a surprising result: auditory stimuli of the form consonant-vowel were consistently rated less pointy than stimuli of the vowel-consonant form within the {ptk} data. It was expected that the softer (voiced, continuant) consonants m and l, and the voiced consonant b, would be perceived to be more similar to vowels than {ptk} sounds, as p, t, and k are all unvoiced, stop consonants. The transposition of consonant and vowel within the {mlb} data would result in less perceived ordering difference than the transposition of consonants and vowels, where the consonants ({ptk}) were distinctly different from the vowels. Therefore, the fact that differences in consonant and vowel ordering produced no effect within the {mlb} data was not a surprise.

The surprising aspect of the significant {ptk} consonant and vowel ordering effect was that vowel-{ptk} pairs were judged to be pointier than {ptk}-vowel auditory stimuli. Despite our intuitions to the contrary, that any difference would prove {ptk}-vowel pairs to be the pointier of the two orderings (due primarily to the consonant's temporal primacy), the opposite rating effect was observed throughout all shape factor levels, and was significant in five out of the seven shape levels. See Table 2. This unexpected result may be explained by the abrupt ending a stop vowel in final position mandates for a sound. Thus, whether a sound has a definite endpoint (hard) or a more arbitrary one (soft) as in sounds with a vowel or non-stop consonant such as m or l in final position, appears to have a strong effect on the perceived pointiness of the sound. Further, one may argue, that this final position stop is more salient than a stop in initial position, which creates an abrupt beginning for a sound. Thus, the overall softness ratings ascribed to both {mlb}-vowel and vowel-{mlb} data, and to a lesser extent to auditory stimuli of the {ptk}-vowel form may be primarily a function of the "- stop" character of the final sound of the stimuli.

The above argument is reasonable, but does not take into account the data discussed in the next section describing a systematic rating difference within the {mlb} sound set. If we take our cue from this next set of analyses, it does not appear that it is the +/-stop feature of the consonant that is of primary saliency in phonesthetic sound discrimination for pointiness. The following analyses suggest, that while the stop feature is one important factor differentiating the experimental sounds from one another, the more predictive linguistic feature for our rating data is the consonant's voicing. The general form of the stop feature explanation for rating differences in consonant and vowel ordering may still hold, but with the augmented interpretation that the causal factor behind the rating differences in {ptk}-vowel versus vowel-{ptk} data is not merely the +/-stop status of the final sound, but is primarily the +/-voicing of the final sound -- two factors which happen to be conflated for all sounds in the {ptk} set. Up to this point, however, the voicing feature alone, does not allow for the observed difference between {ptk}-vowel and {mlb} stimuli.

This emphasis on the voicing feature may support an additional operation as well: that pointiness may be a function of the internal "smoothness" of the sound, and changes in voicing literally disrupt the vibrational status of the vocal cords, resulting in pointier ratings for sounds which contain both voiced and unvoiced components (i.e. {ptk}). However, this internal consistency hypothesis alone does not account for the ordering effect we have encountered in {ptk}, hence we need both the factor of the +/-voice of the final sound and the internal consistency mechanism in order to explain the consonant and vowel order effect in a voice only account.

A focus on the voicing feature has an additional benefit, in that it allows for easy explanation of the large significant differences between the {mlb} and {ptk} data of the first analyses, even though {mlb} contains a stop. The stop feature, however, cannot be disregarded entirely, and it is likely that all three mechanisms described above have some effect in the existence of the {ptk} ordering effect, although their precise relative importance for pointiness remains to be determined. This observation brings us to the next set of analyses.

Rating differences within the {mlb} and {ptk} sets. These analyses were undertaken to test the cohesiveness of each of the {mlb} and {ptk} data sets, and to get a closer look at the subtleties of fit rating choice at each of the auditory phonesthetic extremes for individual consonants. The consonants m, l, b, p, t, and k were chosen from the literature as widely held representatives of some of the phonesthetically roundest and pointiest. (e.g., Wescott, 1971). To justify the use of these sounds in analyses as two (assumed) monotonic sets ({mlb} and {ptk}) requires some investigation into the actual cohesiveness of each of those sets.

For the "pointy" auditory stimuli, there were no significant, systematic differences in the choice of fit ratings between any pair of members in the {ptk} data set for any shape classification level analysis. For the "round" auditory stimuli, however, one significant, consistent difference between the sounds arose: the b sound was systematically judged to be pointier than both the m and l sounds (which were statistically indistinguishable). This {b} to {ml} difference may be easily accounted for by the +stop status of b as opposed to the -stop nature of the other two members of the {mlb} set. This intra-set difference, however, was not large enough to be a statistical detriment to the cohesiveness of the {mlb} set as a whole when compared to the rating data of {ptk}. In short, b was judged to be significantly more different than the other experimental +stop consonants that comprised {ptk} than the members of its own auditory set. The weighted average fit ratings for the consonant b suggest this result, in that {b}'s weighted averages over all shape factor levels were valenced identically to those of {m} and {l}, and opposite those of all members of {ptk}. This result supports the conclusion that voicing is a more important linguistic feature than +/-stop for predicting fit rating choices in comparisons involving the notions round and pointy. The +/-stop feature, however, is important enough to produce a reliable difference within the voiced {mlb} set.

Although not statistically significant, there are two noteworthy trends visible in these data which should be mentioned as possibilities for further experimental elaboration. The weighted average fit ratings for {ptk} suggest a reliable pointiness ordering between the members of this set (from pointiest to roundest): {k}, {t}, {p}. Likewise, in the {mlb} data a similar pointiness trend stands out (again, pointiest to roundest): {b}, {l}, {m}. The linguistic feature, point of articulation, may be hypothesized to play a role here, as there appears to be a trend from labial-round to velar-pointy. The one divergence from this pattern, {b}, is nonetheless consistent as an extension of {ptk}, and its +stop status appears to account for its placement within {mlb}.

Shape factors. The morphological factors presented for experimentation involved measures of circularity, curvature, the ratio of a shape's area to its perimeter, and the shape's orientation. The first three of these factors were originally hypothesized to be good indicators of the visual information attended by subjects in making sound to shape comparisons, where the consonants involved in the comparison were m, l, b, p, t, or k. It was hypothesized that the fourth factor, orientation, would not represent aspects of shape attended by participants in comparisons with these same sounds. Thus, curvature, circularity, and area to perimeter ratio should be good predictors of fit rating, and orientation should be non-predictive. Both hypotheses were confirmed by the data, supporting the assertion that the measures of shape put forward here are effective indicators of visual detail relevant to sound-shape comparisons involving the experimental sounds m, l, b, p, t, and k, and may further provide insight into various perceptual continua that make up the term "pointiness."

There were differences in the predictive strengths of the shape factor levels shown to be effective predictors of fit rating response in the analyses of the {mlb} versus {ptk} data. The strongest morphological predictor of fit rating between shapes and auditory stimuli was neutral curvature, followed by high area to perimeter ratio, positive curvature, low area to perimeter ratio, negative curvature, low circularity, and high circularity. (Although the predictive strengths differed between shape factor levels, all levels presented a reliable effect, $p < .0001$. Refer back to Table 1.)

Orientation. Although the inclusion of orientation in our studies was intended for control purposes as a factor hypothesized to have no effect on participants' fit rating choices, orientation may not necessarily remain unpredictive in all studies of phonetic symbolism. The direction a shape points may contain "speed," "stability," or even "weight" connotations, for example, shapes pointing horizontally may be perceived to be indicating motions which are slower than shapes with vertical orientations, and top-heavy shapes may elicit very different rating responses from subjects than bottom-heavy stimuli. In such instances, orientation may become an important factor for studies comparing "quick" sounds such as z, or v, to "slower" sounds such as d or g, or likewise for studies looking at concepts such as "activity," or "potency." (c.f., Osgood (1960) and Osgood, Suci, and Tannenbaum (1957). In our study of "pointiness," however, orientation has proven to be an unpredictive factor for describing the mapping of shapes to sounds such as m, l, b, p, t, and k.

Final comments. The identification of the morphological and linguistic factors important to these sound to shape mappings, and knowledge about these factors' relative predictive effectiveness will allow for the creation of a parsimonious model for phonesthetically consistent (intuitively correct) fit between sounds and a shapes. Such a model may be put to use in creative interactive computational environments involving visual or auditory input and feedback -- a possibility that provided the primary motivation for this study, and required a precision and consistency in descriptive methods that were unavailable from previous accounts of the phonesthetic sound to shape association. This experiment represents a first step toward the goal of a full model for this type of sound to shape association space, and has verified our ability to formulate morphological measures of perceptual importance, and the usefulness of the tools originated here for its investigation. As the details of the shape component of the comparison space can now be more rigorously defined, the domain consequently becomes capable of sustaining the long required cartographic detail it has always lacked. A "sound" model is finally within sight.

The task. A final comment is in order on the unusual nature of the comparison subjects were asked to make. The judgment of the existence and type of fit between shapes and sounds is not one that we are aware of making in our daily lives. Perhaps it only occurs to us that we have expectations concerning the co-occurrence of morphological and auditory features when those expectations are (rarely) violated, for instance, an encounter with a large, menacing dog with a high, squeaky voice, or a paperclip hitting the floor with an accompanying thud. There is a danger, in conducting experiments like this one, that upon receiving the scant instructions for the comparison they are to make, subjects will simply not know how to proceed, or protest vociferously that the task makes no sense. The modal reaction from subjects before beginning the task was a shrug of the shoulders and a confused or exasperated expression. After the experiment, however, they appeared to have fully apprehended the task, insisting that certain shapes and sounds were obviously made for each other, and arguing among themselves as to the details of these mis/matches. The data appear to support the conclusion that (in addition to the significant between-subject consistency revealed in the data), subjects found some meaning in the comparisons they were making. The option to select a 0 (no fit) rating was available for every trial, and yet appears to have been chosen by our participants less than might be expected. See Figures 7 - 13. Phonesthetic information appears to be widely available and is a potentially important cue to the external world, as such it deserves more scrupulous study. The institution of more precise levels visual and auditory analysis is a first step in the right direction.

Bibliography

Allott, Robin. (1997). Sound Symbolism. http://www.percep.demon.co.uk/soundsmb.htm#ref

Bently, M. and Varon, E. J. (1933). An accessory study of "Phonetic Symbolism." American Journal of Psychology, 45, 76-86.

Brown, R. W., Black, A. H., and Horowitz, A. E. (1955). Phonetic symbolism in natural languages. Journal of Abnormal and Social Psychology, 50, 388-393.

Brown, R. W., and Nuttal, R. (1959). Method in phonetic symbolism experiments. Journal of Abnormal and Social Psychology, 59, 441-445.

Carmichael, L., H. P. Logan, et al. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. Journal of Experimental Psychology, 15, 73-86.

Cohen, R. A. and Izawa, C. (1976). Effects of phonetic symbolism on paired-associate learning. Bulletin of the Psychonomic Society 8(6), 475-478.

Daniel, T. C., and Togli, M. P. (1976). Recognition gradients for random shapes following distinctive or equivalent verbal association training. Journal of Experimental Psychology: Human Learning and Memory 2(4), 467-474.

Denes, G. and Spinaci, M. P. (1981). Influence of association value in recognition of random shapes under dichhaptic presentation. Cortex, 17(4), 597-601.

Firth, J. R. (1964) Tongues of men and speech. London: Oxford University Press.

Greenberg, J. H. (1965). Some generalizations concerning initial and final consonant sequences. Linguistics, 18, 5-34.

Hevner, K. (1935). Experimental Studies of the Affective Value of Colors and Lines. Journal of Applied Psychology, 19, 385-398.

Hinton, L, Hichols, J. and Ohala, J. J. (1994). Sound symbolism. Cambridge, England: Cambridge University Press.

Irwin, W. F. and E. Newland. (1940). A genetic study of the naming of visual figures. Journal of Psychology, 9, 3-16.

Jakobson, R. and Waugh, L. R. (1987). The sound shape of language. New York: Mouton de Gruyter.

Johnson, R. C., Suzuki, N. S., and Ohls, W. K. (1964). Phonetic symbolism in an artificial language. Journal of Abnormal and Social Psychology, 69(2), 233-236.

Köhler, W. (1929/1947). Gestalt Psychology. New York: Liveright.

Lindauer, M. S. (1990). The meanings of the physiognomic stimuli *takete* and *maluma*. Bulletin of the Psychonomic Society, 28(1), 47-50.

Lindauer, M. S. (1990). The effects of the physiognomic stimuli *takete* and *maluma* on the meanings of neutral stimuli. Bulletin of the Psychonomic Society, 28(2), 151-154.

Lundholm, H. (1921). The Feeling Value of Lines. Psychological Review, 28, 43-60.

Malkiel, Y. (1990). Diachronic problems in phonosymbolism. Amsterdam: J. Benjamins Publishing Company.

Marchand, H. (1957). Phonetic symbolism in English word-formation. Indogermanische Forschungen, 64, 146-168, 256-277.

Markell, N. N., and Hamp, E. P. (1960/1961). Connotative meanings of certain phoneme sequences. Studies in Linguistics, 15(1/4), 47-61.

Marks, L. E. (1987). On cross-modal similarity: Audio-visual interactions in speeded discrimination. Journal of Experimental Psychology: Human Perception and Performance, 13(3), 384-394.

Masson, D. I. (1952). Synesthesia and sound spectra. Word, 8(1/4), 39-41.

Nagae, S. (1978). Effects of phonetic symbolism labeling on shape recognition memory. Japanese Journal of Psychology, 48(6), 329-336.

Newman, S. S. (1933). Further Experiments in Phonetic Symbolism. American Journal of Psychology, 45, 53-75.

Orr, J. (1945). One some sound values in English. British Journal of Psychology, 35, 1-18.

Osgood, C. E. (1960). The cross-cultural generality of visual-verbal synesthetic tendencies. Behavioral Science, 5, 146-169.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). The measurement of meaning. Urbana: The University of Illinois Press.

Oyama, T. and Haga, J. (1963). Common factors between figural and phonetic symbolism. Psychologia: An International Journal of Psychology in the Orient, 6, 131-144.

Pullum G. K. and Ladusaw, W. A. (1986). Phonetic symbol guide. Chicago: University of Chicago Press.

Sapir, E. (1929). A Study in Phonetic Symbolism. Journal of Experimental Psychology, 12, 225-239.

Scheerer, M. and Lyons, J. (1957). Line drawings and matching responses to words. Journal of Personality, 25, 251-273.

Smets, G. J. F. and Overbeeke, C. J. (1989). Scent and sound of vision: Expressing scent or sound as visual forms. Perceptual and Motor Skills, 69, 227-233.

Stratton, G. M. (1917). Theophrastus and Greek physiological psychology before Aristotle. London: Allen & Unwin.

Svartdal, F. and Iverson, T. (1989). Consistency in synesthetic experience to vowels and consonants: Five case studies. Scandinavian Journal of Psychology, 30(3), 220-227.

Tarte, R. D. and O'Boyle, M. W. (1982). Semantic judgments of compressed monosyllables: Evidence for phonetic symbolism. Journal of Psycholinguistic Research, 11(3), 183-196.

Taylor, I. K., and Taylor, M. M. (1965). Another look at phonetic symbolism. Psychological Bulletin, 64, 413-427.

Thorndike, E. L. (1945). One the association of certain sounds with pleasant and unpleasant things. Psychological Review, 52, 143-149.

Uznadze, D. (1924). Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung. Psychologische Forschung, 5, 24.43.

Weiss, J. H. (1964). Phonetic symbolism reexamined. Psychological Bulletin, 61, 454-458.

Weiss, J. H. (1964). The role of stimulus meaningfulness in the phonetic symbolism response. Journal of General Psychology, 70, 255-263.

Wertheimer, M. (1958). The relation between the sound of a word and its meaning. American Journal of Psychology, 71, 412-415.

Wescott, R. W. (1971). Linguistic iconism. Language, 47, 416-428.

Whorf, B. L. (1956). Language, Thought, and Reality. Cambridge, Massachusetts: MIT Press.

Zigler, M. J. (1930). Tone Shapes: A Novel Type of Synaesthesia Journal of General Psychology, 3, 227-287.

Zusne, L. (1975). Curved contours and the associative process. Perceptual and Motor Skills, 40(1), 203-208.