Value-Added Text : Where Graphic Design Meets Paralinguistics

Stuart Mealing

Expressive typography is the sine qua non of the graphic designer—font styles and parameters such as size and color are selected to lend additional interpretive potential to a plain text message. When applied by a designer the process is intuitive and is hardwired to a particular text. Value-added text (**VAT**) is an attempt to visually extend the semantic potential of a message still further in a computer-based environment and to render the process both algorithmic and dynamic, its principles being applicable to typographic (and to iconic) text. This paper emphasizes the exploration of potential paralinguistic mappings which exploit and extend the traditional vocabulary of typography. Much that can be communicated in human-to-human language is lost in its transfer to text but paralinguistics—which studies the features of communication that accompany, or substitute for, the bare words used—offers a gateway to an enriched presentation of text. **VAT** therefore proposes automated graphic proxies which communicate more in a typographic message than the literal semantics of the user's native language and also offers potential assistance in cross-language communication.

University of Plymouth
Visible Language 37.1
Mealing, 43–58

Rhode Island School of Design
Providence, Rhode Island
02903

INTRODUCTION

When I first visited France I spoke expressionless schoolboy French and was understood by nobody. My wife on the other hand spoke only English but accompanied it with appropriate gestures and expressions and was universally comprehended. It was clear that the paralinguistic features of her language carried more meaning than the actual words of mine and had transcended the barriers of foreign language. In a similar manner the graphic designer will aim to communicate more in a typographic message than its literal semantics.

Most people have experienced problems responding to emails where the 'tone' of the message is ambiguous. While the word choice and syntax of formal written text gives controlled clues about whether the writer is being factual, cynical, angry or ironic for example, the closer stylistic proximity of email to spoken language does less to protect the recipient from the absence of contextualizing facial expression and intonation.

Elam subtitled her book on expressive typography 'the word as image' and it is the visual characteristics of a message—its image—that **VAT** employs. It replaces some of the meaning lost when rich layers of verbal communication are stripped off in the transposition to naked text. To some extent punctuation marks retain the flavor of spoken language, providing symbolic representation**!?*!**, and typographic options such as **BLOCK CAPITALS** can restore emphasis to the mental elocution that accompanies silently read text. Much that can be communicated in human-to-human language is, however, lost and paralinguistics studies those features of communication that are additional to the bare words used.

PARALINGUISTICS

In direct communication between humans the words spoken are supplemented by, or on occasion replaced by, a range of paralinguistic features such as body language, expression, gesture, intonation, volume, etc. which serve to qualify the bare bones of the message. VAT provides visual mappings for these features, mappings that can then be combined with raw text to restore some of its lost semantic support. These mappings can be abstract, symbolic, representational or mimetic and either static or kinetic. Their application can readily borrow from the grammar, syntax and vocabularies of art, design, film, theater and existing sign languages (subject to cultural variations and familiarity). Their operation within a computer environment provides opportunities for dynamic presentation.

A semiotic division can be made of the major signs and signals of communication into verbal (in which the linguistic element of speech can be separated from the expressive, non-linguistic elements known as supra-segmentals) and non-verbal (which divides again into visual and tactile). Of these the visual elements potentially have a very direct relevance to VAT and the prosodic features of pitch, loudness and tempo (which together contribute to the rhythm of language) offer readily measurable input data that could be automatically mapped to visual characteristics.

Other voice qualities, such as timbre and intonation, can give clues about the emotional state and social group of the speaker as well as about the information structure of the utterance—for instance whispering can tell of secrecy or conspiracy. They are, unfortunately, prone to international variation in interpretation and therefore only useful within a limited language domain.

TYPOGRAPHY

Although VAT was originally conceived in the context of iconic text (Mealing, 1993)—as a means of adding proxies for those features of human communication which could improve its international comprehension—its initial application to typographic text is perhaps more obvious. If consideration of cross-language issues is temporarily suspended then typographic text has some advantages as a starting point. Its features are comprehensively embodied in and easily accessed through a range of technologies, in particular computer appli-

cations, it is universally understood (by the literate) and, in normal practice, it enables clear understanding. Clear, that is, within the terms of reference of written language but without paralinguistic accessories.

Typography is the branch of graphic design which focuses particularly on type—choice of typeface, font size, color, layout and so on. In its 'pure' form it would deal exclusively with type, as opposed to type integrated with other visual material, though even a page of type necessarily has its own image quality as well. The discipline is, however, typically subsumed into the broader armory of the graphic designer where other visual and communicative skills come into play.

While the ability of the graphic designer to add expressive qualities to the literal meaning of text is relevant to VAT, it is interesting to consider the extent to which any such transformation could be described as objective. Boyd Davis (2000) makes enlightening comparisons between the methodologies employed in art schools and those used in the sciences (in the context of design for human-computer interaction) and finds strengths and weaknesses in both disciplines. Talking of computer science he notes that:

> *Conference literature is dominated by papers reporting quantitative assessments made under experimental conditions. There is little emphasis on creativity, but much on learning from research undertaken by others.*

commenting on the lack of objectivity in art and design training Boyd Davis says:

> *This approach emphasises the experiential nature of design [and its practitioners tend to be] weak at using any sort of objective or qualitative approach. … [There is a] resistance to theory of all kinds; and a reliance on a tradition of visual, intuitive solutions which are not based on any form of textural research.*

Tradition embodies rules that might have become explicit in areas such as leading and kerning but not in how to suggest joy, fear or indifference.

Despite that there would probably prove to be a high level of agreement about how to achieve such ends, and a still higher one about the success of such attempts. VAT sets out to derive testable rules from this and similar bodies of subjective agreement in the visual media in order that its subsequent application can be objective.

VISUAL LANGUAGE

While typographic text has inherent visual qualities there are arenas for its use in which the visual qualities are paramount. Concrete poetry, for example, uses the visual form of a poem to convey meaning and Houédard (1965), in text accompanying the exhibition 'Between poetry and painting' explains that:

The area 'between' Poetry and Painting is where they overlap—they do this (a) since all writing originates in painting (writing is painting words) and (b) since it is possible to think in images alone—in diagrams, models, gestures and muscular movements, as well as in words alone, mind is the first place where Poetry and Painting meet.

Writing and painting are ways in which minds communicate—and it is only in the context of communication theory and Wittgenstein's investigations plus the way artists and poets were influenced by Zen concepts of mind and meaning that a critic can see the sort of forces that have drawn poets and painters together in a common field.

The visual presentation of text can, of course, enhance understanding in other contexts. The layout pattern alone of a page of a high-level computer language such as Pascal tells a programmer much about the structure and function of a program and experienced authors claim to be able to tell whether a book is well written by the visual rhythm of its text-block patterns.

Hand-written text clearly speaks of its author and, as well as the 'general impression' it gives the layman—untidy, organized, decorative, etc.—it is open to graphological analysis. While its efficacy is widely doubted, this parascience interprets character and personality from handwriting and is increasingly used in

business as part of the applicant assessment process. Variables such as proportion, evenness, slope, slant angle, size, spacing, width and weight are all considered (West, 1999) and used to extract information beyond the remit of the text.

Type itself is protean, its production involving the specification of a number of variables. Choice of typeface (from many thousands), weight, size, color, kerning, leading and layout are the most obvious and the possibility of customizing or inventing new type families also exists. It will be obvious that some of these parameters might be defined by the procedure used—a typewriter, for instance, prescribing the way in which many of them are set and a display surface defining the maximum available size. The interaction between these variables is what distinguishes at a glance an up market broadsheet newspaper from a populist tabloid.

Many of these variables can also be applied to marks in general, both on a surface and in three-dimensional space. Lines can be heavy or light, wide or thin, smooth or angular, widely or closely spaced. They can also be regular or erratic, mechanical in character or autographic (thus revealing of the person making the mark). Gestures too could be said to have similar characteristics. Rules for the application of mark-variables such as these could be applied equally to the elements of typographic or iconic text and to ancillary features such as a dynamic background.

GESTURE

Gesticulating is instinctive and an integral part of the speech process that may even help to shape the thoughts underlying speech (Hall, 1988), a gesture being conveniently defined as a significant movement of a limb or body and also the use of such movements especially to convey feeling or as a rhetorical device. It has become elaborated and formalized into a number of sign languages, for example British, American and Amerind, all of which have both symbolic and mimetic elements the latter of which are often accessible internationally.

Most interesting in terms of building interactive dialogue systems is the semantic and pragmatic relationship between gesture and

speech. The two channels do not always manifest the same information about an idea, but what they convey is virtually always compatible, both semantically, in that speech and gesture give a consistent view of an overall meaning to be conveyed, and pragmatically, in that speech and gesture mark information about the meaning as advancing the purposes of the conversation in a consistent way. ... The semantic and pragmatic compatibility seen in gesture-speech relationship recalls the interaction of words and graphics in multimodal presentations. (Cassell, 1999)

There is a clear opportunity for the accompaniment of text by dynamic material that simulates or represents the gestures that accompany speech. A symbolic abstraction which embodies the visual rhythms of gesture might be appropriate or a more literal mapping for a gesture which is mimetic. The former might operate in the background and the latter perhaps adjacent to the message. Gesture mapping will, however, be pursued in depth elsewhere.

SEMANTIC GRANULARITY

Spontaneous signing, when used by my wife on a foreign excursion, is likely to accompany a single word or phrase and communicate a single semantic unit but when used by deaf children can be constructed, untaught, into sentences (Goldin-Meadows, 1998). Gestural communication is most likely to map to a word when the form of an object is mimed.

Semantic units in a message (or dialogue) will often not match discrete grammatical units of written language and it is necessary for VAT to decide the granularity of meaning at which a visual match is to be attempted. A 'top down' approach would attempt a single mapping for the message in its entirety while a 'bottom up' approach would seek to separately match each meaningful item. It might be tempting to start with a word-level mapping but the expression to 'kick the bucket' (a slang expression in England meaning 'die') clearly shows one weakness of such an approach while waving a fist demonstrates another.

The smallest contrastive unit of grammar is a morpheme and it would be convenient to find a matching entity which was the smallest contrastive unit of visual or pictorial

meaning—perhaps a 'picteme' (Mealing, 2000)—to which it could be mapped. Since the picteme is currently hypothetical a more reasonable aim might be to map to a lexeme (the smallest contrastive unit of a semantic system) but how often would a message's gestalt be sacrificed by such a forced division?

MESSAGE PRESENTATION

It is not the purpose of VAT to try to imitate spoken language but to add meaning to the text which is its visual incarnation. We can therefore make our own design decisions about how the text is best presented. This freedom is, of course, subject to the constraints of a given medium. The sort of text message considered here is likely to be either a piece of constructed prose, assembled over a period of time before transmission, or a real-time translation of natural language transmitted as it is spoken. In the latter case the application of VAT would necessarily also be in real-time and would be mapped directly from the selected inputs provided by the speaker. However when listening to speech only a single moment of sound wave is 'heard' at any point in time with the moments accumulating and acquiring meaning in the brain. It follows that the length of time that transmitted words linger and accumulate on a computer screen is an important factor in their presentation.

In the case of a constructed message it could either be presented as a single block, as an email is, or in sequential chunks—chunk size corresponding to some meaningful unit of either written or spoken text. Existing presentational models include the ticker tape machine or teleprinter (delivering the message character-by-character at the typing speed of the operator) and bullet points sliding into place in computer-based presentations, typically word-by-word or line-by-line; scrolling allows a message to be run through at a user's own pace while subtitled films present text in units roughly matched to short bursts of speech, usually semantically coherent and with punctuation used to aid comprehension. Other, perhaps more adventurous, dynamic options can also be considered.

VAT features, applied to whatever chunk size is deemed appropriate, can remain hardwired to the chunk and active for the life of the message on screen or can be

applied temporarily to chunks in sequence. For example if volume was to be mapped to font size then a word (or chunk) spoken loudly, perhaps for emphasis, would be represented in a larger font size than the surrounding text. In hard-wired application the increased size would be constant for the life of the message. In the case of temporary application it would either increase its size only at its first appearance (in the case of a streaming message) or, if the message was presented as a single block, then the size increase might be part of a sequential application mirroring the timing of the message's speech pattern—a sort of linguistic 'Mexican wave.'

SYNTACTIC OPTIONS

Any direct mapping of spoken text would necessarily reflect the syntax of the verbal message. It is not the case, however, that English syntax represents the only available model (or that its model is static) and different languages organize the grammatical elements of Subject, Verb and Object into different word orders. While the order SVO is common to seventy-five percent of the world's languages (Crystal, 1989), it is not always the case. Sign language, for instance, often starts with the Object and Spanish precedes questions with a Question mark. Yoda ("Strong am I with the force"), a character from the Star Wars series of films, uses an OSV construction and the contextual clarity of starting with an Object is tempting both in an unfamiliar linguistic environment and for an iconic language.

SYSTEM INPUTS

The mapping process around which VAT is formed uses inputs derived from speech and its accessories. These will, through the process of its application, find visual form. Readily quantifiable features such as volume, pitch and tempo have already been identified as the source of potential inputs while other equally apparent but less obviously quantifiable features such as expression, gesture and intonation present themselves. Instinctive gesturing presents rhythms (likely to be synchronized with the meter of speech) and dynamic spatial locations of the hands that provide relevant input data. In addition, future consideration could be made of sensory inputs which match inflections of verbal presentation—changes to pulse rate and its strength, for instance, might be indicators of anger while skin capacitance is measured as an indicator in lie detector tests.

A text message can be created either by real-time translation from speech using readily available software; in close to real-time by dictation to a keyboard operator; or it can be constructed over a longer period of time. In the first and second cases the message could either be completed before transmission or could be transmitted as it was being created. In the last case the authoring period suggests that it would be transmitted only after completion. The application of VAT features might therefore be required either on-the-fly, in which case automatic mapping suggests itself, or in retrospect as part of a post-composition editing process.

In this context it is unlikely that cool, post-composition application of VAT would offer exactly the same range of mappings as its 'hot' application in real-time. The expression of anger, for example, might provide real-time inputs of raised volume, quickened pulse, aggressive gesture or tremulous enunciation. Adding retrospective qualifiers for the same emotion would be strangely calculating. A secretary taking dictation could note expressive symptoms and transcribe them with VAT but the process lacks conviction. For this reason it is likely that some VAT features would best map to live inputs while others would fit more formal application. The latter are likely to provide a better basis for any future migration of VAT from screen to paper.

MAPPING OPTIONS

Given that we can identify and quantify salient features of language and have suitable variables to which they can be mapped then a semantic and presentational relationship between the two needs to be established. A convenient starting point is to map color space to prosodic space; a clearly simple but potentially rewarding exercise. The three prosodic dimensions of pitch, loudness and tempo can be mapped to a number of three-dimensional color models that exist—RGB (red, green, blue), CMY (cyan, magenta, yellow), HSV (hue, saturation, value), HWB (hue, whiteness, blackness), YIQ and YUV (luminance, two color differences), DLP (dominant wavelength, luminance and purity) and others. Each model has its own appeal as a mapping target but RGB has the obvious convenience of being the model used for computer display screens.

If loudness is mapped to red, pitch to blue and tempo to green on a scale of 0 to 256 then the combination of quiet, slow and high-pitched (50,50,200) produces deep blue; loud, fast and low-pitched (200,200,50) produces light greeny-yellow; very loud, fast and high-pitched (250, 200, 200) produces a light, warm gray; three maximum values produces white and three zero values (or silence) produces black. This is, however, not as visually intuitive as using HSV since it would seem appropriate to map volume to visual tone (quiet = light, loud = dark perhaps) but this does not equate to any single dimension of RGB space. While there would be few firm 'translations' from prosodic intention to colored representation the mapping provides an alternative contextualization whose meaning would accrue with exposure.

A further step in the project is to further research associational links, coded or instinctive, for instance between color and emotion and to test out a range of mapping options. Is silence black or white? Is anger red or blue? Is tenderness pink? Volume could alternatively be mapped to size or weight or capitalization or all three:

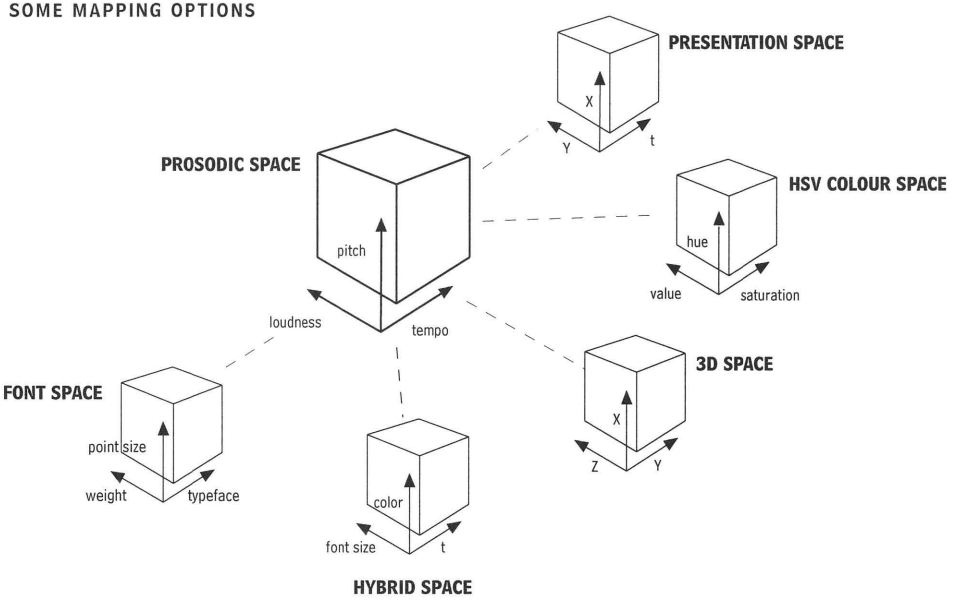
ok, **OK!**

Additionally in this simple example, the comma and ensuing space are diagrammatically suggestive of a pause, while the exclamation mark adds emphasis. You might merely read the first 'ok' but you almost hear the second.

A third option would be to map volume to spatial depth. Typographic convention offers each text chunk a canonical position in page space that can be subverted to good effect. In XYZ space mapping the Z-axis to volume suggests a closer (= larger) position for loud and a distant (= smaller) position for quiet. Three-dimensional manipulations such as this lead to interesting problems of overlap with which conventional leading and kerning are ill equipped to deal. The X- and Y-axes remain available for other mapping.

An emotional state which 'colors' a conversation or message could literally color text chunks or, perhaps more helpfully, the background of a text block. What dif-

SOME MAPPING OPTIONS



ferences would there be in the interpretation of a message on a rich red background as opposed to one on pale blue? The means of diagnosing an emotional state in order that it can be mapped using VAT is unresolved but since it is possible for an observer to recognize the existence of such a state in a speaker then symptoms clearly exist. Measured combinations of data from inputs already identified may prove sufficient for the recognition of quantifiable states; indeed mapping those individual inputs to visual variables might, of itself, give rise to a visual combination that could be recognized as representing the state.

One of the most evocative typographic features is the type family itself, each carrying hints and associations which can give clear inflection to a message. These are a few fonts that came installed on my current computer:

charcoal
apple chancery
arial
sand
comic sans
gadget
monotype.com
palatino
techno
textile

They can be modified in even the most basic word processors to: plain, bold, *plain italic*, *bold italic* and underlined. These simple modifications, combined with the few typefaces above, together produce fifty variations that could be selected by input mapping. The principles of line quality embodied in a sample range of fonts such as this could also be applied to the style of mark used in creating iconic or pictorial text.

METHOD

There are a number of distinct and separate elements to be dealt with in the realization of VAT, many of which can be developed in parallel. The speed and pattern of VAT's development will be partly dependent on cross-disciplinary collaborations. Contact is developing with departments in this and other universities regarding the range of useful data channels that can be derived directly from speech and from associated corporeal activity. Promising but far more speculative discussion is on the potential use of motion capture technologies to derive data (such as speed, scale and rhythm) from live gesturing and on the possible generation of data from expression by the inverse application of computer character animation techniques to facial analysis. Specific gestural semantics will be considered later in more depth.

Macromedia Director proves convenient for prototyping options with **Lingo** (its internal programming language) being called on for tasks such as the dynamic modification of font variables. Visual decisions such as those on choice of font and color derive from the extension of existing graphic conventions. Following informal user testing

the elements are then built into an application written in **REALbasic** which allows simple, flexible control and patching of input speech variables to screen-based output variables. This application will be used for rigorous user testing of the various options. At this stage a standard analogue-to-digital converter provides a small, hardwired range of data sets from pre-recorded texts and conversations.

With so many potential variables there is a clear need for constraint to prevent VAT from expanding into a fairground of visual information which could overpower rather than support the text it supplements. This will be treated as a graphic design decision by which active input channels will be restricted and/or the extremes of output parameters limited. The technical means by which VAT might be applied to computer-based transmissions such as email will not be considered until an effective mapping model has been proved.

OF DOGS, BABIES AND FOREIGNERS

Animals and babies do not understand the meaning of the words in spoken language. They are, however, likely to understand something of a spoken message addressed to them because it is often accompanied by exaggerated paralinguistic features in acknowledgement of their limited lexical comprehension. Firm hand signals accompany "SIT!", wide eyes and a dramatically large smile accompany "What a good little girl." Dog sits and baby smiles.

In the same way that television does not become rendered incomprehensible if the sound is turned off, so VAT is likely to convey some level of meaning when separated from the text it is designed to accompany. If therefore, VAT is successful in mapping paralinguistic features within its visual language then it is likely that they alone will elicit some response in receivers bereft of natural language. While not a primary objective of the project, the ability of the added features to communicate when divorced from text will be investigated.

Initial experiments with VAT are being conducted using English alone but if response to the visual accessories alone proves to communicate some useful part of, or context for, the message then its potential as an aid to cross-language communication would clearly be enhanced.

CONCLUSION

Written language can differ greatly in structure and function from speech but can also attempt, as often with casual email messages, to be a close substitute for the spoken word. Written text survives to be re-read while speech carries with it the preciousness of the ephemeral. VAT seeks to reinstate the semantic nuances lost when transient 'phonic substance' is translated to immutable 'graphic substance' (Crystal, 1989); to rehumanize text derived from spoken language through computer mediation.

Anything added to a message by VAT must add to its understanding or, at worst, be neutral to it. The VAT features must therefore be intuitively understood or their meaning comfortably acquired by exposure over time without a conscious learning process interfering with understanding of the underlying text message. It is not, therefore, necessary to concentrate on learning VAT but merely to experience it and let its meaning accrue. Even if no value is added all that is sacrificed is bandwidth.

References

Boyd Davis, Stephen. 2000. Educating the Multimedia Designer. In: Dudley E. and S. Mealing, editors. *Becoming Designers—Education and Influence*. Exeter, UK: Intellect.

Cassell, Justine and Mathew Stone. 1999. *Psychological Theories about Speech and Gesture in Interactive Dialogue Systems*. American Association for Artificial Intelligence.

Crystal, David. 1989. *The Cambridge Encyclopedia of Language*. Cambridge, UK: Cambridge University Press.

Goldin-Meadows, Susan and Carolyn Mylander. 1998. Spontaneous Sign Systems Created by Deaf Children in Two Cultures, *Nature*, 391.

Hall, Sarah. 1998. The Instinctive Gesture. *The Guardian* (newspaper), Nov 19th.

Houédard, Dom Sylvester. 1965. Chronology. Accompanying: 'Between Poetry and Painting,' an exhibition at the Institute of Contemporary Arts, London.

Mealing, Stuart. 1993. Talking Pictures. In Yazdani, M, editor. *Multilingual Multimedia*. Bristol, UK: Intellect.

Mealing, Stuart. 1994. A Computer HINTerface. *Intelligent Tutoring Media*, 5:2.

Mealing, Stuart. 2000. Do You See What I'm Saying?, In Yazdani, M. and P. Barker, editors. *Iconic Communication*. Bristol, UK: Intellect.

West, Peter. 1999. *Graphology*. Shaftesbury, UK: Element Books.

Author Notes

STUART MEALING is a Reader in Computers & Drawing at the University of Plymouth. With a first degree in Fine Art he exhibited widely before taking a post-graduate degree in Computing. Since then he has had five books and a range of papers published and holds research interests in visual language and in computer drawing. For many years he has taught art and design courses in the United Kingdom, was a founder of the Centre for Visual Computing at the University of Plymouth, a founder editor of the journal Digital Creativity and has been an honorary Research Fellow in computer science at Exeter University.