2016

Iconic Properties are Lost when Translating Visual Graphics to Text for Accessibility

Coppin, Peter and Li, Abrose and Carnevale, Michael

Iconic Properties are Lost when Translating Visual Graphics to Text for Accessibility

Peter Coppin, Ambrose Li, and Michael Carnevale

For many blind and low-vision individuals, accessing charts and graphs often means accessing a text description of the graphics, usually aurally. However, in doing so, parts of the charts that are not originally conveyed textually are lost in the translation into text. By synthesizing ideas from the science and philosophy of perception and cognition, diagrammatic reasoning, and semiotics, this essay makes the case that translating charts into text descriptions results in the loss of iconic properties of the graphics, and proposes that non-linguistic sonification can be recruited to preserve such properties. The essay concludes by proposing how predictions based on this synthesis can inform design.

Keywords: Graphic representation, diagrammatic representation, sentential representation, pictorial representation, symbolic representation, iconic representation, graphic-linguistic distinction, iconic-symbolic distinction, semiotics, cognitive semiotics, cognitive science, cognitive neuroscience

1. Introduction

Imagine a blind or low-vision individual who needs to access a graphic, for example a financial chart (Figure 1a). Unlike a sighted individual, who can see the actual chart, what the blind or low-vision individual accesses, usually aurally, is often its text description. Both individuals are accessing a depiction of rising and falling stock prices over time. However, whereas the sighted individual sees words and undulating shapes, all the screen reader user hears are words (Figure 1d).

Such is the state of the art in accessible graphics: Many blind and low-vision individuals depend on approaches like the Web Content Accessibility Guidelines (WCAG) to provide them with *text descriptions* – translations of the graphics into text which these individuals can then access via a screen reader, usually aurally via text-to-speech. Text descriptions are essentially interpretations meant to convey the meaning intended by the author of the graphic. However, if a text description can fully convey the meaning of the graphic, then why did the author create the graphic in the first place?

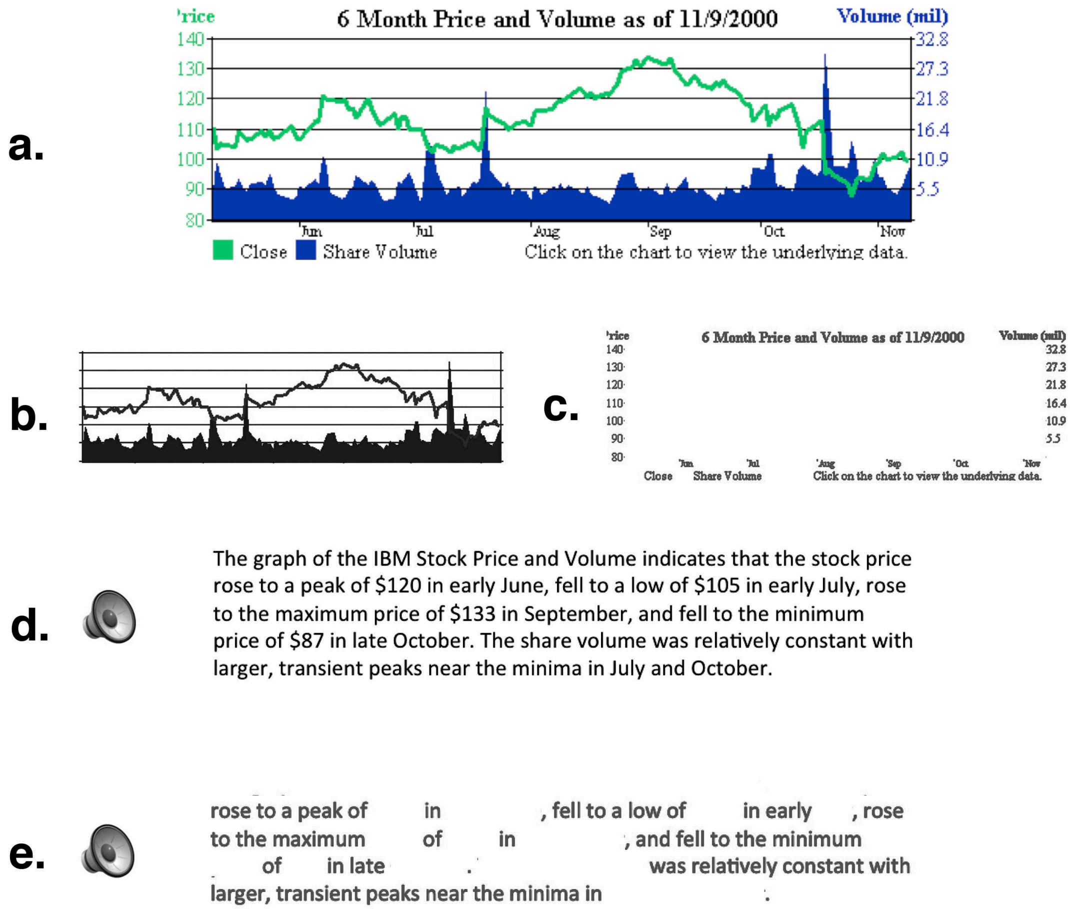

Figure 1. Deconstruction of a financial chart. The original chart (a) comprises parts conveyed via shapes (b) and parts conveyed via text (c), but in a text description (d) parts originally conveyed via shapes are also conveyed via text (e). Adapted from "Web Accessibility Best Practices: Graphs" by Campus Information Technologies and Educational Services (CITES) and Disability Resources and Educational Services (DRES), University of Illinois at Urbana/Champaign. Copyright 2005 by University of Illinois at Urbana/Champaign.

1.1. Formulating a Design Problem From the Scenario

In the foregoing scenario, parts of the chart, predominantly numerical values and labels (Figure 1c), are already text and will naturally carry over to the text description. However, I claim that the rest of the chart (Figure 1b) is not and cannot be adequately[1] translated into text because these parts that are originally conveyed via shapes are fundamentally different from the parts originally conveyed via text. Shimojima (1999) calls this fundamental difference the "graphic-linguistic distinction," while a semiotician might see this as Lessing's dichotomy of "pictures" versus "literature" (cf. Wellbery's understanding of Lessing, as elucidated by Sonesson, 1988); in either case, one might predict the parts originally conveyed via shapes to be "lost in translation."

Suppose a designer wants to improve the state of the art and design a system of accessible graphics that minimizes what is "lost in translation." Let us call this our *target design problem*. Because the state of the art is often aural, we claim that a sound-based solution is appropriate given our target design problem.

Although touch (cf. Kennedy, 1993) might be a more natural fit than sound, if the state of the art is primarily aural, a sound-based solution is probably more immediately applicable and probably aligns better with our target design problem. Also, although screen readers can deliver text descriptions tactilely as braille, refreshable braille displays are expensive relative to audio hardware. As well, although touch screens might be tactile, Klatsky, Giudice, Benntt, and Loomis (2014) have shown that conveying information tactilely via touch screens suffers from challenges related to haptic perception. Touch screens are also not as ubiquitous and inexpensive as audio hardware.

1.2. Recasting the Design Problem as a Theoretical Question

Now, to solve our target design problem we will first need to conceptualize what is actually lost, but identifying what is lost will identify distinct and common properties of graphics relative to text. Theoretically, then, our true goal is not our target design problem, but the conceptualization of what is lost during the translation from visual graphics to text descriptions – in other words, a theoretical model. Practically, however, this model can then inform the design of new approaches

[1] We are adapting Toury's (as cited in Palumbo, 2009, p. 8) definition of adequacy here, so by adequately we mean being able to subscribe to the norms of the source culture (of visual graphics) and to express the relations expressed in the source culture. This definition is consistent with the understanding that text descriptions should convey the author's intent.

for conveying properties of graphically represented shapes in non-visual perception modes, which in our target design problem would be sound.

1.3. How Such a Theoretical Model Could be Tested

While we will not discuss specific plans to test the model, since the translation of graphics to text is a translation process, the concept of "back-translation" (cf. Palumbo, 2009, p. 14) can be adapted to serve as the basis of experiments to test and challenge the model itself.

1.4. How This Essay is Organized

Our conceptualization of what is lost will be based on the science of perception and cognition, but expressed in terms compatible with semiotics. Although our destination is cognitive semiotics, we will begin in an area in computer science where research and practice relevant to our target design problem is transpiring, but, in our view, lacks a means for describing what is lost.

1.4.1 *The Graphic-Linguistic Distinction*

We will begin by reviewing several leading accounts of the "graphic-linguistic distinction" from the field of diagrammatic reasoning. What will emerge is that charts and graphs designed for visual perceivers are mostly composed of items labelled via text, but located in relation to other labels through visually perceived spatial, geometric, and topological relationships. This section will discuss how spatial properties of sound could be recruited to convey spatial, geometric, or topological relations among labelled items that are currently lost in translation.

1.4.2 *From semiotics: iconic and symbolic properties of visually or aurally perceived items*

Synthesizing the various accounts of the graphic-linguistic distinction results in a pattern that suggests a more effective means to translate what is lost, but the terms for conveying spatial-topological-geometric properties are unwieldy compared to corresponding terms in semiotics. Recruiting well established terms and concepts from semiotics will partially replace these unwieldy phrases.

At this point, diagrams can then be described as iconically conveyed relations among symbolically conveyed items or objects. However, what principles could guide a designer in distinguishing between symbolic and iconic properties of a graphic? Also, how could a designer identify appropriate mappings from iconic properties of visual graphics to those of sound to convey the same relations?

1.4.3 *Perceptual-Cognitive Properties of Pictures, Diagrams, and Text (Coppin, 2014)*

To answer these two questions, we will introduce Coppin's (2014) perceptual-cognitive model, which distinguishes between what he calls the pictorial and symbolic properties of graphics, and where he conceptualizes diagrams as pictorial relations among symbolic objects. Pictorial and symbolic properties in this model will be shown to be respectively akin to iconic and symbolic properties from semiotics.

1.4.4 *Application*

With this synthesis in place, the final section applies the synthesized model to the target design problem to precisely identify what is lost in translation and offers a solution to inform translation.

2. The Graphic-Linguistic Distinction: Implications for Sonic Interface Design

We will now examine the "graphic-linguistic distinction." Out of seven candidate distinctions that Shimojima (1999) identifies we will discuss four, in relation to how they extend the idea of this distinction into sound. Because Shimojima's discussion assumes that "generally, pictures, images, and diagrams are graphical representations" (p. 313), any characterization of graphical representations here applies also to diagrams.

2.1. 2D Versus Sequential

The first distinction to review comes from Larkin and Simon (1987), who define a diagrammatic representation as a "data structure in which information is indexed by two-dimensional location" and a sentential representation as "a data structure in which elements appear in a single sequence" (p. 68). By definition, a text description (Figure 1d) in its visual form is therefore sentential, because text is arranged as a linear sequence of marks, whereas the original chart (Figure 1a) is diagrammatic because financial values are indicated via marks indexed to a 2D grid. Larkin and Simon also state that diagrammatic representations "preserve explicitly the information about the topographical and geometric relations among the components of the problem" (Figure 2, upper right) whereas sentential representations do not (p. 66). Thus, the spatial relations among the labelled marks enable one to visually perceive the contour of the line or the relative positions of marks scattered across the 2D surface, thereby inferring values and trends that are not explicitly conveyed via labels (cf. Barwise and Etchemendy, 1995).

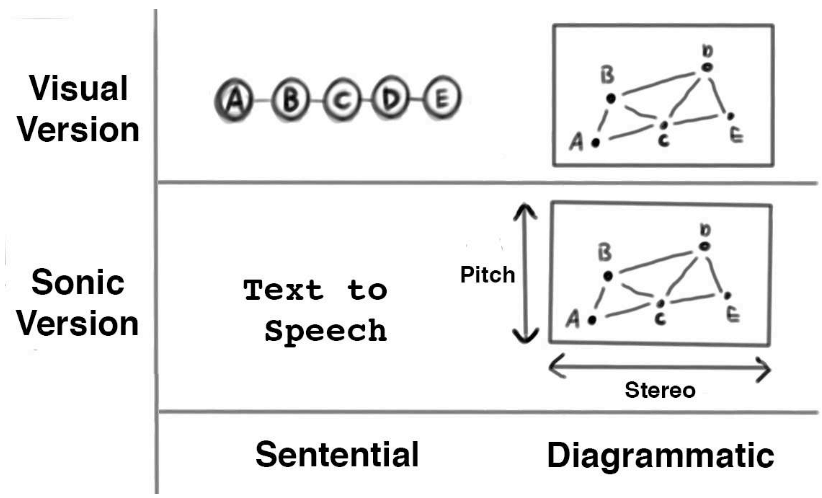

Figure 2. Diagrammatic versus sentential representations via sonic and visual perceptual modes.

2.1.1 *Implications for sonic charts and graphs*

Because Larkin and Simon's definitions do not require the information in the data structures to be visual, text descriptions in their text-to-speech form are already by definition sentential. Now note that forming a 2D space in the sonic domain will require two properties of sound that can be independently manipulated and perceived. Identifying such sonic properties should enable designers to construct sonic external representations that are diagrammatic (as defined by Larkin and Simon), which when used to translate visually perceived diagrammatic structures should enable conveying topographical and geometric relations that cannot be conveyed via text-to-speech translations. We can imagine, for example, how in Figure 2 (lower right) a blind or low-vision user could press arrow keys to move an "audio cursor" along the *x*-axis of an imagined 2D space to perceive the contours of the graph.

2.2. Relation Symbols and Object Symbols

The second distinction to examine comes from Russell (1923), who proposes that in sentences "words which mean relations are not themselves relations," whereas in maps, for instance, "a relation is represented by a relation" (p. 90). For example, a financial chart (Figure 1a) conveys higher and lower monetary values via marks at higher and lower elevations, respectively. This convention enables the visually perceived spatial relationships between the marks to represent relationships among monetary values over time.

2.2.1 *Implications for sonic charts and graphs*

To consider how relations can be conveyed sonically, consider two tones A and B in the foregoing example, where A has a lower pitch than B (Figure 3). Their sonic relation is then their perceptible difference in pitch. Now if Tone A denotes a stock price at an earlier point in time, and Tone B a stock price at a later point in time, then the perceptible difference between their pitches can convey the difference in price over time. Moving the sonic cursor from left to right would then result in a higher pitch, conveying the relationship between the stock price over time via a sonic relation.

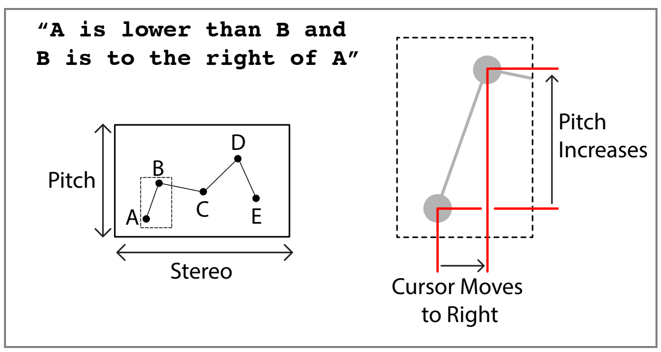

Figure 3. By scrubbing a "sonic cursor" along an axis, audiences could access sonically conveyed relations through changes in pitch and via stereo.

2.3. Analogue Versus Digital

We next explore the analogue-versus-digital distinction, most commonly associated with Goodman (1968). Analogue systems are dense throughout, where a system being dense refers to the ability to place a new element between any two elements in a representation ad infinitum, and any new element added can change the meaning of the representation. Digital systems, in contrast, are discontinuous and differentiated throughout. Goodman claims that pictorial representations are also defined by their degree of repleteness, which refers to the number of possible perceptual features that can vary before changing the representation's meaning. Thus pictures are analogue and more replete, diagrams are also analogue but less replete, but linguistic systems are partially digital.

In other words, an analogue representation can have an infinite number of variations within a representational space that can carry unique meanings. If a representation is digital, however, the number of meanings are limited.

2.3.1 *Implications for sonic charts and graphs*

Using the analogue instead of the digital properties of visually perceived graphics appears to require two interrelated capabilities: lower-level capabilities to perceptually process features from an environment or representation, and higher-level capabilities to recognize the linguistic categories of the perceptually processed features (Mandler, 2006). For example, discerning the values on a financial chart requires perceptually processing the light reflected from the chart to observe lines in relation to textual labels. Discerning the same values aurally would require the same set of interrelated capabilities: lower-level capabilities to process varying frequencies, timbre, and so on, as well as higher-level capabilities to recognize the linguistic meanings of the sounds. The current text-to-speech approach exploits only the digital properties of language; designers could produce more effective translations sonically by exploiting the analogue properties of sound.

2.4. Intrinsic Versus Extrinsic Constraints

Last to examine is the "intrinsic versus extrinsic constraints" distinction (Shimojima, 1999, p. 328) where "representations obeying inherent constraints" are considered graphical (p. 332). Let us, however, examine what he quoted from Barwise and Etchemendy (1990):

> Diagrams are physical situations. They must be, since we can see them. As such, they obey their own set of constraints . . . By choosing a representational scheme appropriately, so that the constraints on the diagrams have a good match with the constraints on the described situation, the diagram can generate a lot of information that the user never need infer. Rather, the user can simply read off facts from the diagram as needed. This situation is in stark contrast to sentential inference, where even the most trivial consequence needs to be inferred explicitly. (p. 22)

To illustrate that "diagrams are physical situations," consider how a picture or diagram conveys topological and geometric information through visual perception enables the illustration in Figure 4a to describe many relationships. However, each description conveys a different story about what is shown visually and therefore affords different inferences. Barwise and Etchemendy (1990, p. 22) thus note that a diagram can show "countless facts" (by which they mean one can construct multiple sentences from a diagram).

2.4.1 *Implications for sonic charts and graphs*

When Barwise and Etchemendy (1990, p. 22) refer to diagrams as "physical situations," they are referring to properties (and affordances) of diagrams that can interact with a human perception system. Designers seeking to extend the affordances of visual diagrams to the sonic domain are challenged to identify properties or dimensions of sound that similarly make use of "physical situations" (that interact with human perception) to enable multiple stories to be described about the relationships that are shown sonically.

Extending the example in Figure 3, the hybrid stereo–varying frequency interface should enable one to "hear the shape" of a contour. If text-to-speech labels are indexed to the contour, then the user should be able to form multiple sentences about the geometric and/or topological relations among the labelled elements.

2.5. Summarizing Extensions of the Graphic-Linguistic Distinction Into the Sonic Domain

Extending these classic graphic-linguistic distinctions thus suggests the following design opportunities when designing sonic versions of visual charts and graphs:

- The 2D versus sequential distinction suggests the need to identify perceptually distinguishable spatial properties of sound in order to afford the communication of spatial, geometric, or topological information.
- The analogue versus digital distinction suggests that analogue properties of sound, such as frequency, timbre, stereo, and echo could convey analogue properties of visual graphs.
- The relation symbols versus object symbols distinction suggests that analogue and spatial properties of sound noted previously could be recruited to map numerical values to perceptual dimensions.

- The intrinsic versus extrinsic constraints distinction suggests the need to identify "physical situations" that naturally emerge via the human perceptual processing of sound so that "countless facts" (building upon Barwise and Etchemendy, 1990, p. 22) can be inferred from those sonically conveyed physical situations.

2.6. Recasting the Graphic-Linguistic Distinction in Semiotics Terms

Although Shimojima is dissecting the characteristics of graphical and linguistic properties from the perspective of diagrammatic reasoning, semiotics has long been tackling this very problem. Semiotics can therefore provide us with a developed terminology and further insights regarding the differences between picture and language.

One tradition that can inform us comes from Lessing, who in *Laokoon* points out that while the link between the signs of language and their objects is arbitrary, pictures employ qualities of expression that carry similarities to their object. Scholars including Bayer, Wellbery, and Sonesson further developed Lessing's distinction and helped define the properties and communicative limits of language versus picture.

Another relevant tradition is Peirce's, especially his idea of the icon. Although iconicity is usually defined in terms of similarity to its object, Stjernfelt (2000) points out that Peirce has actually provided what Stjernfelt calls an operational definition or criterion for iconicity: "by the direct observation of it other truths concerning its object can be discovered than those which suffice to determine its construction" (Syllabus 2.279 as cited by Stjernfelt, 2000).

Our target design problem then can be characterized as creating, in the sonic domain, accessible graphics that is iconic in character. A key point here about Peirce's operational criterion is that it does not privilege vision; iconicity is thus in principle relevant also to audition, tactition, and so on.

Peirce's icon is one among the semiotic triangle of icon, index, and symbol. An index points to its object through an indirect relationship (e.g., smoke to mean fire), whereas a symbol's relation to its object is arbitrary (e.g., words and their meanings). In relating to Shimojima's graphic-linguistic distinction, graphics then correspond most naturally with icons, while linguistic representations correspond to symbols. In particular:

- "Spatial, geometric, or topological information" of a visual or sonic information display is akin to iconic properties of a visual or sonic display, and

- “Intrinsic constraints” (physical situations, as stated in Barwise and Etchemendy, 1990, p. 22) of a visual or sonic information display are akin to iconic properties of a visual or sonic display.

Peirce also developed a notion of diagram that was later interpreted by Stjernfeldt’s (2000). This notion of diagram will be compared and contrasted with this essay’s emerging notion in Section 3.4 (Part 5).

3. Provisional Model

I now introduce a basic model of perception and action, with the goal of integrating these ideas into a coherent system to inform design and aid understanding.

3.1. Part 1: Perception-Reaction to Environmental Change and Variation

Suppose an individual reaches for and grasps an object, such as a cup on a table (Figure 4a). Reflected light from the cup (Figure 4d) and its surrounding environment is picked up by retinal detectors and perceptually processed to inform a reaching and grasping action with fingers and hand positioned to grasp both the proximal and distal sides of the cup (Figures 5b–c).

This *perception-reaction loop* (cf. Gibson, 1986) comprises two relevant interrelated aspects: First, because the proximal side is visible, reflected light from the proximal side of the cup is picked up (cf. Gibson, 1986) and processed by sensory receptors. But because the distal side is invisible, the other aspect is the capability to anticipate, predict, or *simulate* (cf. Barsalou, 2009) the curvature of the distal side of the cup to inform a hand–finger orientation that is sufficient to grip the unseen distal surface (Hockema, 2004; Anstis, Verstraten and Mather, 1998; Goldstone, 1998; Freyd, 1992).

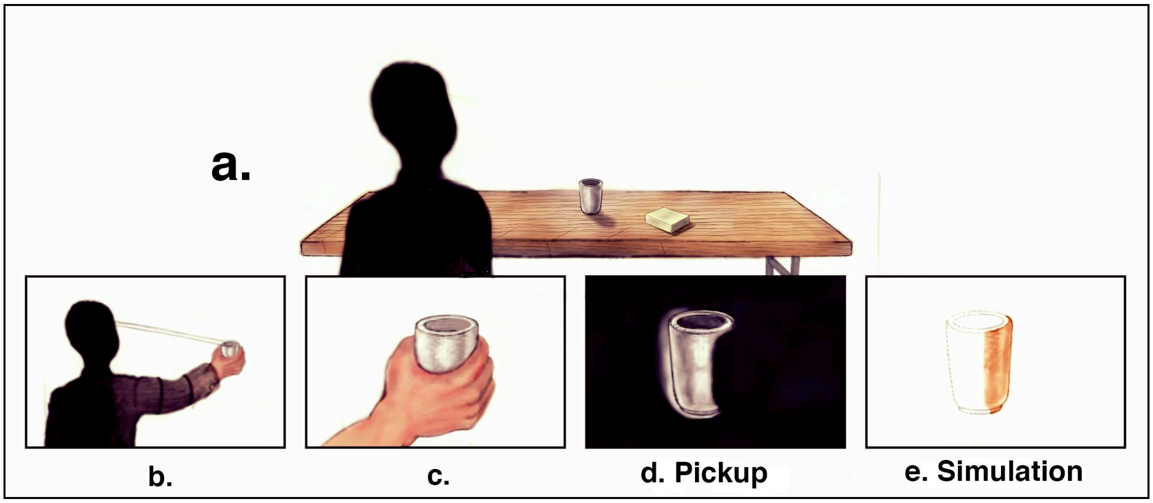

Figure 4. Information pickup versus simulation.

Because individuals predict or simulate an author's intended meaning when they recognize features of an external representation, prediction or simulation capabilities can enable individuals to use external graphic representations (Coppin, 2014). For example, based on prior experience in Western culture, an individual looking at a religious painting in a European church might predict that a flying white dove is intended to have a religious significance. Similarly, an individual reading the written word "cat" might predict that the author intended to convey the conceptual category of CAT (see Coppin, 2014, Chapter 3). This has been termed the "conventionalized" account of representation (Kulvicki, 2010).

Although this conventionalized account is uncontroversial for describing how individuals infer meanings intended by authors of written graphics such as text, applying the conventionalized account to picture perception is controversial (Gibson, 1960, 1971, 1978; Gombrich, Arnheim, and Gibson, 1971; Goodman, 1968; Kennedy, 1974; Kulvicki, 2010; Coppin, 2011, 2014). Many researchers have, instead, claimed that picture perception recruits unlearned, innate, or inherited biologically grounded capabilities to perceive and react in environments composed of occluded surfaces and edges (Gibson, 1960, 1971, 1978; Kennedy, 1974). Rather than describing picture perception capabilities in terms of innateness, Coppin (2014) describes how pictures make use of capabilities that inherently develop when learning to perceive and react within a physical environment composed of surfaces and edges.

3.2. Part 2: The Anatomy of Perception-Reaction

Recall that grasping the cup requires capabilities to pick up reflected light and to simulate the distal side. Memory traces of past perception-reactions (conjunctive neurons) are the resources from which simulations are constructed.

Simulation involves many of the same neural systems used during perception (Kosslyn, Ganis, and Thompson, 2001). If I simulate (imagine) a jet flying from left to right, I use many of the same processes and systems for perceiving an actual jet (Kosslyn, Ganis, and Thompson, 2001). In the cup example, when I perceive the cup, I also inform potential action (reaching for and grasping the proximal and distal sides of the cup). Thus, perception and simulation are integrated aspects of perception-reaction within a physical environment.

3.2.1 *Pictorial properties of graphics (and comparison to the semiotic notion of iconic)*

Let us now apply this simple model to external graphic representation. Suppose a viewer looks at the European painting described previously. Ambient light is reflected from the painted surface to produce optic arrays that are picked up by retinal detectors. The optic arrays are processed by lower-level perceptual categories and lower-level simulators to enable perception of a dove – a depicted object other than the painted surface. Even if the viewer has never seen a bird (or does not have a conceptual category for one), she can still perceptually process the spherical shape of the head, the cone-like shape of the beak, the hemispherical shape of the eyes, and so on, because she has spent a lifetime developing the capabilities to pick up and perceptually process the kinds of optic arrays that the artist has artificially produced via markings. These are the *pictorial properties* of a graphic.

Pictorial properties make use of lower-level *perceptual categorization* capabilities and simulators developed to perceive and react within environments composed of occluded surfaces and edges (geometric and topological relationships of an environment). Thus, pictorial properties inherently convey geometric and topological relationships associated with Larkin and Simon's (1989) diagrammatic definition. Pictorial properties are also defined as that which is picked up at the sensory surface when light reflects from a graphic. Thus, pictorial properties can be processed and interpreted for identification under multiple conceptual categories. In other words, pictorial properties are clearly on the analogue side of the analogue–digital spectrum. Finally, pictorial properties of graphics are clearly physical situations: The marks configured by the artist are precisely what produces the perceptual structure that the viewer picks up.

Iconicity is often described in terms of similarity to the represented object. However, as previously discussed, iconic representations also have an internal logic or set of rules. This is akin to Coppin's (2014) perceptual-cognitive approach to the pictorial-symbolic distinction because, in that account, graphic representations are composed of both pictorial and symbolic properties but to varying degrees. Returning to the dove example: During perceptual processing of the optic arrays from the painted surface, lower-level simulators enable perceptual processing of the depiction as a 3D shape. These are the same capabilities that enable perceptual processing of the proximal side in the cup example to engender simulations of the distal side of the cup. In other words, pictorial properties include an "internal logic" or "set of rules" (the simulations of the distal side of the cup or bird).

This integrates Coppin's perceptual-cognitive conceptualization of pictorial properties with the notion of iconicity from semiotics. To prevent confusion, this integrated notion of iconicity will be referred to as iconic′[2].

3.2.2 *Symbolic properties of graphics (and comparison to the semiotic notion of symbolic)*

In Coppin's model, symbolic properties make use of capabilities and simulators developed to categorize concrete structures configured by an author (for example, by marking a surface) and perceptually processed from an environment. Thus, symbolic properties inherently convey higher-level conceptual categories most often associated with and best communicated via language. If language is temporal (cf. Sonesson, 1988, p. 91) and temporality implies sequentiality, symbolic properties then more closely correspond to Larkin and Simon's (1989) sentential definition. Symbolic properties are also defined as the simulations intended by an author that fall under the author's intended conceptual categories, thus placing symbolic properties on the digital side of the analogue–digital spectrum. Finally, symbolic properties of graphics are the inverse of physical situations: They are less easily mapped back to what could be picked up from the physical world through lower-level perceptual categorization capabilities and simulators.

---

[2] In mathematics, the ′ (prime) symbol is often suffixed to an existing symbol to denote a related concept. We are borrowing the symbol here to suggest that although the integrated concept is closely related to iconicity, the two might not be identical.

Note that in Coppin's model, a perceptual system can still pick up the pictorial properties of writing from an unfamiliar language, but that individual may be unable produce simulations necessary for the symbolic properties (simulations of the author's probable intended meaning; see Coppin, 2014, Chapter 3). Because symbolic properties make use of higher-level simulators at the convergence of sensory modes and are more amodal, these simulations fall under conceptual categories that do not correspond to what could be perceptually processed from the physical world. Symbolic properties will therefore be arbitrarily related to their objects, coinciding with how symbols in semiotics are also understood in terms of their arbitrary (conventionalized) relationship to their objects.

Coppin's perceptual-cognitive conceptualization of symbolic properties is therefore quite consistent with the notion of symbolicity from semiotics. However, to prevent confusion, this integrated notion of symbolicity will be referred to as symbolic′.

3.3. Part 4: Model Extended to Sound (and Cross Modal Representation)

During perception-reaction, when memory traces of an Item A become encoded (the cup in the previous example), Item A is encoded *within a context* with other items (Barsalou, 2009), such as the table (Item B) that the cup was sitting on, my memories of how the cup felt when I grasped it (Item C), how it tasted (Item D), and how it sounded (Item E). Thus, as lower-level perceptual simulators and perceptual categories for vision develop over a lifetime, they develop *in networks with* other lower-level simulators and perceptual categories for hearing, touching, and seeing.

This results in what is known as cross modal correspondence. For example, in our physical environment, smaller objects vibrate at higher frequencies and larger objects at lower frequencies, so individuals developing in such an environment might thus "naturally" associate smaller objects with higher-frequency sounds and larger objects with lower-frequency sounds (Spence, 2011). Such natural associations could explain conventions that have emerged for representing sound that go far into human history, such as music notation systems (Figure 5), where higher-pitched sounds are represented by marks at higher elevations, and recent psychophysical research supports this claim (Parise, Knorre, and Ernst, 2014).

Figure 5. Music notations use a convention where marks at higher elevations represent higher-frequency sounds. *Note.* Bach, J. S. (n.d.). *Suite in G minor,* BWV 995 [Lute score]. Retrieved January 15, 2015 from http://commons.wikimedia.org/wiki/File:Bachlut1.png

3.3.1 *Sonic interface properties*

Let us now consider how iconic′ properties of sound can be conceptualized as the aural equivalents of the iconic′ properties of visual graphics (pictorial properties, according to Coppin, 2014). Similar to how light reflected from a marked surface of a graphic picked up by retinal detectors was conceptualized as the iconic′ properties of visual graphics (pictorial properties; Part 3), this section conceptualizes iconic′ properties of sound as the sound vibrations propagated from an object picked up by sensory receptors of the ear and perceptually processed as objects in an environment. Similar to how iconic′ properties of visual graphics (pictorial properties) are transduced into nerve signals and processed by lower-level perceptual categorization and simulation capabilities that developed over a lifespan to enable perception of occluded surfaces and edges, sound vibrations are transduced into nerve signals and processed by lower-level perceptual categories and lower-level simulators that developed over a lifespan to enable perception-reaction within physical environments with topological and geometric properties. Thus, similar to how a visual designer can configure a marked surface of a graphic to produce iconic′ properties of visual graphics (pictorial properties) that make use of lower-level perceptual categories and simulators to enable perceptual processing of surfaces and edges that are other than a marked surface, a sound designer can configure an audio device to produce iconic′

properties of sound that make use of lower-level sound perceptual categories and simulators that enable perception-reaction within a physical world with topological and geometric properties, to enable an individual to perceptually process topological and geometric relationships (iconic′; perhaps via the Doppler effect, stereo, or echo) that are other than the device that is producing the sounds ("sonic pictures").

Let us next recruit the iconic′ and symbolic′ definitions to more carefully conceptualize sonic versions of visual charts and graphs.

3.4. Part 5: Applying the Extended Model to the Graphic-Linguistic Distinction

Applying these iconic′ and symbolic′ definitions to extend Larkin and Simon's diagrammatic and sentential definitions requires replacing *elements* with *symbols′* (as in Section 2.6) and sequentiality with symbolized′ relations.

In a visual or sonic diagram, the relations among symbols′ are conveyed iconically′ (see Table 1) because they are picked up by sensory receptors via light/sounds reflected from a surface, and what is meaningful about what is perceived is the perceptually processed relations among the symbols′. In contrast, sentential relations are conveyed symbolically′ (see Table 1) because, although iconic′ properties of visual graphics are picked up by sensory receptors via light reflected from a surface, what is meaningful about what is perceptually processed are the author's intended simulation, and the conceptual category that the intended simulation is intended to fall under.

Table 1. Diagrams are composed of iconically′ represented relations among symbolically′ represented objects and sentences convey symbolically′ represented relations among symbolically′ represented objects.

	Diagram	*Sentence*
Relations	Iconically′ represented	Symbolically′ represented
Objects	Symbolically′ represented	Symbolically′ represented

Let us now compare this essay's notion of diagram developed thus far with Peirce's definition. In Peirce's system, diagrams are a type of icon, and are defined by their "skeleton-like sketch of relations between parts" (p. 363). The notion of diagram developed in this essay thus far, where relations among objects are conveyed via an iconic′ perceptual feature (e.g., visuospatial location, pitch, tactile size, etc.), while objects are denoted symbolically′, is potentially synergistic with the Peirce-Stjernfeldt account because iconically′ conveyed relations could be seen as akin to aforementioned skeleton-like sketch of relations among parts. The word diagrammatic, could thus be a term that refers to iconically′ conveyed relations among parts.

3.5. Applying The Model to an Example Design Problem

Let us now return to the WCAG text description example from Figure 1 in order to demonstrate how this model aids understanding and can inform design. The problem with the text description (Figure 1d) is that all content is conveyed symbolically′ (via text-to-speech) whereas the original visual graphic conveys much of the content via iconic′ (pictorial) properties of visual graphics. If a designer aims to present the chart sonically, how can the designer decide which aspects should be conveyed via symbolic′ properties (text-to-speech) and which aspects should be conveyed via iconic′ properties of sound (spatial sound)? Recall the pictorial and iconic′ definitions, where pictorial and iconic′ properties are predicted to afford the communication of concrete structures more effectively than symbolic′-linguistic properties, and an aspect of a graphic can be identified as more concrete if it produces a perceptual structure that corresponds to what could be picked up and perceptually processed from a physical environment (Part 3). The shape contour (Figure 1b) is thus primarily pictorial-iconic, and is therefore more appropriate for translation using iconic′ properties of spatial sound.

To determine the aspects of the graphic that should be conveyed via text-to-speech, recall the symbolic′ definition, where symbolic′ properties are predicted to afford the communication of abstract conceptual categories more effectively than pictorial or iconic′ properties, and a concept can be identified as more abstract if it is more amodal (Part 3), in that it is less easily mapped back to a structure that could be picked up and perceptually processed from a physical environment. The numbers that label increments on the *x*- and *y*-axis are thus more symbolic′

because they cannot be mapped back to a perceptual structure that could be picked up from a physical environment.

Figure 7 shows a spark line visualization of financial cost over time[3] (top left) and the corresponding spectral display of its sonic translation (top right). A spark line could be considered an abstraction (e.g., the value of a commodity) mapped to a concrete structure (e.g., a mountain's varying elevation from left to right). A text description of the mountain's shape would convey a string of conceptual categories that could refer to many possible concrete mountain shapes. However, sound frequencies that rise and fall with the elevation (spatial structure) of the drawn mountain (e.g., if panning from left to right) could translate the pictorially (iconically′) conveyed spatial structure of the mountain more accurately. Two additions to the varying sound frequency could aid shape comprehension: A baseline tone at middle C (akin to a horizon line) and "shading" under contours via bandpass-filtered white noise cut off at the varying frequency and at the horizon line.

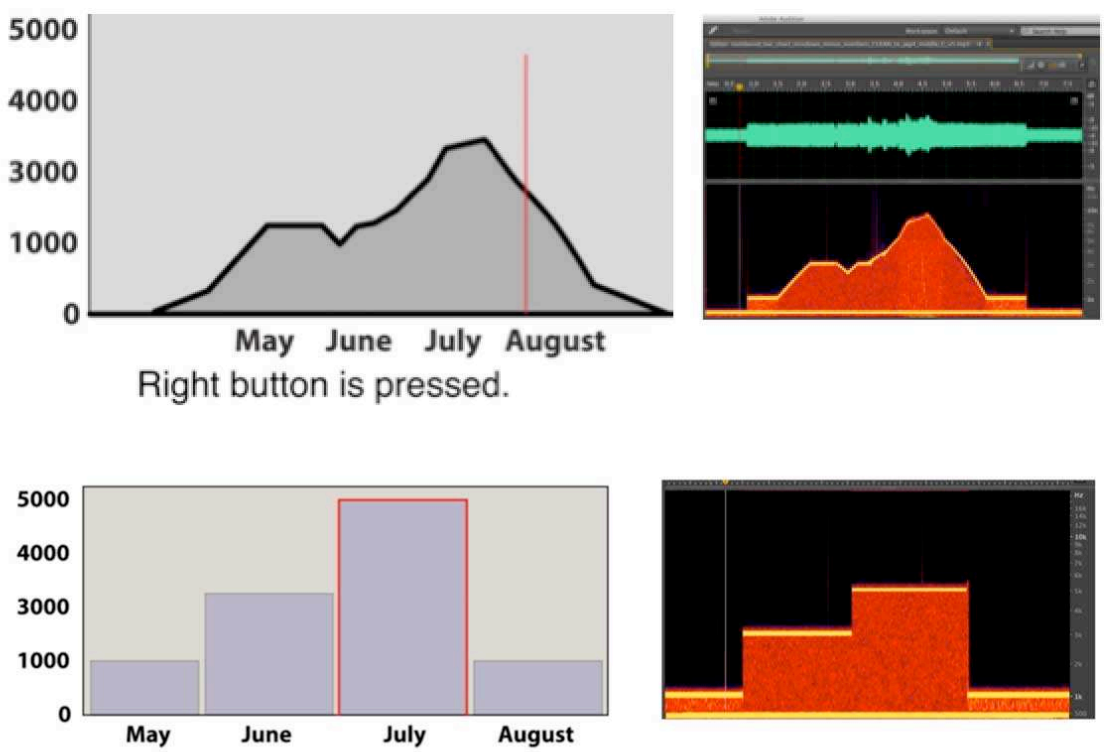

Figure 4. Spectral translations (right) of visuals (left).

3 For a movie that demonstrates this prototype, please go to http://perceptualartifacts.org/agi/research/sonification/

4. Conclusion

I have proposed a provisional model to underpin the various accounts of the graphic-linguistic and iconic-symbolic distinctions described in the literature. The model allows us to extend the graphic-linguistic and iconic-symbolic distinctions into aural and cross-modal domains.

The model distinguishes between two interrelated capabilities: Lower-level perceptual capabilities to pick up and perceptually process concrete structures of an environment, and higher-level capabilities to process and interpret how perceptually processed structures fall under more abstract conceptual categories.

The model conceptualizes iconic′ properties of both visual graphics and sound as what is picked up by sensory receptors and processed by lower-level perceptual categories and simulators that develop over a lifespan to enable individuals to perceptually process occluded surfaces and edges (topographical and geometric relationships) of a physical environment, thus enabling the perceptual processing of geometrical and topographical relationships that are other than the surface of the external representation.

The model thus predicts iconic′ properties to afford the communication of concrete structures more effectively than symbolic′ or linguistic properties. Also, symbolic′ properties are thought to emerge when an individual perceptually processes iconic′ properties of an external representation that an author intentionally configures to cause the perceiver to have a simulation that falls under the author's intended conceptual category. The model thus predicts symbolic′ properties to afford the communication of abstract conceptual categories more effectively than iconic′ properties.

By reverse engineering the classic graphic-linguistic distinction to more fundamental perceptual principles, I have introduced a means to understand how the distinction applies to sonic representations.

The proposed model streamlines definitions that distinguish diagrammatic from sentential structures: Text-sentences comprise symbolically′ represented relations among symbolically′ represented objects, whereas diagrams are iconically′ represented relations among symbolically′ represented objects. A sonic diagram is thus conceptualized as iconically′ conveyed sonic relations among objects linguistically (symbolicly′) conveyed via text-to-speech.

This proposed model enables researchers and designers to generate testable predictions for converting visual graphics into non-visual perceptual modes (other than the text-to-speech approach proposed by WCAG, which ignores the pictorial properties of graphics).

Acknowledgements

This research was supported in part by grants from the Centre for Innovation in Data-Driven Design and the Graphics Animation and New Media Centre for Excellence. I would like to thank research assistants Ambrose Li, Michael Carnevale, and Damon Pfaff as well as Dr. David Steinman.

References

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences, 2*(3), 111–117. doi:10.1016/S1364-6613(98)01142-5

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1435), 1177–1187. doi:10.1098/rstb.2003.1319

Barsalou, L. W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Carnegie Symposium Series: Building object categories in Developmental Time* (pp. 389–431). Majwah, NJ: Erlbaum.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*(1521): 1281–1289. doi:10.1098/rstb.2008.0319

Barwise, J., & Etchemendy, J. (1990). Visual information and valid reasoning. In W. Zimmerman (Ed.), *Visualization in mathematics* (pp. 8–23). Washington, DC: Mathematical Association of America.

Barwise, J., & Etchemendy, J. (1995). Heterogeneous logic. In B. Chandrasekaran, J. Glasgow, and N. H. Narayanan (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives,* (pp. 211–234). Menlo Park, CA: AAAI Press.

Coppin, P. W. (2011). Reconciling competing accounts of picture perception from art theory and perceptual psychology via the dual route hypothesis. In B. Kokinov, A. Karmiloff-Smith, & N.

J. Nersessian (Eds.), *European Perspectives on Cognitive Science: Proceedings of the European Conference on Cognitive Science EuroCogSci 2011*. New Bulgarian University. Sofia, Bulgaria: New Bulgarian University Press. Retrieved from http://nbu.bg/cogs /eurocogsci2011/proceedings/pdfs/EuroCogSci-paper200.pdf

Coppin, P. W. (2014). *Perceptual-cognitive properties of pictures, diagrams, and sentences: Toward a science of visual information design* (Doctoral dissertation, University of Toronto, Toronto, Canada). Retrieved from https://tspace.library.utoronto.ca/handle/1807/44108

Freyd, J. J. (1992). Dynamic representations guiding adaptive behavior. In F. Macar, V. Pouthas and W. Friedman (Eds.), *Time, action, and cognition: Towards bridging the gap* (pp. 309–323). Netherlands: Springer. doi:10.1007/978-94-017-3536-0_32

Gibson, J. J. (1960). Review of Art and Illusion. *The American Journal of Psychology, 73*(4), 653–654. Retrieved from http://www.jstor.org/stable/1419971

Gibson, J. J. (1971). The information available in pictures. *Leonardo, 4*(1), 27–35. Retrieved from http://www.jstor.org/stable/1572228

Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo, 11*(3), 227–235. Retrieved from http://www.jstor.org/stable/1574154

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*(1), 585–612.

Gombrich, E. H., Arnheim, R., and Gibson, J. J. (1971). On information available in pictures. *Leonardo, 4*(2), 27–35. Retrieved from http://www.jstor.org/stable/1572214

Goodman, N. (1968). *Languages of Art: An Approach to a Theory of Symbols.* Indianapolis, IN: Bobbs-Merrill Company.

Hockema, S. A. (2004). *Perception as prediction* (Doctoral dissertation, Indiana University). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3134120)

Kennedy, J. M. (1974). *A psychology of picture perception*. San Francisco, CA: Jossey-Bass Publishers.

Kennedy, J. M. (1993). *Drawing and the blind*. New Haven, CT: Yale University Press.

Klatzky, R. L., Giudice, N. A., Bennett, C. R., and Loomis, J. M. (2014). Touch-Screen Technology for the Dynamic Display of 2D Spatial Information Without Vision: Promise and Progress. *Multisensory Research, 27*, 359–378. doi:10.1163/22134808-00002447

Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*(9), 635–642. doi:10.1038/35090055

Kulvicki, J. (2010). Knowing with images: Medium and message. *Philosophy of Science, 77*(2), 295–313. doi:10.1086/651321

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65–99. doi:10.1111/j.1551-6708.1987.tb00863.x

Larkin, J. H., & Simon, H. A. (1989). Learning through growth of skill in mental modeling. In H. A. Simon (Ed.), *Models of thought. Volume II* (pp. 134–144). New Haven, CT: Yale University Press.

Mandler, J. M. (2006). Categorization, development of. In *Encyclopedia of Cognitive Science*. doi:10.1002/0470018860.s00516

Palumbo, G. (2009). *Key terms in translation studies*. New York, NY: Continuum.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences, 111*(16), 6104–6108. doi:10.1073/pnas.1322705111

Russell, B. (1923). Vagueness. *Australasian Journal of Psychology and Philosophy, 1*(2), 84–92. doi:10.1080/00048402308540623

Shimojima, A. (1999). The graphic-linguistic distinction: Exploring alternatives. *Artificial Intelligence Review, 13*(4), 313–335.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995. doi:10.3758/s13414-010-0073-7

Sonesson, G. (1988). *Methods and Models in Pictorial Semiotics.* (Report 3 from the project Pictorial meanings in the society of information). Retrieved from http://www.lunduniversity.lu.se/lup/publication/525910

Stjernfelt, F. (2000). Diagrams as a centerpiece of a Peircean epistemology. *Transactions of the Charles S. Peirce Society, 36*(3), pp. 357–384.

Author address

Peter Coppin, Faculty of Design, OCAD University, 100 McCaul Street, Toronto, Ontario, Canada, M5T 1W1. Email: pcoppin@faculty.ocadu.ca

About the author

Peter Coppin is an Assistant Professor in the Faculty of Design at OCAD University in Toronto, Canada, where he teaches inclusive design, industrial design, and directs the Perceptual Artifacts Lab (PAL). His research seeks to improve information displays through a better understanding of how designs make use of perceptual-motor capabilities. He received his PhD from the University of Toronto, Toronto, Canada and MFA from Carnegie Mellon University, Pittsburgh, PA, USA.

Ambrose Li is a Research Assistant with the Perceptual Artifacts Lab at OCAD University in Toronto, Canada. He received his master's degree from OCAD University.

Michael Carnevale is a Research Assistant with the Perceptual Artifacts Lab at OCAD University. He studied perceptual neuroscience at the Centre for Vision Research at York University. His research experience and interests include psychophysical methods, multisensory integration, and perceptual illusions.