

# Prosodic Font: Translating speech into graphics

**Tara Rosenberger**  
MIT Media Laboratory  
20 Ames Street  
Cambridge, MA 02139 USA  
+1 617 253 4564  
tara@media.mit.edu

**Ronald L. MacNeil**  
MIT Center for Advanced Visual Studies  
265 Massachusetts Avenue  
Cambridge, MA 02139  
+1 617 253 4416  
ronmac@media.mit.edu

## ABSTRACT

The proliferation of speech recognition as input to Computer Mediated Communication (CMC) systems opens up new possibilities for the design of typographic forms. Designers can use the musical expressiveness of the speaking voice to shape letterforms in real time. Letters formed by speech are more representative of the emotional and contextualized person speaking than are fonts now. Prosodic Font is an object-oriented font that assumes a dynamic, temporal form. It emulates the tonal and rhythmic motion in the speaking voice. Preliminary user testing results show that people are able to identify Prosodic Fonts as representative of particular prosodic variations.

## KEYWORDS

Dynamic Typography, Glyph design, Prosody, Speech.

## INTRODUCTION

*“Compared to the richness of speech, writing is a meager system. A speaker uses stress, pitch, rate, pauses, voice qualities, and a host of other sound patterns not even vaguely defined to communicate a message as well as attitudes and feelings about what he is saying. Writing can barely achieve such a repertoire.”* [2].

Writing can begin to exhibit the expressive repertoire of speech if it uses speech itself in the design and display of typographic forms. Type is not designed to communicate to the reader how the writer was feeling at the time, the emphasis (s)he placed on certain words and entire thoughts, nor the rhythmic regularity heard in speech. Yet, with the common use of speech recognition systems as input, we can use speech as data, expressive of a particular person, in the design and presentation of words.

*Prosody* is the tune, tone and rhythm of the speaking voice. Prosodic variation conveys the speaker’s affective disposition and her feelings towards who’s she’s talking to and what she’s talking about [1, 3]; as well as emphasizes and clarifies interpretations of syntactic and discourse structures [5]. Speech prosody can change pitch, intensity, voice quality, and rhythm during a single syllable.

Prosodic Font is an example of using speech parameters in the real-time formation of a font. We describe below the design of the animated font, the notation of the affective speech examples, and the testing we have performed on people’s ability to relate Prosodic Font to speech prosody.

## FONT STRUCTURE

Previous work in text layout design demonstrated words presented serially through time as an alternative to the linear-spatial layout of text [7]. This work notes that words presented serially assume the characteristics of a “voice.”

Prosodic Font builds upon this work by dynamically changing the size, height, and width, and stem-thickness of *glyphs*, the visual form of the alphabetic letters, dynamically through time, in accordance with the prosodic changes in a speech file. Words are displayed for a duration that is related to the time it took the speaker to say it. Consequently, some words appear longer than others.

Designing a typographic font to move is different than designing fonts destined for paper. We hypothesize that decorative serifs, very fine lines and multiple forms of punctuation may make reading a temporal font more difficult. We simplified the letterforms to the minimum necessary to distinguish letters from one another.

Conversely, a prosodic font does need some things non-prosodic fonts do not. It must have means of representing the letters that describe a single sound, a *phoneme*, such as “th,” “sh,” and “ng,” as a single glyph. We call these multiple-letter forms, “phonetic ligatures.” In this way, phonemes are directly related to orthographic form. Unusual phonetic emphasis given a phoneme can be reflected graphically in the glyph representing that sound.

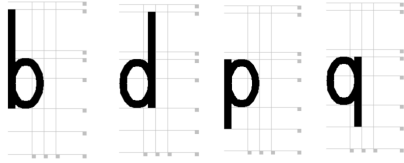
We designed a letterform system using simple Object Oriented structures. Existing continuous degree of freedom fonts, such as Adobe’s Multiple Master fonts or Knuth’s MetaFont [4], have too many parameters (in excess of seventy) to map systematically to changes in speech prosody. Additionally, prosodic fonts may not require the amount or kind of detail provided by traditional fonts. Each Prosodic Font glyph is comprised of one or more primitives shown in figure 1.



**Figure 1. These marks are the basic building blocks for any letterform glyph in Prosodic Font.**

These strokes are placed within a typographer’s grid space to create a single identifiable letterform as shown in

figure 2. To perform graphic transforms on a letter, we change the detailed grid parameters.



**Figure 2. Differences between glyphs ‘b’, ‘d’, ‘p’ and ‘q’ involve which primitive (the horizontal bar or circle) has precedence, and its vertical grid constraints.**

### SPEECH PROSODY NOTATION

Prosodic Font has no form nor duration unless given a description of speech; it is a computational description.

We created text descriptions of speech examples we collected. The sound corpus is excerpted from two speakers – one male and one female – who are telling stories about four different emotional experiences (joyful, sad, contented, angry) which produce extreme prosodic variation [3].

We labeled the speech using a combination of automatically generated Tilt prosody descriptions [6] which consist of a tune (abstraction of the F0 curve) and pitch accent types. In addition, we labeled by hand the highest legitimate amplitude found for each syllable, each syllable’s duration in hundredths of a second, and unusual phonetic emphases.

### RESULTS

Prosodic Font displays words sequentially, either left or centrally aligned on the screen. The “voice” appears to emanate from the point of text alignment.



**Figure 3. “Oh wow” was an excited exclamation that began softly and ended softly. This is a composite image.**

There are two variations on how we apply the prosodic notations to the font: (1) the prosody is applied per word, or (2) the prosody is applied per syllable, creating a word collage. The syllabic styled text can be read as a temporal whole, similar to how we hear speech.

Soft speech produces small, light fonts; loud speech produces large, weighted fonts. High speech lengthens the letterforms while low speech widens them. The words

stretch and lighten while the pitch rises, and squash and widen when the pitch lowers.

We studied people’s ability to associate Prosodic Font graphic motion to speech prosody. After playing a Prosodic Font file, we asked subjects to choose the speech file that “best matched” the Font. Twelve out of fourteen people matched the two Prosodic Font files to the correct speech file out of three possible choices. We observed that people paid especial attention to the similarity of rhythm and pacing across the mediums. As a result, any inconsistencies in computer animation speed will negatively affect people’s ability to understand the prosody correctly.



**Figure 4. “The sight of this huge beautiful setting sunset” was a part of a contented story. The words did not exhibit as much variation in size and shape than the “excited” words, and were lighter due to the soft tone.**

### FUTURE RESEARCH DIRECTIONS

Much work remains in both prosody and prosodic font characterization. To represent the voice, we need models of speaker voice quality, prosodic emphasis over and above pronunciation requirements, and means of associating these effects with particular syllables and sounds.

### REFERENCES

1. Bolinger, D. (1989). *Intonation and its Uses*. Stanford University Press, Stanford.
2. Gibson, E., Levin, H. (1975). *The Psychology of Reading*. Cambridge, MA: MIT Press.
3. Kappas, A., et. al. (1991). “Voice and Emotion” in *Fundamentals of Nonverbal Behaviour*. Eds. Feldman and Rime. Cambridge: Cambridge U. Press.
4. Knuth, D. (1986). *METAFONT: The Program*. Reading, MA: Addison Wesley Publishing Co.
5. Ladd, D. (1996). *Intonational Phonology*. Studies in Linguistics 79. Cambridge: Cambridge University Press.
6. Taylor, P. (1998). *Analysis and Synthesis of Intonation using the Tilt Model*. Draft paper. Available at: <http://www.cstr.ed.ac.uk/~pault/papers.html>.
7. Wong, Y.Y. (1995). *Temporal Typography: Characterization of time-varying typographic forms*. M.S. Thesis, MIT, Media Laboratory.

