

CAPTIONING PROSODY: EXPERIENCE AS A BASIS FOR
TYPOGRAPHIC REPRESENTATIONS OF HOW THINGS ARE SAID

By

Casey Irvin

Bachelor of Arts
University of Waterloo
Waterloo, Ontario, Canada
2008

A thesis presented to

Ryerson University and York University

In partial fulfillment of the
requirements for the degree of

Master of Arts

in the Program of

Communication and Culture

Toronto, Ontario, Canada, 2012

©Casey Irvin 2012

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

CAPTIONING PROSODY: EXPERIENCE AS A BASIS FOR TYPOGRAPHIC REPRESENTATIONS OF HOW THINGS ARE SAID

Master of Arts, 2012

Casey Irvin

Communication and Culture

Ryerson University and York University

This project explores a potential framework for expressing prosody in typefaces used for captioning video. The work employs C. S. Peirce's triadic form of the sign, specifically the icon and index; Theo van Leeuwen's exploration of the semiotics of typography and the voice; and George Lakoff and Mark Johnson's idea of experiential metaphors to form a theoretical underpinning that explains the meaning of speech and typography in terms of physical, bodily experiences. Seven typefaces were designed to show shouted, whispered, quick, slow, tense, relaxed, and trembling ways of speaking respectively. A series of three focus groups with deaf, hard of hearing, and hearing participants were held to evaluate the usefulness of these typefaces and, based on the results of a questionnaire and group discussion, alterations were made to the designs after each focus group. Bodily experience is found to be a potentially suitable groundwork for showing prosody in video captions.

Acknowledgements

I would like to acknowledge the hard work and guidance of my supervisor Jan Hadlaw, as well as my committee members Stéphanie Walsh Matthews and Richard Hunt. The support—both academic and otherwise—of these great people has been invaluable and has not gone unnoticed. Thank you.

Dedication

This thesis is dedicated to my partner, Emily Webb, for every time I woke her up in the middle of the night and every time she convinced me to relax. I would also like to dedicate this work to my parents, for always supporting my pursuits.

Table of Contents

1. Introduction.....	viii
“Prosody”	1
Goals and Limitations of this work	2
2. Past Research on Emotive & Prosodic Captioning	5
Debates, Pros/Cons	5
Past Research on New Ways of Captioning	5
Aspects Missing from Past Research.....	7
Current Practices in Place Around the World.....	9
Captioning Practices	9
Two Relevant Non-Captioning Cases	12
3. Semiotic Approach for the Project	19
Indexicality: C. S. Peirce	19
Triadic Form of Signification	19
The Semiotics of Typography	23
Typography and the Voice in the Work of Theo van Leeuwen	23
Experiential Metaphors.....	23
4. Typeface Development.....	27
The Potential of Type Design	27
The Ideal of Typography as the Crystal Goblet.....	27
Humanist Sans Serif Typefaces	27
<i>Open Sans</i>	29
The Design Process	29
Handwritten Samples.....	29
Designing a Digital Typeface	30
The Initial Designs	31
Loudness-Focused Typefaces: ”Shouting” and “Whispering”	32
Tempo-Focused Typefaces: “Quick” and “Slow”	34
Rhythm- & Pitch-Focused Typefaces: “Tense” and “Relaxed”	36
Timbre-Focused Typeface: “Trembling”	38

5. User Testing.....	39
Focus Groups Research Instruments	39
Video.....	39
Questionnaire.....	40
Group Discussion.....	41
Focus Group One	41
Key Findings in Focus Group One	41
Changes Made to Typefaces After Focus Group One	43
Problems with Implementation.....	46
Focus Group Two	47
Key Findings in Focus Group Two	47
Changes Made to Typefaces After Focus Group Two	49
Focus Group Three	51
Key Findings in Focus Group Three	51
Changes Made to Typefaces After Focus Group Three	53
6. Conclusions	55
Outcomes of Testing.....	55
The Potential of Prosodic Captioning.....	55
The Success of Showing Different Prosodic Elements	58
Considerations for Future Work.....	60
Final Summation.....	64
Works Cited	97

List of Tables

Table 1.....	32
---------------------	-----------

List of Appendices

Appendix A: Examples	65
Appendix B: Glossary of Typographic Terms	79
Appendix C: Study Questionnaire.....	80
Appendix D: Prosodic Type Samples	88

This page has been intentionally left blank.

Chapter One: Introduction

This project investigates the proposition that the design of a typeface can effectively communicate the prosodic information carried by speech. This is tested within the context of video captioning.

“Prosody”

In this work, I employ a definition of language that relies on social semiotician Theo van Leeuwen’s definition of semiotic resources.¹ For van Leeuwen, “semiotic resources are the actions, materials and artefacts we use for communicative purposes” coupled with the “ways in which these resources can be organized” (*Introducing* 3). Our everyday experiences show us that languages are not simply written and spoken as they may include gestural, musical, tactile, or olfactory elements, or be communicated in countless other fashion. A language could include a diversity of actions, materials, and artefacts strung together in a variety of combinations to create meaning. These resources can be just about anything. They can be created physiologically, such as with the voice, or with technology, such as with a writing instrument (3).

The term prosody, in the study of poetry, describes the systematic analysis of meter, rhyme, and stanza. When that scope is extended to literature more broadly, it also includes “speech-sound patterns and effects such as alliteration, assonance, euphony, and onomatopoeia” (Abrams and Harpham 256). In this work, I focus on an understanding of prosody based more on spoken language and linguistics that takes it to mean the suprasegmental² elements of speech

¹ Van Leeuwen states that “semiotic resource” is the social semiotic equivalent of the term *langue*, which he says is “the traditional semiotic term for the system of language” (*Introducing* 279).

² Suprasegmental elements of speech are expressed across multiple, temporally ordered discrete segments of speech (such as vowels, consonants, and syllables) (Dachkovsky and Sandler 289).

such as pitch, loudness, stress, rhythm, and timbre—*how* something is said. Prosody is an important element of spoken language.

Prosody is also part of the domain of non-verbal communication. Sign languages often rely quite heavily on non-word elements to convey the nuances of meaning. Facial expression, posture, rhythm/speed of speech, and the range of movement are no doubt chief among these prosodic elements of sign languages. In Israeli Sign Language, as Dachkovsky and Sandler have found, there is a similarity between the postlexical³ intonational melodies used in audible speech to indicate sentence type (such the rise in pitch at the end of an English question) and movements of the upper-half of the face (291). Prosody appears to be an important aspect in all non-written, linguistic human communication.

Goals and Limitations of this work

Captioning offers a superb opportunity for testing ideas around the typographic representation of prosody because it is, at its root, a visual translation of audible sound, especially speech. In setting the legal definition in policy documents, the Canadian Radio-television Telecommunication Commission (CRTC), the government agency charged with regulating the broadcast and telecommunications industries across the country, defines captioning as “the audio component of a television program in textual form” (CRTC 2009-430 2). This definition of captions should be contrasted with that of subtitles, the foreign-language counterpart to captioning. Whereas captions attempt to reproduce the whole auditory experience of a video, subtitles are simply a written translation of a different language (Brown 157). At their heart, subtitles are not a complete visual reproduction of sound, because the suprasegmental

³ Postlexical information conveys “functions, meanings, and relations such as sentence type, speech act, focus, and other aspects of information structure at the level of phrases, utterances, or the discourse as a whole” (Dachkovsky and Sandler 289).

meaning of the speech is still largely dependent on the viewer's ability to hear. Beginning with the understanding that captions have the potential to reproduce the whole field of sound, leaves open a space for prosody.

Ignoring the importance of prosody when captioning has the potential to confuse and frustrate captioning viewers because “without continual access to conversation modifiers and access to the paralinguistic components of human conversation [viewers may] misinterpret or misunderstand the semantics of television and come away from their viewing unsatisfied” (Fels *et al.* 2331). This confusion is liable to occur especially when a speaker is either not facing the camera, or off-frame entirely. Whispering and shouting narrators receive the same visual treatment in currently practiced ways of captioning. Further, while the effect of spoken prosody can on some occasions be read from the gestures, posture, or facial expression of a speaker on camera, any assumption that this is sufficient is naïve. As Fourney and Fels note, sounds (including prosody) can be used to oppose these typical markers of meaning for the sake of irony or sarcasm (571).

Investigating the potential of expressing prosody typographically is also interesting as its own endeavour. Whereas captioning provides a solid and practical basis for testing these ideas, and is as such the focus of this project, the outcomes of this research could also expand knowledge in the areas of literature (especially graphic novels), visual arts, design, and human-computer interaction, as well as anything else touched by a translation from the audible to the visual and *vice versa*.

I am placing certain limits on the scope of my work. Firstly, I will be focusing on a relatively small set of Roman characters. Each typeface includes the characters required to write in English with conventional punctuation. Extending the ideas developed in this project into

further languages will be possible in the future, but at this time working with a simpler sets of characters is more flexible and quicker to adapt. This allows for a greater range of ideas and theories to be tested. Also, my work does not aim to typographically convey the audible qualities of non-speech sounds (but may still do so serendipitously). For instance, the typefaces developed in this project are not intended to help show the difference between the honk of a car horn and the honk of a goose's call.

Chapter Two: Past Research on Emotive & Prosodic Captioning

Debates, Pros/Cons

Past Research on New Ways of Captioning

Other researchers have tried to provide the emotive elements of speech through the use of animation (Forlizzi, Lee, and Hudson; Rashid *et al.*; Fels *et al.*; Malik, Aitken, and Waalen; Rosenberger, and MacNeil), colour (Fels *et al.*; Fourney, and Fels), and non-character graphics and icons (Fels *et al.*; Civera, and Orero). However, in each case, the design of the individual letterforms has received little critical attention. These studies make little note of the effect that the typeface chosen has on the viewer's experience of captioned content. They also raise interesting considerations regarding three groups of users: deaf individuals, hard of hearing individuals, and hearing individuals.⁴

Rashid *et al.* begin their work from the proposition that kinetic typography⁵ has the potential to express emotion, mood, and tone of voice. They point to the extensive use of animated text in film and television title sequences as evidence for this potential (507), but focus on expanding ideas from the kinetic typography work done by others into the domain of video captioning. The past work they cite has shown that animations can enhance the emotional interpretation of written words (Wang, Prendinger, and Igarashi; Bodine, and Pignol) and that certain properties of animation, such as increases to size or direction of movement, can correspond to the prosodic elements of the spoken voice (Forlizzi, Lee and Hudson). Rashid *et al.* only included hearing and hard of hearing individuals in their study because the dialogue of

⁴ This division into three groups is an oversimplification. For instance, the Canadian Hearing society works to improve the lives of “culturally Deaf, oral deaf, deafened and hard of hearing” individuals (“About CHS”). In the case of the studies addressed and my own study, culturally deaf, oral deaf and deafened are grouped together under the term deaf.

⁵ This term, kinetic typography, is equivalent to animated typography.

the video captioned was signed rather than audibly spoken and, as the authors note, reading captions while watching sign language is confusing and redundant (508). The exclusion of deaf participants is unfortunate, but the researchers did find that hard of hearing individuals desired more sound information to be presented visually and that animated text could be a way to provide this (518).

A study by Fels *et al.* in 2005 found that while both deaf and hard of hearing participants expressed a desire for more details on the sound of a video to be presented, there was a good deal of discrepancy between these groups regarding what should be presented and how that presentation should be accomplished (2336). For instance, hard of hearing participants were found to like the use of graphics, icons, and colour as a means of providing some of the sound information, whereas deaf participants were generally unaccepting of these approaches to captioning (2336). A similar study from 2007, by Lee, Fels, and Udo, found significant differences between the opinions of hard of hearing and deaf participants; the opinions hard of hearing individuals were generally more favourable towards ‘enhanced’ captions than the opinions of deaf participants. In Fels *et al.*’s study, hard of hearing participants were more interested in the use of graphics, icons and colour than deaf participants. In the opinions of the deaf participants, colour was unacceptable unless it was part of the text (Fels *et al.* 2336). This preference for colour only in-text may suggest that a strictly static typographic approach might be preferred over the use of external icons or animation. I believe that this may be the case because with either typography or in-text colour, the viewer is not expected to attend to a third element on screen (the video and textual captions being the two pre-existent elements) or attend to extra space on the screen for captions—when the animated captions move outside the space required for a static line of text. Locating prosody strictly within a part of captioning that is

already present is a more economical and less distracting use of screen space. CRTC guidelines around captioning explicitly deal with the issue of on-screen interference between captions, and the other visual elements including the video, images, and other text (CRTC 2009-430 86). Icons and animations that add more, get in the way, and exaggerate this identified problem.

Aspects Missing from Past Research

Putting a greater emphasis on the typography may be a more successful way to convey the meaning found in prosody because it does not add new—potentially distracting—graphical elements to the captioned video as do animation or non-character graphics. Icons and typography also offer a greater range of potential choices than simply adding colour does, which can interfere with legibility when there is not enough contrast between the colour of the letters and the background colour. At the same time, my research fits with the findings of the past studies addressed above because its findings not preclude any of these other new approaches. Colour, animation and icons can easily complement the typographic approach I am suggesting, if appropriate.

The starkest difference between my work and the studies discussed above is the tendency of the latter to conflate emotion with prosody to ultimately privilege emotion. I have chosen to separate these two elements and focus on prosody over emotion for two reasons. The first reason is that prosody is easier to objectively codify. All spoken prosodic information can be reduced to the level of a sound wave. Stated simply, sound waves of a greater frequency are higher in pitch than those of a low frequency, and higher amplitude waves are louder than lower amplitude waves. Minute changes to these elements affect the shape of waveforms, which give a sound its particular timbre. Different sound waves mixing together can also create different timbres. Larger changes will give a sense of rhythm. This is too reductive to be practical in terms of

emulating human speech prosody typographically at this time, but it is important to note that there is no equally simple reduction that can be made for emotion. Loudness and pitch are quantitatively measurable in a way that happiness and sadness are not. Emotions therefore require a far greater amount of interpretation on the part of the person creating the captions. More interpretation by someone other than the viewer leaves open a space for more errors and ultimately achieves the opposite of its goal: less involvement from the viewer.

Following from this, the second reason to privilege prosody is that coding emotion takes the act of interpretation out of the hands of the viewer—a maneuver that is problematic because it runs contrary to the spirit of equal access, an important facet of accessibility. Focusing on prosody rather than emotion means that the act of interpretation remains much more in hands of the viewer. The emotional content of speech is often expressed in the prosodic features of that speech. In a way, prosody is a code for understanding the emotional state of a speaker. However, when new ways of captioning aspire to convey the emotional state of a speaker rather than the prosody of the voice, too much of the act of interpreting the meaningful emotional state of a speaker is taken away from the viewer. In my opinion, removing the opportunity for individuals to interpret the emotional states of speakers for themselves runs contrary to the spirit of equal access that Canadian society is moving towards. This movement is seen in official policies, such as the protections of persons with disabilities in the Canadian Charter of Rights and Freedoms, the Accessibility for Ontarians with Disabilities Act, and the CRTC guidelines discussed below. In their study that looked at the use of graphics, colour, icons and animation, Fels *et al.* noted that when producing captions for study purposes, their work became “an exercise in interpretation rather than verbatim translation” (2331). Any future improvements to captioning practices outside the laboratory that are based on Fels *et al.*’s approach would probably also need

to be non-verbatim reductions and therefore conflicts with Canadian broadcast regulations that state captions must be “verbatim representations of the audio” (CRTC 2011-741, Appendix ii). The fact that a viewer is unable to hear a speaker’s voice should not be grounds to assume that the viewer does not wish to, or is unable to, understand prosodic features within context and from that form their own conclusions about the emotional state of that speaker. A prosody-based approach aims at a degree of accessibility for the viewer of captioned video that is more-or-less equal to that of hearing individuals; it does not attempt to tell captioning viewers how they should interpret the emotional state of the speaker. I am certainly not claiming to have designed a set of typefaces that are as nuanced as the human voice, or devised a process that pre-empts interpretation by the captioner, but I feel that the experience of interpreting the emotion of a speaker is too important an act to be overly-determined by anyone other than the viewer.

Current Practices in Place Around the World

Captioning Practices

CRTC Regulations

The CRTC’s *Broadcasting and Telecom Regulatory Policy CRTC 2009-430: Accessibility of Telecommunications and Broadcasting Services* outlines the most recent set of goals and regulations concerning a series of stakeholder submitted issues. These include: relay services, which are operator assisted services to allow individuals with speech and hearing disabilities to make telephone calls; emergency telecommunications services; accessible wireless services; descriptive video; and closed captioning (CRTC 2009-430 2). Following from this document, two closed captioning working groups consisting of broadcasters, deaf/hard of hearing advocacy groups and other stakeholders were struck to suggest possible regulations for

Canadian closed captioning. One of the working groups oversees French-language broadcasting and the other English-language broadcasting.

In 2011, the French-language working group completed a set of standards that were adopted in *Broadcasting Regulatory Policy CRTC 2011-741: Quality Standards for French-Language Closed Captioning*, and changed their mandate to focus more on specific captioning standards for youth and adults over 55 years of age (CRTC 2011-741 28). The English-language working group has to settle a number of matters, but the most recent document, CRTC 2011-488, points many of these unsettled issues to the French working group's equivalent solutions. With these documents, Canada is moving towards increasingly accessible broadcast television. Regardless, the policies are still focused on the less ground-breaking, albeit important, aspects of captioning such as the acceptable percentage of misspelled words or whether a broadcaster needs to be concerned with ensuring captions and other on-screen information avoid visually conflicting or obscuring each other.

Additionally, the recent move from the Line 21 to CEA-708 captioning technology in Canada and the United States has brought with it a number of improvements to the technical possibilities of closed captioned video.

Line 21 and CEA-708 Captioning

The black box filled with white, pixelated, uppercase letters that most North Americans are familiar with for closed captioning is called Line 21 (see Appendix A, Example 5). This captioning standard received its name because it is transmitted on the 21st line—the last before the actual image begins—of the 525 lines used in an analogue NTSC television picture (Strauss 206–7). A properly equipped television decodes this 21st line of information and translates it into the black boxes and white words that the viewer sees. While it was remarkable at the time it was

first developed in the early 1970s (206), Line 21 is terribly limited. All Line 21 captions use the same typeface and small set of characters (“CEA-608 and CEA-708”). On 31 August 2011 most Canadian over-the-air broadcast markets switched to wholly digital signals (“On August 31”). Because of this, CEA-708, the captioning standard used in digital television, is now more prominent than Line 21. CEA-708 has a number of significant improvements over its predecessor. Most importantly here is: the addition of seven possible typefaces, as opposed to just one; drop-shadows, raised, depressed, and uniform “character edges;” support for more languages, including some that use a non-roman alphabet; coloured and translucent backgrounds for captions; and the possibility of multiple captions on-screen at the same time (“CEA-608 and CEA-708”). While the trajectory of these technological advancements shows that there is a general movement towards a more expressive approach to captions, the technology is still limited to only the set of typefaces that come imbedded in a CEA-708 equipped television or decoder.⁶

The Potential for Online, Open Captions

I frame my work within a less limiting range of technology. Online video, which is becoming increasingly accessible even on televisions, is potentially free of the need to adhere to the technical restrictions of closed-captioning for television. Rather than hiding captioning information within the video signal (as was done in Line 21 captions) or linking a video with a captioning file that is then superimposed over the video, online video has the potential for two separate video files to be created—one captioned, one uncaptioned. This means that there is the possibility of using captions that are free from the requirement that they can easily be hidden. Certainly, there is a greater efficiency provided by hosting a large video file and a small

⁶ Each television set includes a single typeface for each of these categories: monospaced with serif, proportional with serif, monospaced sans serif, proportional sans serif, casual, cursive, and small capitals (“CEA-608 and CEA-708”).

captioning file, rather than two large video files, but these new ideas must be imagined and tested at the level of where they could be, not where they are. Until that more efficient system can handle the approach that I propose here, these ideas should be considered applicable to only open-captioned videos. In this open framework, where the potential to convey prosody on screen does not need to fit many technological limitations, there are a few key systems for representing prosody in written texts that are used by millions of people on a daily basis. Exploring these can help build an understanding of how prosody is already shown visually. I point to Quốc Ngữ, a Vietnamese writing system, and modern musical notation as two informative examples.

Two Relevant Non-Captioning Cases

Vietnamese Alphabet

Tonal languages provide an interesting point to begin considering the possibilities (and limitations) of adding more prosodic information to the typographic word. Tonal languages employ, as part of their system of lexical meaning, elements such as pitch and timbre. While there are many tonal languages spoken, Vietnamese provides a particularly relevant example because it one of the few tonal languages that uses, like English, the Roman alphabet. *Quốc Ngữ*, literally meaning “National Language,” is a method of writing that has its roots in a Vietnamese/Portuguese/Latin dictionary first developed by the Jesuit Missionary Alexandre de Rhodes in 1651 (Nguyễn) to replace the ideographic (and difficult to learn) Chinese system that had historically been used to write Vietnamese. In the mid-1920, no more than 5% of the Vietnamese population were literate enough to read a newspaper, but by the end of the 1930s that percentage had doubled (Marr 34). Today, 90% of the Vietnamese population is literate (Alves 5). This huge

increase in literacy levels over just a few generations is remarkable and largely due to the increased reliance on *Quốc Ngữ*.⁷

There are five tonemarks attached to vowels in *Quốc Ngữ*—*Huyền*, *Hỏi*, *Ngã*, *Sắc*, and *Nặng*—with a sixth, unmarked tone—*Ngang*.⁸ Depending on the tone used in pronouncing a word, a great variety of meanings are possible. For instance, the letters *ma* when spoken can mean, “ghost,” “mother,” “but,” “tomb,” “horse,” or “rice seedling” depending on the tone used in the pronunciation (Nguyễn). Pitch is a very obvious component to the differing tones in spoken Vietnamese, but “the laryngeal features of creakiness and breathiness are primary in signalling tone and... pitch height is derived from these features and from features describing tonal shape” (Pham 2). That is to say, these diacritics denote a complex mixture of pitch, timbre, and loudness.

Quốc Ngữ is an example of one approach that could be taken to show prosodic elements of speech. This project could have taken a similar approach by complementing a plain style of text with additional information in the form of new characters or diacritics, but I have chosen to follow a different route because new elements could cause similar problems to those found in the icon-based approach of the studies addressed above. With the addition of a second level of characters to a modified set of Roman letters, a far greater number of Vietnamese speakers than ever before were able to gain literacy skills. But, teaching someone to read for the first time is a far cry from drastically *changing* the way that someone reads. Age and experience with standard captioning have been noted as a potential barrier to accepting new and more expressive ways of captioning (Lee, Fels, and Udo 12). Adding to a person’s reading process additional characters to

⁷ These advances could, of course, not have also been possible without to a series of literacy campaigns and improvements to educational institutions.

⁸ See Appendix A, Example 3 for a diagram and description of the pitch changes of each tone.

attend to could be at this time distracting, especially to an older individual with decades of reading experience in a particular language. As I stated above in regards to icons, animation, and colour: diacritics may still be acceptable for conveying prosody. However, my work aims for an approach that will not be as problematic for established readers, and can therefore form the basis for future ways of captioning that could potentially include meaningful diacritics.

Showing captioned prosody in an way based on the diacritics used in *Quốc Ngữ* could be especially distracting to those established readers because *Quốc Ngữ* marks tone at the level of the vowel. While this is a familiar practice for readers of *Quốc Ngữ*, the English readers in my study will be unfamiliar with it. An approach this fine-grained could mean that if an entire sentence is shouted, all syllables would have the same diacritic attached to represent the increased loudness and slight raspiness they share. A basic solution to this is the use of diacritics only at regular intervals, such as the beginning of each line, or only where changes to prosody occur. This latter option is tempting and does align with the current practice of marking changes to who is speaking in captions at each occurrence, but it is necessary to explore both these ideas more fully. To do so, I turn to the system of modern musical notation.

Musical Notation

In *The French Revolution: A History*, 19th century historian Thomas Carlyle writes about Jeanne-Marie Phlipon, a noblewoman on her way to her death. Carlyle includes an account of her written by a man named Riouffe. According to Carlyle, Riouffe describes Jeanne-Marie Phlipon as a person that “expressed herself with a harmony and prosody that made her language like music, of which the ear could never have enough” (289). Even in witnessing the experiences of a person headed for the guillotine, it was clear to Riouffe that music and speech cannot often be wholly separated. Following this idea, contemporary musical notation is largely a written

approach to notating prosody. In fact, the etymology of “prosody” traces back to the ancient Greek word *προσῳδία*, *prosōidia*, meaning a song sung to music (prosody, n.).

I wish to ground this discussion in a concrete example. Appendix A, Example 3 shows three lines of vocal music from a French Opera version of Shakespeare’s *Hamlet*, with music by Abroise Thomas and words by M. Carré, and J. Barbier. Looking at this short section we can see very explicit directions regarding the tempo, pitch, loudness, timbre, and rhythm of Hamlet’s voice. In musical notation, the elements of prosody are treated individually and in a variety of ways. I see essentially three ways of providing prosodic information in musical notation. There are elements that are marked with a regular frequency, elements that are marked when they change, and elements that are being marked continuously.

I want to begin here by highlighting the elements that receive regular, metered demarcation and what these types of elements might say about captioning prosody in video. In musical notation, clefs designate a particular centre for the range of pitches. Bass, tenor, alto, and treble clefs⁹ all align with an expected range of notes. Just as a person’s voice falls naturally within a range, different clefs afford different ranges of notes. In this example, the bass clef is used.¹⁰ Along with the key signature, it is customary for the clef to be rewritten at the beginning of each line of music. I can see value in this approach for musicians rehearsing and learning a piece of music, but not for marking prosody in captions. When reading a longer score from anywhere other than the first bar, regular demarcation removes the need for a musician to refer

⁹ There are more clefs, but these are the commonly used clefs for modern music.

¹⁰ While any note can be written in any musical clef through adding extra lines above or below the staff at the particular note, not all notes are represented on the staff (the five horizontal lines and four spaces between those lines) in a specific clef. Notes are easier to read when they fall on the staff, so the composer/arranger’s choice of clef is intended to help a musician more easily read the music. Certain instruments are conventionally, but not necessarily, written in a particular clef.

back to the beginning of a piece or search for the last change to a particular parameter before they begin. This is unnecessary with video, which is generally watched from the very beginning.

A second set of prosodic information is marked when a relevant change is made to it. The most salient elements in this category are tempo, loudness, and timbre. In this piece, gradual increases to loudness, crescendos, are marked with either “*cresc.*” or a symbol that look similar to an elongated <, while gradual decreases, decrescendos, are shown with a narrowing, elongated >. Finite points in loudness are indicated with *f* for fortissimo (loud) and *p* for piano (quiet). These types of changes are marked once and then assumed to continue until a new loudness is indicated. This approach is translatable to captioning and could be employed in interesting ways, but it is an unsuitable model for my work because it still requires the experienced reader to learn a new set of icons/markers before it can be useful.

The final set of information is in a constant state of flux. As long as the score continues, pitch and rhythm are marked with different notes. Breaks and rests, where no note is played, are included in this system too. Written notes are a multimodal communicative act. That is, the notes represent the expression of multiple semiotic modes (van Leeuwen, *Introducing* 281)—in this case rhythm and pitch—in a single character. The typographic word has a similar multimodal quality (42); it expresses a lexical mode (which letter is it, what word do the letters form) and a typographic mode (how do the letters look). In this way, the notes on a written piece of music stand as an example of effective multimodal characters. It could very well be the case that just as a musician can easily interpret two types of information simultaneously, the reader of a typeface can interpret the lexical meaning and typographic meaning of a word or phrase and understand them together. With my proposed typographic approach to representing prosody, different typographic meanings are perceived at the exact same time and in the exact same eye movements

as the lexical meaning are.¹¹ This aligns with Fels *et al.*'s finding that in-text colour is preferred over colour appearing as an extra element outside the text.

As an instructive example for this research, I take musical notation as both a caution away from over-directing the reading experience and as an example of successful multimodal written communication. While the basics of musical notation can be learned in a few hours, ultimately it is a complex system that takes years to master. It breaks apart distinct elements of prosody and provides instructions to the reader in different ways for each of rhythm, pitch, timbre, tempo, and loudness. A novice player might be able to understand simple melodic lines with only a small amount of practice and instruction, but once they are asked to attend to changes in tempo, loudness, or key, in addition to melody and rhythm, the complexity of a score can quickly become too much for the beginner. Captioned video is also already an experience of multiple types of information at once. Video, textual, and typographic information (as well as auditory for some viewers) combines into a single experience and adding a new stream of information to that may be disruptive to established readers. A thin sliver of what is included in written music, the notes, illustrates a fundamental aspect of the approach I take in this project: communicating two types of information effectively with a single character. This project works to expand the typographic choices available to video captioners. By improving the typographic mode that is already in use, my project offers a more elegant and less invasive basis to conveying

¹¹ It would be possible to mark changes to prosody only at the instances where they occur through changes to typography—i.e. the first shouted word would be given a different typeface and shouting is assumed to continue until another word is marked as a whisper, at which point whispering is assumed to continue. This will give more or less visual salience to the words at the points where prosody changes and is therefore potentially problematic. This may also be disruptive to the reading process in a similar fashion to added icons and diacritics if the eye is drawn towards these words with extra emphasis applied to them.

prosody (and not emotion) than does adding animation, diacritics, or icons, but without necessarily blocking those other approaches from complimenting my work.

Chapter Three: Semiotic Approach for the Project

Indexicality: C. S. Peirce

A semiotic approach is used in this project as a theoretical basis for conceptualizing the translation that must be made between audible prosody and its visual representation. I begin my thinking here with the work of Charles Sanders Peirce. Best known for his philosophical system of pragmatism and his typology of signs, C. S. Peirce was a philosopher, logician, and mathematician from Cambridge, Massachusetts (Perron and Danesi 23).

Triadic Form of Signification

In my work, I employ Peirce's definition of the sign. He defines the sign as "something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign" (Peirce, *CP* vol. 2 par. 228). From this, Peirce understood the process of signification in a triadic form comprised of the representamen (also referred to as the sign), object, and interpretant. The object is the something that a sign stands for (Peirce, *CP* vol. 1 par. 339). This could be a physical thing or a concept. Finally, the interpretant is mental effect of contact with that sign, the sign in the mind (Peirce, *CP* vol. 1 par. 564). The letters *C A T* together are a sign that stands for an object, a feline animal, and in my mind this produces an interpretant, the notion of a feline animal.

Within Peirce's framework, signs are determined by the object that they stand for in three different ways. Firstly, signs that are determined by a degree of likeness are called icons. These are connected to their objects because they resemble that object in some way (Peirce *EP* vol. 2 pag. 460–1). For instance, a paint chip is an icon for the paint inside a can because it resembles the paint's colour. The second way that an object can determine a sign is at the level of the index,

where there is a real and direct causal relationship between the object and the sign (*EP* vol. 2 pag. 461). As an example, a weathervane is an index of the wind because there is a causal relationship between the wind blowing in a direction and the weathervane turning to point that direction. The third and final way that an object can determine a sign is in the form of a symbol (*EP* vol. 2 pag. 461). Conventions and laws are what make this connection. Written language is symbolic because it is by socio-cultural agreements that we have come to connect the letters written on a page and the meanings that those words and phrases signify. Another example of a symbol is a stop sign. A red octagon has no connection to the idea of stopping a vehicle other than by social conventions that dictate that meaning to be present in red octagons.¹²

Peirce in *Typography and the Spoken Voice*

My approach focuses on the indexical nature of both spoken prosody and the typographic word. Prosody is an index of the production of sound in the human body. Higher pitches are a result of tightening vocal cords; louder speech comes from stronger muscle contractions and more air being forced out of the lungs. On the other hand, the indexicality of typography can be traced through the history of typographic letterforms based in the body and the materiality of pens, brushes, chisels, and anything else used to make a mark on a surface.

Movable type took root in a western context around 1450 with the work of Johannes Gutenberg in Mainz, Germany (Füssel 14–5). A style of ceramic type had been developed in China and Korea as early as the eleventh century, but the 10,000 plus characters required to write the languages meant that printing was too time-consuming and complicated with this technology

¹² It is a simplification to think that a stop sign is solely a symbol. The colour red is often used to alarm people—such as in the case of fire trucks, alarm bells, and emergency exit signs—so its use is indexical of the need to alarm someone into stopping. The colour red on a stop sign also has iconic significance because its use draws a resemblance between the stop sign and other things, like those listed above, that are intended to grab a person's attention. The same sign can therefore be a mixture of Peirce's three types of signification.

for it to take any serious hold (14). In contrast, most languages in Europe could be printed with two sets of 26 letters—majuscules and minuscules—and a few punctuation marks. Gutenberg made a crucial choice when cutting and casting his characters that would resonate in type design for centuries. He chose to attempt to reproduce the entire appearance of the manuscript pages that scribes were producing at the time. To do this he replicated the double column layout and margin proportions of the written manuscript that his 1282 page Bible was based on, and more central to my work here, created a whole set of characters to mimic the variation that individual letterforms had when written by a scribe (17–8). To that end, Gutenberg’s set of characters included 47 capitals, 63 lowercase, 92 abbreviations, 83 ligatures, and 5 punctuation marks (18) in the sharp, angular style of writing known as Blackletter (see Appendix A, Example 3 for a sample of a Blackletter typeface).

Two centuries prior to Gutenberg’s breakthrough, scribes in other parts of Europe, especially Italy, were tending to cleaner, less ornamented forms based on a revival of the “Roman” letters introduced by the Emperor Charlemagne in the 8th century (Hutchings 18–19). These far more closely resemble the serif typefaces we are familiar with today. In 1470, the brothers Johannes and Wendelin da Spira in Vienna began printing in a typeface based on these Roman forms (Chappell 68). This typeface and those that followed by the da Spira brothers’ contemporaries such as Nicholas Jenson (68) were based on the handwriting of Italian humanist scholars participating in this typographic revival (Willen and Strals 10). These humanist typefaces (see Appendix A, Example 4 for a sample of a humanist typeface based on the type of the 15th century), as they have come to be called, “are closely connected to calligraphy and the movement of the hand” (Lupton 46). There is an indexical connection between the hand and the shapes of these letters.

While Gutenberg's characters are based on the handwriting style used by scribes in northern German at the time and the Venetian printers emulated a humanist style of handwriting, it is crucial to note that the typefaces are actually removed by a degree from the indexical signification of the movement of a hand. Typography can always be traced back to shapes originally created with the movement of some part of the human body. However, as type design has evolved over the five and a half centuries since Gutenberg, the index of the hand has received varying degrees of prominence. It is important to understand that the index of the hand has been able to remain present in typography even as technology moved modern type design away from hand tools like the pen or chisel and into the world of computer screens and keyboards. This is possible, in a Peircian framework, because of unlimited semiosis.

In discussing this process, Peirce notes that the interpretant of a sign is capable of becoming the representamen for a further process of signification (*CP* vol. 5 par. 138). It can create in the mind "an equivalent sign, or perhaps a more developed sign" (vol. 2 par. 228) as he writes. For instance, if I am looking at a photograph of a forest canopy and in this photograph smoke is rising from a section of trees, the photograph as a whole is an icon of that scene. The photograph resembles the actual forest. The smoke in the picture is an icon of smoke, but that smoke in my mind is also an index of the fire that it is rising from. A typeface may only be iconic of the handwritten word, but in the mind of a reader that handwritten word that a typeface is an icon of, is indexical of a hand writing. In the first level of signification, a humanist typeface is the representamen that iconically resembles an object, handwriting. This produces an interpretant in the mind of the viewer. That interpretant, the mental effect of experiencing a typeface as iconic of handwriting becomes the representamen for a second degree of signification. This representamen, handwriting, is indexically connected to an object, the

movement of the hand. That further has the mental effect of bringing another interpretant into the mind of the viewer.¹³ While not immediately present, there is a connection between a typeface—even if shown as pixels on a screen—and the movement of a hand across a surface. Eventually, once sufficiently inscribed into the social consensus ruled by law and convention, the indexical or iconic representation becomes a symbol.

The Semiotics of Typography

Typography and the Voice in the Work of Theo van Leeuwen

Theo van Leeuwen is a theorist whose work focuses on the semiotics of the voice and the semiotics of typography. Two key pieces of his writing strongly inform my work in this project. Van Leeuwen discusses the semiotics of the voice (“Voice”) and typography (“Typography”), but stops short of any bold steps to connect the two. My project can be taken as an attempt to connect a semiotics of the voice with a semiotics of typography. Van Leeuwen relies largely on the same theoretical basis to understand both the experience of reading the typographic word and listening to the spoken voice: George Lakoff and Mark Johnson’s idea of the experiential metaphor.

Experiential Metaphors

My definition of metaphor begins with the definition used by Murray Knowles and Rosamond Mood. These authors define it as “the use of language to refer to something other than what it was originally applied to, or what it ‘literally’ means, in order to suggest some resemblance or make a connection” (2). I hone this definition slightly with van Leeuwen’s idea that, the fundamental quality of metaphor is “the idea of ‘transference,’ of transferring something

¹³ See Example 5 for a diagram illustrating this process of signification.

from one place to another, on the basis of a perceived similarity between the two ‘places’” (*Introducing* 30). Metaphor, then, can be understood as a way of explaining something in terms of something else by transferring ideas from one area into another based on similarity.¹⁴ This means that to explain prosody in terms of typography, there must be a suitable way to transfer ideas between the two. To find those appropriate means of transference, I turn to the idea of experiential metaphors and Lakoff and Johnson’s assertion that “no metaphor can ever be comprehended or even adequately represented independently of its experiential basis” (19).

Lakoff and Johnson outline a series of common conceptual metaphors found in Western (American) English. One conceptual metaphor they look at is more-is-up (19–20). Commonly used expressions for the more-is-up metaphor are ones such as “my income *rose* last year,” “the number of errors he made is incredibly *low*,” “he is *underage*,” and “if you’re too hot, turn the heat *down*” (16, emphasis in original). Connecting the two sides of this metaphor, more and up, requires some sort of experiential basis. For instance, when more of something is put into a pile, that pile grows and becomes higher; held upright, the mercury in a thermometer rises higher as temperatures increase. These are experiences that we can use to understand the connection between two sides of a metaphor.

Discussing the argument-is-war metaphor that Lakoff and Johnson employ in terms of Peircian semiotics, R. Lance Factor claims that “metaphors are a kind of implicit formula for what would be a perceived similarity, if we were to pass from the situation or event as described to the situation or event as actually encountered in experience” (232). The similarity means that the one end of the metaphor is iconic of the other end (232). So, even though I have never been

¹⁴ This is a more transparent definition of metaphor than Peirce provides in his discussion of a particular type of icons, the hypoicon. Here he writes that hypoicons “which represent the representative character of a representamen by representing a parallelism in something else, are metaphors” (*CP* vol. 2 par. 227).

to war, if it is described to me in terms of a situation or event I have experienced (an argument), my experience with arguments gives me a basis to understand war— that is, if war shares a resemblance with argument. Relating this to typography and spoken prosody, the typography-as-prosody metaphor is understandable if there is a perceived similarity between the *experience of type* and the *experience of prosody*. Even for an individual that has never heard spoken prosody before, if the experience of a typeface is similar to the experience of a prosodic feature, then the connection between that typeface and that prosodic feature will be more implicitly understandable—and therefore more meaningful—than an *arbitrary* connection between a chosen typeface and a prosodic feature.

While Ellen Lupton’s claim that “typography is what language looks like” (1) might be an over-simplification ignorant to gestures and sign languages, it does point to an important metaphorical connection between audible speech and the visible word. Representing vocal prosody in terms of an experiential phenomenon that can be understood by individuals who can hear as well as by individuals who have been deaf their entire lives is a complex matter. It is certainly true that many deaf individuals are familiar with the experience of speaking and the physical sensation of pushing air through vibrating vocal cords, but without the experience of hearing others speak, I fear that moving directly from speech to typography will be ineffective way to show prosody. The word metaphor is actually a metaphor itself: originally the term meant ‘transport’ (van Leeuwen, *Introducing* 30). Here I suggest that the experience of handwriting (while not universal) is a common enough experience to act as way to transport between the prosodic dimension of audible speech and the written word.

I will provide an example to better explain thoughts expressed above. Take, for instance, a voice that is trembling. The most salient aspect of a trembling voice is that the volume

oscillates quickly between being louder and softer. This quality is produced by the flow of air being pushed with varying strength through the vocal cords and out the mouth. As I speak with a trembling voice, I can feel that it is the muscles in my abdomen, chest, and throat contracting and relaxing quickly that produces this audible quality. The spastic action of these muscles is an experience that resembles and is resembled by (iconically signifies) the experience of muscle spasms in my arm. When handwriting, a muscle spasm will cause the lines drawn to be wavy. With this in mind, I can then create a typeface that resembles this waviness iconically. Since the typeface is iconic of my spastic handwriting, which in turn is indexical of the movement of my hand, and the movement of my hand is experientially iconic (metaphorical) of a trembling voice, there is a connection that can be made between the typeface and a prosodic quality.

Chapter Four: Typeface Development

The Potential of Type Design

The Ideal of Typography as the Crystal Goblet

In 1932, the typographer, writer, and historian Beatrice Warde gave an influential speech to the Society of Typographic Designers in London, England (Swanson 91), where she drew the comparison between the ideal of typography and a crystal wine goblet (Warde 91–3). According to Warde, just as the goblet is designed “to *reveal* rather than to hide the beautiful thing which it was meant to *contain*,” so too should typography strive to present the meanings of words and phrases while in no way imparting its own character on the experience of reading (91, emphasis in original). Warde’s idea is problematic because she takes a very narrow view of the possibility of reading—and wine drinking, for that matter. Her views become even stronger when she further compares typographic style to the experience of listening to another person speak. She claims that “if you begin listening to the inflections and speaking rhythms of a voice from a platform, you are falling asleep” and that “type well used is invisible *as* type, just as the perfect talking voice is the unnoticed vehicle for the transmission of words, ideas” (92, emphasis in original). As I have discussed above, a great deal of the meaning we derive from spoken communication comes from being attentive to those very inflections and speaking rhythms. My design work strives to contradict Warde’s arguments on the nature of typography. It is not the case that *all* meaning must be presented lexically in the words; there is ample space for us to *also* employ the meaning of type treatments.

Humanist Sans Serif Typefaces

Around the 1920s, a style of typefaces called geometric sans serifs began to appear (Cheng 15). These represent a distinct movement away from the role of the body in creating

letters. Based on “an often naïve or rigid use of geometric shapes such as circles” (Baines and Haslam 82), these typefaces were favoured by the modernists of the time for their “perceived alliance with ‘machine age’” (Cheng 15). The way letters are constructed in geometric sans serifs has more to do with the use of rulers and compasses—which were later replaced by computers—than it does with brushes and pens. It is certainly true that perfectly drawn circles and straight lines were employed in the design of even older typefaces, however in these predecessors to geometric sans serifs, these precision tools appear more obviously in the design of smaller components, such as the curve of a serif, than they do in the shape of the letters as a whole. The geometric face *Futura*¹⁵ (See Appendix A, Example 6 for a sample of this typeface) for example, has many (near) perfectly round letters, and the weight of almost all strokes across all letters is equal.

While this movement obscured the role of the human in the creation of letters, another style beginning around the same time maintained the appearance of the hand in a sans serif style. Early humanist sans serifs, such as *Johnson* or *Gill Sans* (See Appendix A, Example 7 for a sample of Gill Sans), have features that in some ways emulate the humanist typefaces discussed earlier and have “calligraphic structures (particularly the *a*, *g*, and *t*” (Cheng 15) but do not wholly look like they were handwritten. That is to say, they iconically signify the indexical movement of the hand creating letters. While maintaining a simple structure, free of serifs, that can allow for a wide range of manipulations.

¹⁵ Please note that there are no conventions for citing typefaces or type families in MLA. Therefore, I have adopted the approach of showing families in italics and the names of specific typefaces in quotation marks, as is done with albums and songs or anthologies and poems. Therefore, *Futura* and *Times New Roman* would be type families, but “Futura Condensed Medium” and “Times New Roman Bold” would be specific typefaces.

Open Sans

To create my prosodic typefaces, I began with an already existing typeface called *Open Sans* that is released under Apache License Version 2.0 (see Appendix D for an example of the regular weight). This typeface was developed by a designer named Steve Matteson, the type director at Ascender Corp (“*Open Sans*”). *Open Sans* is described as a “humanist sans serif typeface.... with an upright stress, open forms and a neutral, yet friendly appearance” (“*Open Sans*”). It has been optimized for print and screen reading and has excellent legibility (“*Open Sans*”). The Apache License allows me to freely develop and distribute derivative works based on Matteson’s compiled files, which are provided through Google. While there are protections in Canada that would allow for this work as fair dealing within an academic context, by starting my creation of prosodic typefaces from font software released under a license that specifically allows the creation and distribution of derivative work, I am ensuring that the knowledge developed here has a much greater chance of moving beyond theory and could potentially see some use in captioning beyond the project. This was an important consideration for undertaking this work.

The Design Process

Handwritten Samples

I return to the example of a trembling voice that I have discussed above to illustrate the flow of my design process. Part of my focus was on the use of the hand and the experience of the body in typography, so actually handwriting lines of text was an important starting point for my designs. This step often began for me with thinking about the experience of my own voice and prosody. I would consider how, as I spoke, the experience felt within my chest, lungs, throat, mouth, lips, face, and anywhere else affected by speaking in that prosodic style. With a

trembling voice I could feel my diaphragm and vocal cords quickly oscillating between contractions and slight relaxations. With these muscular sensations in mind, I tried to consider and reproduce what I saw as their analogue in my arm. That is, I tried to write in a way that could be understood as a metaphor for this sensation. I wrote words and phrases with my arm, hand, and wrist muscles quickly moving between different degrees of contraction to iconically show trembling. For different styles, I found that different writing instruments could enhance the iconic value of my handwriting, so I tried a variety of pencils, pens, brushes, and markers at this stage (see Appendix A, Example 8 for samples of these drafts).

Designing a Digital Typeface

The computerized side of designing was not a simple matter of reproducing of my own handwriting as a useable font file. I still wanted to ground my work in a clear and effective typeface that would have more visual clarity than my own handwriting does and provide a consistent set of letterforms to tie different prosodic styles together under a single family. To do this, I had to distil from my handwritten samples the elements of the letters that showed the index of a particular bodily sensation the best and try to find ways of adding those elements to the regular weight of Steve Matteson's *Open Sans*. With "Trembling" that meant deforming the letterforms to give different strokes the undulating, unsteady appearance found in my handwriting.

The software I used to create the font files of my typeface designs is called Glyphs. When designing in this program, a main window is coupled with a preview window below. This allows a designer working in Glyphs to zoom in close enough to a letter in order to make very precise changes to its shape at the same time as previewing it in the context of other letters and at a different size (see Appendix A, Example 9 for a screen capture of this). This was an extremely

helpful part of my workflow because it allowed me to work on my designs while seeing in real time how my letters looked at the size used for captioning and in the context of complete sentences.

Glyphs also allow the simple creation of multiple master fonts. Multiple master fonts are typically be used for generating interpolations between different master styles. For instance, a designer might create condensed and bold versions of a particular typeface and then use Glyphs’ interpolation feature to calculate and automatically generate a bold-condensed weight. While I did not really use this feature in the way it is intended, I was able to employ it to my advantage. I used multiple masters to be able to quickly compare different degrees of the same manipulation. For example, I created three different degrees of wavy deformation as three different masters for “Trembling.” This allowed me to simply and quickly contrast different amounts of manipulation applied to the same sample text without even switching program windows on my computer.

The Initial Designs

I have designed seven different typefaces for this work.¹⁶ This is not an exhaustive list of how something can be said, but I have chosen these different styles because they each help to highlight the potential for including certain aspects of prosody in type design. I am calling these seven typefaces “Shouting,” “Whispering,” “Tense,” “Relaxed,” “Quick,” “Slow,” and “Trembling” (See Appendix D for a complete sample of these in their finalized form). Below in Table 1 is a brief description of the prosodic dimensions that I feel are emphasized in each of the different typefaces. I have selected these particular styles of speaking because they represent both a wide range of prosody and are distinct enough to avoid any confusion among each other.

¹⁶ A Glossary of typographic terms is provided as Appendix B.

For instance, screaming is not included at this point because it shares too many similarities with shouted and trembling speech. I will begin here with a discussion of my first version of the typefaces.

	Loudness	Rhythm	Tempo	Pitch Range	Timbre
Shouting	Loud				Slightly Raspy
Whispering	Quiet				Breathy
Quick			Quick		
Slow			Slow		
Trembling	Shaky			Shaky	Shaky
Tense		Jerky		Stressed	
Relaxed		Even		Gliding	Smooth

Table 1

Loudness-Focused Typefaces: "Shouting" and "Whispering"

"Shouting"

When a person is shouting, their voice is louder and rougher. The muscles used to push air out of their lungs exert more force and the timbre of their voice becomes rougher from this strain. I have split my discussion of "Shouting" along these lines—loudness and timbre—to better explain how the typeface conveys the prosodic features of a shout.

As Nina Nørgaard points out in her elaboration of van Leeuwen's typographic work into literary texts, the connection between visual salience and sonic salience is a very obvious part of the "iconic meaning-potential of typography" (150). As an example, she notes how the exclusive use of majuscules (writing in all caps) clearly conveys sonic salience iconically in terms of visual salience (150). But, salience requires context. The exclusive use of majuscules is only visually salient if the reader comes to them with the context that most writing does not look this way. The visual salience of a typeface can therefore be based on just about anything, so long as it looks

sufficiently different. Nørgaard stops short of offering an explanation for why the visual salience of ALL CAPS resembles sonic salience, but in my framework I propose that one reason for this could have something to do with materiality. The sonic salience of increased loudness can be expressed metaphorically by increasing the weight of strokes used to draw letters because in both cases there is an indexical connection between the amount of material used in the communicative act and its salience. The material of audible speech is air. In verbal communication, the flow of air is manipulated in the lungs, throat, and mouth before entering the world and traveling to the listener's ear. Loudness is therefore an indexical trace of the amount of air used. This is similar to the physical experience of using a writing instrument where more ink (or graphite, or paint, or *etc.*) means more visual salience. In either case, the increased salience of the communicative act is indexical of the increased use of a material resource. To show this, I manipulated the letterforms to appear as though they had been written with a thicker instrument than the regular typeface was.

In looking at the voices of hard rock singers, John Shepherd noted that the shouting/screaming style they employ is both loud and rough (cited in van Leeuwen, "Voice" 431). Vocal roughness is indexical of irregular vibrations of the vocal cords (Laver 128). These vibrations are not unlike a marker that is running out of ink. Erratic vibrations of the vocal cords impede an even flow of air from the lungs. A marker going dry is similar in that there is not an even flow of ink from the tip. We know through our experiences that "roughness comes from wear and tear" (van Leeuwen "Voice" 429) and that the idea of wear and tear includes a sense of use over time and expendable resources. So, when I write something with a dying marker, I understand that there is a physical difficulty to leaving behind the resource (the ink) that typically evidences the marker's indexical movement across a surface. This experience is

metaphorical of the roughness that accompanies loudness when one shouts. To show this, the letter forms were made to look distressed. Strokes are not perfect and small notches are missing from the letters.

“Whispering”

A whispering style of speech was added after the first focus group. The design of the typeface was an attempt at the letters being present while not being as salient as the other typefaces, especially “Shouting.” To accomplish this, the letters were created with hatchings of small rounded rectangles at a five-degree angle. This has the effect of making the letters appear lighter in colour when rendered on screen at the size used for captioning. This lightness is meant to call towards a hand writing without much pressure applied. It is as though a pencil was used, but the experience is such that the letters could be imperceptible at a distance or without special attention. The muscles of the arm applying a light amount of pressure to the page is metaphorical of the diaphragm applying a light amount of pressure to the lungs. Here, visual salience is based on using less material than the regular typeface does. Once the whole letters had been hatched, some of the hatch marks were removed to further increase the sense that the letters are not wholly there.

Tempo-Focused Typefaces: “Quick” and “Slow”

Obliqueness

The technique of slanting characters was common between the “Quick” typeface and the “Slow” typeface in my initial designs. This idea was largely rooted in the way that we view italics in writing. In his discussion of metaphoric projections of movement in italicized letters, Phil Jones claims the “users creatively construct meaning from italics using the repertoire of bodily experiences, image schemas, conceptual metaphors and conventionalized structures

available to them” (260). For Jones, this creative construction involves a blending of the typographic treatment of a word and its lexical meaning. He focuses on slanting the words “run” and “car” to highlight this meaningful mixture. The experience of movement in a car involves a force from below the body at the wheels that pushes backwards on the motorist, and therefore a left-leaning slant seems to convey movement in a left-to-right reading language (260). However, when we consider the experience of running, a person’s upper body leans forward, so a right-leaning slant seems more natural in this case (260). Slanting is a powerful typographic treatment for conveying a sense of speed, but since there is room for contention between the lexical meaning of a word and the metaphoric representation of movement in typefaces, it may be imperfect, for instance, when “run” slants to the left. The effort to create a slanted typeface that can convey quickness, regardless of the lexical meaning of the word required me to make a choice. I feel that the physical experience of running ties more directly to the body than does being in a car, where speed has to do with the hidden inner working of a drivetrain not the physical sensation of swinging one’s arms and legs back and forth. Being a more bodily experience makes running better for understanding the metaphorical connection between quick speech and my “Quick” typeface.¹⁷ So, was “Quick” slanted to the right, while the slow typeface, in contrast to this, was slanted to the left.

“Quick”

With “Quick” I also attempted to show a sense of speed with the appearance of efficient handwriting forms. That is, some letters connected in an almost cursive fashion when beside each other. These forms are indexical of a hand moving to create letterforms without taking the time to lift the writing instrument for each stroke. I describe the style as “almost cursive”

¹⁷ I concede that the experience of running might be a bit of an ableist argument on which to base these thoughts, but driving is also ableist and somewhat classist.

because I wanted to stop short of creating a typeface that looked overly like script-style handwriting. I did this to try and steer clear of any of the connotations of class that might be associated with ornate scripts and the garishly cursive. The letters that connected in version one are *he*, *ne*, *me*, and *te*. The terminals of some letters were made into sharp points to show the quick pull of a writing instrument off of the writing surface.

“Slow”

In addition to slanting left, to show a counterpoint to the rightward slant of “Quick,” the “Slow” typeface was based on the idea of a thick, inky way of writing. The initial design was an aim to show a viscosity in the way that the letters are created—almost as though the words were written in molasses. Many corners were rounded, the strokes were thicker here than in all of the typefaces other than “Shouting,” and the strokes at the bottoms of letters were thickened even more than those at the tops of the letters. Furthermore, some letters were also given an even more pronounced appearance with especially thick strokes. Particularly, the *o*, *e*, and *g* appeared darker than other letters. This unevenness was intended to invoke an indexical connection between a material that is thick and the inability of the hand to move through this thickness with ease.

Rhythm- & Pitch-Focused Typefaces: “Tense” and “Relaxed”

“Tense”

The letters of “Tense” were created to display the indexical movement of a ridged arm with quick, straight strokes and sudden changes to direction. The main concept here was that the letters appear to have been written by a hand and arm with tightly contracted muscles. This aligns iconically with the style of speech that is produced through clenched teeth or with the strong, sudden bursts of sound indicative of a tense diaphragm and throat. As van Leeuwen

writes, “The sound that results from tensing not only is tense it also means tense and makes tense” (*Introducing* 33). Whole words appeared jagged in this typeface because the angles used in drawing were very sharp and many of the letters were not perfectly aligned to the standard baseline and x-height of the typeface. That is, some letters were intentionally moved vertically off of the baseline they would normally sit on, while other letters went above or below the x-height in places they typically should not. Overall this design idea was to make the words appear sharp and angular.

“Relaxed”

In contrast to the “Tense” letterforms, I created a relaxed typeface. According to van Leeuwen, “the opposite of the rough voice is the clean, smooth, ‘well oiled’ voice from which all noisiness has been removed” (“Voice” 429). When the hand writes smoothly, it acts in much the same way. Fluid and unrestricted movements of the muscles in both the throat and the arm have a similar effect. When I speak with a relaxed voice, my vocal cords do not impart the same rough quality heard in a shout. In relaxed speech the different elements of prosody glide between degrees. Loudness gradually moves between loud and quiet, pitch slides between different frequencies, and the rhythm of speech is even in a relaxed voice. This is because the muscles in the body do not exert force or change tension in a jerky way. This is similar to a relaxed arm writing; the lines are smooth and curves are easily drawn. Tensing the arm muscles and drawing a circle will leave a shape that is not as smooth or circular as a relaxed arm can.

To show this relaxed quality, I designed curvy letterforms with no sharp corners. Even the ends of strokes were rounded off. The lower right terminals of the *h*, *m*, and *n*; and the upper left terminals of the *u*, and *y* curve towards the centre of the letter and then back out, extending below or above the rest of the letter respectively. The *o* was drawn in a more circular

shape here than in the other styles to show the relaxed arm's natural proficiency for drawing a more circular shape. The counters of the *P* and the *R*, and the lower bowl of the *B* were all made larger to give a larger radius to the curved strokes that created them. The *a* and *g* were both drawn as single storey versions of the letter. This had the effect of allowing for a larger, more rounded bowl.¹⁸

Timbre-Focused Typeface: "Trembling"

"Trembling"

My thoughts on the appearance of a trembling voice in typography have been explained above, but to quickly reiterate: trembling vocal cords are analogous to a trembling hand and arm. This results in an uneven line when a letter is drawn by hand.

¹⁸ It is also the case that these single storey letterforms have a connotative connection to childhood letting guides, harking back to the relaxed innocence of childhood.

Chapter Five: User Testing

Focus Groups Research Instruments

All Focus groups consisted of three parts: a written questionnaire, a short captioned video, and a group discussion. An American Sign Language interpreter was hired through Ontario Interpreting Services for the first two focus groups and written notes were kept on a digital projector to help the participants understand and follow the comments made by myself and other participants, regardless of their knowledge of ASL or their hearing ability. There were 13 participants in total, six male and seven female, ranging in age from 21 to 75. Four participants were deaf or deaf cultured, five were either hard of hearing or severely hard of hearing, three were hearing, and one participant expressed that she was deafened.¹⁹

Video

The video that participants watched was a set of four short clips from the film *The Straight Story* directed by David Lynch captioned prosodically (see Appendix A, Example 10 for an still from this study video). This film was chosen because it is a G-rated, live action film. In the first clip, an older man named Bud is looking for his friend Alvin, whom he finds lying on a kitchen floor—Alvin is unhurt. A neighbour and Alvin's daughter Rose are also present. In the second scene Rose has a conversation at the checkout of a grocery store with a grocery store employee. In the third scene, a woman is upset because she has hit a deer with her car. In the fourth clip, Alvin discusses the price of repairing a lawnmower with two mechanics and another man. The third scene was added after the first focus group.

¹⁹ Please note that the discussion does not attempt to meet the rigors of statistical significance because the focus groups only included a small number of participants (either 4 or 5 people) and the changes made to the typefaces between focus groups make statistical analysis that overarches multiple focus groups irrelevant.

Caption styles

As is suggested in CRTC 2009-430 and carried through CRTC 2011-488, and CRTC 2011-741, pop-up style captions are preferred for pre-recorded programs. Pop-up style captions appear on the screen as complete blocks of one to three lines of text. Words in roll-up captions, in contrast, appear as single characters. When a whole line is complete, it moves up and a new line is entered. This continues, but once three lines are on-screen, entering a fourth line will remove the oldest line (at the top of the block of captions). Roll-up captions are only used because they can be entered live, as a program airs. Therefore, they are still considered acceptable for live broadcasts or productions with an extremely short turn-around time between filming and airing (CRTC 2011-488 39). For this study pop-up style captions were suitable and therefore that style was used.

Questionnaire

Before watching the video, participants filled in the first page of the questionnaire (see Appendix C for a copy of the questionnaire). This page asks the participant about: basic demographic information (age, gender, level of hearing); the number of hours that they spend watching captioned and uncaptioned video each week; their level of satisfaction with the way videos are usually captioned; what, if any, elements they feel are not adequately conveyed in captioning; and any other comments they have regarding captioning.

After watching the captioned video, participants filled in the remaining pages. The second page of the questionnaire asks about the general readability and experience of watching the video with prosodic captioning. Pages three through five ask about the different typefaces out of context of the video. Pages six and seven ask about the typefaces within the context of the video

and the final question is an open space for participants to give any other feedback they would like to about the typefaces and captioning.

Group Discussion

The group discussions began by looking at the typefaces independently of each other. Participants were asked to describe what the typeface in question made them think of. Because this was a discussion, participants were able to build on each other's ideas. The second section of the discussion was a comparison of different typefaces. Here, the same word was shown on screen in three different typefaces and participants were asked about which of the typefaces best fit the term.

Focus Group One

The first focus group was held on June 13, 2012 at 7:00 p.m. at Ryerson University. There were four participants—two male and two female—and the participants were 28, 29, 30, and 69 years old. One participant was as deaf cultured, one hearing, one hard of hearing, and the final participant expressed that she had lost her hearing ten years ago but now has a cochlear implant.

Key Findings in Focus Group One

There was a good degree of agreement between my design ideas and the interpretations of the “Shouting” typeface. Participants noted during discussion that the text sample looked “loud,” “harsh,” “broken,” and “hard.” On the questionnaires, three of the four participants thought that “loud” was a good descriptor for the design of the typeface—no other typeface had more than a single participant mark it as associated with the term “loud.” One participant also mentioned,

without being given any indication of my attempt to iconically convey an imperfect flow of ink, that the typeface had the appearance of “old-fashioned ink fading.”

In some ways, participants were able to see my intended meaning with the “Relaxed” typeface, but overall the findings from the first focus group were not extremely conclusive. Participants ascribed descriptive terms ranging from “raspy” and “stressful” to “rolling” and “gentle.” No participants saw “relaxed” as a descriptive term for the typeface, but two participants did rate it as looking more relaxed than either the “Trembling” or “Quick” typeface. Two participants thought that the “Quick” typeface was more relaxed. When presented with a set of descriptive phrases for the typeface, two participants thought that the typeface was intended to convey that the speaker is singing and two participants thought that the typeface shows the pitch of the speaker’s voice changing gradually.

Two of four participants agreed that the term “tense” was fitting for the “Tense” typeface and three of four participants agreed “tense” was a better term for this typeface than either “Trembling” or the regular version of *Open Sans*. In the group discussion, it was described as “cartoonish,” looking like it is “vibrating,” and “too charged.”

The typeface that I designed to show a quickened tempo of speech was the least well received. Two participants commented outright that they did not like the typeface and felt that it was unwelcoming and visually unfriendly. It was also noted by a participant that the spacing appeared too close. This was felt to contribute to problems with readability. Two participants regarded it as “quick” and two as “gentle.” I believe that the visual unfriendliness may have contributed to the fact that participants seemed to have little else to say about the typeface.

On the other hand, “Slow” was viewed strongly in terms of movement, which I take as a valuable connection. This connection though, did not materialize in the way I had hoped. My use

of a left-leaning slant in the “Slow” typeface to contrast the right-leaning slant found in the “Quick” typeface was not very successful. Three of the four participants chose the “Quick” typeface over the “Slow” typeface as a match for the term “fast” when the two typefaces were presented side-by-side, but when “Slow” was presented alone it was described as being “quick,” denoting speed, and as useful to advertise car races. This point about car racing is especially important. It shows that the leftward oblique angle for the “Slow” typeface, while perhaps useful in direct contrast the “Quick” typeface, could actually be read as a way of conveying fast movement as Jones contends. It appears to be problematic in this study that a person must be familiar with one typeface to understand another.

The “Tremble” typeface was described often as “spooky,” “chilling,” and “ghost-like.” While these are not directly prosodic terms, a single suitably frightful experience can make it patently clear to a person that there is a strong connection between being scared and a trembling body. Additionally, some strong conventions exist in mainstream western society around the sounds a ghost makes, whether it is the vibrato sound of a Theremin or the shaking voice that is often heard in horror movies. Three of four participants also saw it as “raspy” and “breathy.” I did not intend for these terms to be associated with the typeface, but they perhaps point to the symbolic meaning carried by other typography that shares a resemblance to “Trembling.” Packing for Halloween products and the way the word “Boo!” is often drawn in other visual media like comic books seem to have guided participants into reading these symbolic meanings in the typeface.

Changes Made to Typefaces After Focus Group One

A participant noted that the “Tense” typeface had the appearance of “vibrating.” I attribute this to my design choice to move some elements of a letter that would usually stop at a

particular metric (such as the baseline or x-height) off of that metric. This created the appearance of vertical movement across the letters as they were read, whereas I was actually aiming at a sense of jaggedness across the whole word (See Appendix A, Example 11). Reading normally involves letter-by-letter recognition, the gross shape of words, and context within the sentence (Pelli and Tillman 1). Moving the letters off standard metrics changes the gross shape of words and therefore interferes with a person's ability to employ word shape recognition in reading (see Appendix A, Example 12). This idea is congruent with a method that psycholinguistic researchers use to control for a subject's ability to recognize word shapes while reading. Changing the word shape by varying the size (Perea and Rosa 788) or the case of letters within a word, sUcH aS tHiS, is used to force a reading study subject to rely on only letter-by-letter recognition and sentence context (Pelli and Tillman 3). Giving an impression of jaggedness, as I had intended to show in "Tense," may require better gross word shape recognition, whereas vibration fits more with letter-by-letter recognition. This is because jaggedness is a spatial attribute of a whole (the broken glass is jagged), but vibration is movement in space over time (the cellphone is vibrating). Gross word shape recognition is only spatial, but letter-by-letter reading is spatial and temporal (first I read *c*, then I read *a*, finally I read *t* to read cat). Therefore, when gross word shape recognition was disrupted, leaving a reader with only letter-by-letter and sentence context—which is also spatial and temporal—a reader may be more primed to see an iconic signification that is both spatial and temporal (vibration) than to see a resemblance to something that is only spatial (jaggedness). Vibration fits better with its synonym trembling than it does with the idea of tension. So, in an effort to strengthen the tenseness of this typeface and highlight the vibrating aspect of the "Trembling," I removed this attribute from "Tense" (see Appendix A, Example 13) and added it to "Trembling" (see Appendix A, Example 14).

Certain aspects of the “Slow” typeface’s design also raised interesting ideas that would be better employed in another typeface. The *o*, *e*, and *g* all stood out as more visually salient than the other letters. This design element was intended to show a degree of inkiness or viscosity—a thickness to that material that the letters appear to be substantiated in—but the actual result was that participants saw a halting or laboured style of speech. They said it was as though the speech was being held back. Participants used the term “stop-and-go” more than once to express this idea. It was also described as “intermittent,” “undecided,” “deliberative,” and “meditative.” This stuttering and broken feeling is better suited to the “Tense” typeface, where speech is more forced. So, I removed the unevenness from “Slow” (see Appendix A, Example 15) and increased the weight of a few letters (*e*, *h*, *n*, *o*, *r*, and *u*) in “Tense” (see Appendix A, Example 13).

The “Quick” typeface was updated after the first focus group in a number of ways. The angle of the slant of the letters seemed to be too extreme, so a version with a more slight oblique was created. The letter spacing was also increased as participants had suggested. The letters that connect to give a cursive-like feeling benefitted especially from the increased letter spacing, because this made each component letter more recognizable. Overall, reducing the slant and increasing the spacing made the typeface much easier to read. The x-height of the letters was reduced as well for the second version of the typeface. This had two purposes. First, the x-height became noticeably lower on the “Quick” typeface than any of the other styles, which helps distinguish it and make it more easily recognized by the viewer. This was a needed addition once the slant was so greatly diminished. Second, the lower x-height gives the lowercase letter the appearance of a more equal ratio of width to height. This could help avoid a feeling of visual unfriendliness because it makes words appear less as imposing blocks of colour and more as

distinguishable sets of letters. In addition to reducing the x-height, many of the pointed terminals were removed or replaced with thicker, less thorny terminals (see Appendix A, Example 16).

The “Relaxed” typeface remained relatively unchanged aside from a few terminals being increased in size to make them more noticeable. The “Shouting” typeface received no design changes because it was so well recognized by the participants. However, in the test video for the second and third focus groups, an additional clip was added to showcase it further in the hope that more useful ideas will be gained or that the typeface’s value will be even more reinforced.

Problems with Implementation

While the prosody of all speech is important at least on some level, my first focus group showed that there could be a conflict between marking prosody and the viewer’s ease with identifying the correct speaker in a scene. In some cases, speaker identification could be improved by marking all lines of text prosodically—for instance, when a conversation is between a person who always shouts and another person who always whispers—however, actual conversations are rarely that simple. If, in this example, the shouter abruptly began to whisper and the whisperer began to shout, there is the potential for the roles to be confused. In the video that participants watched, the vast majority of the captioning was presented in a prosodic typeface and only a small number of lines appeared in the unaltered (regular) version. A hearing participant in the first focus group responded that this made speaker identification more difficult to track. Another participant, who was born hearing, but lost the ability to hear and now has a cochlear implant noted that “for a deaf person who hasn’t heard sound, [using different typefaces] might be helpful to express the intensity of the situation, but I would think changing [the typeface] too much wouldn’t be helpful.” This issue could potentially have less of an effect on the experience of viewing captioned videos as people become accustomed to changing

typefaces, but the methodology required to study that is beyond my scope here. As a result of these observations, then I showed the video for the second and third focus groups I made a far greater use of the regular typeface and only highlighted the more extreme prosodic features typographically (see Appendix A, Example 25 for still images representative of this change).

Focus Group Two

The second focus group was held at Ryerson University on June 28, 2012 at 7:00 p.m. and it consisted on two male participants and two female participants, two hard of hearing participants and two deaf participants. The age of participants was 31, 42, 47, and 59. Not all sections of the questionnaire were completed during the second focus group because two of the four participants needed more interpretation of written English than I encountered in the first focus group or expected. This slowed the general pace of the study, so questions 17 through 20 on the questionnaire and the portion of the group discussion where typefaces are looked at in contrast to each other needed to be passed over. This is unfortunate, but the most crucial aspects of the focus group were still completed.

Key Findings in Focus Group Two

There was a noticeable difference between the reaction of participants in the second focus group and the first focus group with regards to the “Quick” typeface. It was also the typeface I revised the most. Whereas participants disliked the first iteration of the design and found it to be visually unfriendly, the updated version that I presented at the second focus group was far better received. The only negative attribute expressed about it during the discussion was by a participant with a visual impairment who said that the slant was visually distracting, but this came with the caveat that the slant also made the typeface seem more emphatic. I think that the emphatic aspect coincides with the italic-like quality participants noticed. Italics are commonly

employed to give emphasis. In HTML code, the tag (emphasis) forces a typeface into its italic version. The typeface was also described fairly well in terms related to writing and the body. Participants said that it made them think of handwriting and “wannabe calligraphy,” a phrase that I feel agrees with my idea that the typeface looks script-like without giving strong overtones of elegance or class.

Missing from participants’ responses during focus group 2 was any indication they saw a sense of speed in “Quick.” None of the four participants²⁰ responded on the questionnaire that they would associate the typeface with the term “quick”—or “slow” for that matter. It seems that I may have traded too much speediness for readability in my alterations after the first focus group.

The result of removing the slant form “Slow” seemed to be that it became too bland. Participants regarded it as “normal,” and “boring.” On the questionnaire, two participants did still indicate that they regarded it as fitting with the term “quick.” I take this as an indication that at some level there is still a sense of speed inherent in the designs, even if it is the wrong amount of speed and not seen by all participants. This typeface was considered especially visually clear, possibly because of the large letter spacing and the relatively large width of letters.

A hard of hearing participant was very quick to suggest that “Whispering” conveys a sense of whispering. Other terms that participants used to describe it were “light,” “faint,” “not bold,” “dream-like,” and “skeletal.” All told, I took this as a strong indication that I was on the correct path towards showing a whispering style of speech. Even with showing more shouted dialogue there was no similarly impactful findings for “Shouting.”

²⁰ One of these four participants only responded to a single question (the first) about specific typefaces on the questionnaire

Participants described “Trembling” as “weak” and showing a “gravelly voice, like an old man.” These are suitable terms for the index of shaking or vibrating as fatigue and physical stress both cause muscles to shake when (over)used and age is correlated with increased shakiness from essential tremors, and Parkinson’s disease (Mayo Clinic Staff). Participants also stated that they thought the typeface was suitable for horror movies or Halloween, which has similar connotations to the ghostliness described in Focus Group 1. Overall, I felt comfortable that my designs at this point were headed in the right direction.

“Tense” elicited a good amount of feedback from participants. It was described as appearing “angry,” “fussy,” and “like speaking through clenched teeth.” These ideas are all relatable to the experience of being physically tense—especially the “clenching” thought. This typeface was also seen as being “distorted” because of the uneven weight of the strokes used in the different letters. There were no significant opinions regarding “Relaxed.”

Changes Made to Typefaces After Focus Group Two

The degree to which stroke weight varies in “Tense” was made more even after focus group two. Whereas participants saw a stopping and starting cadence in this design feature when it was part of “Slow,” the implementation of it into “Tense” was perhaps a bit too heavy-handed. I therefore reduced to contrast between stroke weights for the third iteration of the typeface (see Appendix A, Example 17).

The “Whispering” typeface featured small missing portions of letters to stress the feeling of airy breathiness and not-wholly-there-ness. This aspect may have contributed to issues around the readability of the typefaces, as they were perhaps a bit too faint as a result. To help avoid this, all missing sections were replaced. Also, the hatchings used to create the letters were

increased in size (see Appendix A, Example 18). This had the effect of making the letters appear darker, and therefore more easily readable when sized for captioning.

As further attempt at distinguishing “Relaxed” and highlighting the roundness that both makes it unique and iconically resembles the indexical movement of a relaxed arm, the bowls of the *a*, *b*, *d*, *p*, and *q* were made even wider and rounder. The hope was that this would help spur more conversation about this typeface (see Appendix A, Example 19).

Six degrees more of a slant to the right was added to “Quick” in an effort to bring forward the feeling of speed that may have been lost after the first focus group. This put the slant of the typeface at roughly 12 degrees to the right (see Appendix A, Example 20). More connections between letters were added as well because this feature was not showing up frequently enough in test texts. In this iteration of the typeface, *as*, *ae*, *ce*, *cs*, *es*, *ge*, *gs*, *hn*, *ms*, *ns*, *te*, *ue*, *us*, and *ts* were all pairs of letters that appeared to be written with a connecting stroke.

In an attempt to make “Slow” look less “normal” and “boring,” I took an approach that used obliques again. This time, however, I slanted the left and right sides of the characters in opposing ways. So, the left sides of the characters were given a right leaning slant and the right sides were slanted to the left. Thinner letters, specifically *i*, *j*, *t*, *I*, and *J* were only slanted on either the right or the left because slanting both sides made these letters too pointed (see Appendix A, Example 21). My hope was that this would give a sense of speed, as slanting has been shown to do, but without also giving any obvious directionality to the letters that might imply quickness over slowness.

Some participants felt that the letters of the “Trembling” were not aligned to the baseline well enough and they saw this as making the typeface more difficult to read. This agrees with research noted above (Pelli and Tillman 3) because my designs may have disrupted the ability of

readers to use this gross word shape, therefore taking away one of the important tools needed to quickly comprehend words. Gross word shape, letter-by-letter decoding, and sentence context are all required for a person to read at their best (Pelli and Tillman 3). Because participant's expressed concerns with the unevenness of letters' vertical placement off of the baseline in "Tremble," this aspect was made less severe. Letters still had a degree of vertical movement and gross word shape recognition may have been slightly disrupted, but the level of vertical displacement was dialled down to be less distracting.

Focus Group Three

The third and final focus group was held on July 10, 2012 at 7:00 p.m. at Ryerson University. There were two male participants and three female participants. The participants were 21, 22, 57, 58 and 75 years old. Two participants were hearing; one was hard of hearing; one was severely hard of hearing; and a one was deaf, but used a cochlear implant.

Key Findings in Focus Group Three

"Quick" was described by four of the participants as "relaxed." I believe that this could have something to do with the almost cursive feeling I was aiming for. Four participants associated the quick typeface with a sense of speed. Two participants marked the typeface as "quick" and two marked it as "slow." While this is not ideal, it backs up the idea that slanting can be useful for indicating the prosodic element of tempo because the six degree increase to the slant was the most obvious change made between focus groups two and three. My attempt to reintroduce a sense of speed to the slow typeface appears to have been mildly successful. Two participants did note that they saw "slow" as a term associated with the typeface—an improvement over the two that saw it as "quick" in the second focus group—but one participant

explicitly ruled this impression by stating that they saw the typeface as “deliberate, but not slow.”

The design of “Slow” in the third focus group may have posed problem with identifying the intended meaning of “Shouting.” When the two typefaces were presented side-by-side (along with “Whispering”) three participants saw “Shouting” and two saw “Slow” as the best fit for the term “shouting.” Participants also described “Shouting” as “angry,” “irritated,” and “irritating.” Four of five participants indicated on the questionnaire that “Whispering” appeared both “quiet” and “gentle.” No other terms were as frequently associated with this typeface as these two were.

There was a great deal of agreement between participants around words associated with “Tense.” All five participants agreed the that term “harsh” was fitting of the typeface, three thought it looked “intense,” and three marked it as “loud.” Participants also saw the typeface as “stern,” “didactic,” and “authoritative.” Van Leeuwen points out that for both men and women, speaking in unnatural pitch registers plays a role in being assertive and dominant. He states that “men use higher regions of their pitch range to assert themselves and dominate” and that “women, on the other hand use the lower end of their pitch range to be assertive” (“Voice” 427). These connections between assertiveness, unnatural pitch, and tension are a positive indication my work showing an unnatural or stressed feeling to the pitch of a speaker is coming through in the typeface. The terms “brusque” and “brisk” were also used during the discussion to describe it. This points towards a sharp abruptness to the speech that also agrees with my design plan.

My typeface intended to show a relaxed way of speaking was still fairly unsuccessful at appearing relaxed. It was however associated with the terms “happy,” “fun,” “innuendo,” “sly,” and “sarcastic.” I believe that these results could be hinting towards a better use for this typeface.

It could be useful for showing a joking or laughing tone. “Trembling” continued its trend of being associated with shakiness, scariness, and old age.

Changes Made to Typefaces After Focus Group Three

The standalone versions of *c*, *e*, *g*, and *j* (those not attached to another letter in a ligature) were adjusted so that they would look better in this form. Some of the sharper terminals that were removed after focus group one—to give a more visually friendly appearance—were reintroduced, but in a much less stark form. Again, this is intended to show the act of quickly lifting the writing instrument from the page at the end of a stroke. The width of all letters for “Slow” was slightly increased and letter spacing was adjusted to compensate (see Appendix A, Example 22). Word spacing was increased to give the appearance of more time between words.

For the final version of “Shouting,” the exclamation marks were made larger (see Appendix A, Example 23). The idea here is that making these punctuation marks, which already indicate a number of the prosodic features at play in shouting, more visually powerful may highlight the meaning they already carry. That is, if increasing the size of any character gives the impression of loudness, then increasing the size of exclamation marks even further may make these characters especially prosodic.

“Whispering” was made slightly darker by thickening the hatch marks the letters are drawn with. Removing and shortening hatch marks in some letters, which was also done in the first version of this typeface, then offset this extra darkness. This was an attempt to show breathiness more, which was not something participants attributed to my designs in the last focus group.

At this point, I created ligatures of common English language letter combinations where the tremblingness may have been problematic because there was too stark a difference between

the vertical alignments of the letters. I did this for the combinations *an*, *ae*, *er*, *in*, *nd*, *re*, *st*, and *the*. For instance, when the letters *e* and *r* are typed together without the ligature, the discrepancy between the vertical placements of them is too great; *e* is especially high above the baseline and *r* is especially low. By creating a ligature I was able to move the characters so that when they appear in this combination the contrast is less stark. Repeated letters were also problematic in this typeface. Placing a letter beside its duplicate created the appearance that there was no vibration because both characters are the same distance off of the baseline. I adjusted the height of one instance of a repeated letter in ligatures for *ee*, *ff*, *ll*, *oo*, and *tt* to overcome this issue in these common cases (see Appendix A, Example 24).

For the “Tense” typeface, the letters with thicker strokes that give the start-and-stop feeling were increased, but still not to the level used prior to focus group two. No significant changes were made to “Relaxed” after focus group three.

Chapter Six: Conclusions

Outcomes of Testing

The Potential of Prosodic Captioning

I consider my work presented here to be a successful first attempt at exploring a groundwork for prosodic captioning rooted in bodily experience. Overall, the iterative process of holding three focus groups and making updates to the typefaces based on the responses after each session was effective and there was a clear trend towards better typefaces as the study progressed. The benefits of these refinements can be seen in the improved scores participants gave to questions about the overall readability and impact of the typefaces. On a five-point Likert scale²¹ marked as very disagreeable (one), neutral (three), and very enjoyable (five), most participants in the first two focus groups rated the experience of watching a video captioned with the prosodic typefaces (question nine on the questionnaire) between the range of very disagreeable and neutral. In the last focus group however, this trend had shifted and most participants rated the experience in the range between neutral and very enjoyable. Another important change was that in focus group three, all of the participants responded that the use of the fonts did not make understanding the scenes more difficult at all (question ten on the questionnaire). This contrasts with the first two focus groups, where the range of responses on the same question was between the fonts making the scene “somewhat” (three) more difficult to understand and “very much” (five) more difficult to understand. These results suggest that by working in an iterative process of development and testing, I was able to move towards typographic treatments of prosody that were more agreeable to the viewer and improved the

²¹ As mentioned earlier, statistical significance was not tested for these values because the small number of participants in each focus group made this type of analysis inaccurate.

viewer's ability to make sense of the scenes.²² For a new way of captioning to be successful it must, as a baseline, be sufficiently enjoyable and not impede a viewer's understanding of a scene, because no matter how meaningful an approach it might be, if it is not enjoyable and makes understanding the video more difficult, it will probably not be preferred or even used. A positive reading experience must be present first and foremost in captioning. My iterative design and testing process made reaching that baseline possible. Without such a procedure, my work may not have been an acceptably enjoyable or difficulty free experience. With that said, it is important to note that these typefaces might still not be suited to everyone—two points raised by participants during the second focus group made this clear.

One of the participants in the second focus group had concerns with the typefaces because she also has diabetes related vision loss. For her, some typefaces were preferable because they were generally easier for her to read, such as “Shouting” which she described as very clear, large, and bold, but “Relaxed” she simply said that she detested. Another participant expressed that the use of different typefaces was not desired. He explained that he prefers the way captioning is currently done and that he does not want the typeface to change. This participant also noted that when watching videos he uses both captions and hearing aids, which allows him to read the dialogue while still being able to pick up on how that dialogue is delivered prosodically. This participant's preference, along with the problems the typefaces might cause for some viewers with visual impairments, speaks to a need for more diversity in how video is captioned. Prosodic captioning cannot simply replace all captioning, but it can be another option

²² Of course, these two results could be linked. Typography that makes a scene difficult to understand will probably be less liked. At the same time, if a typeface is sufficiently disliked, that could be distracting to the point that understanding a scene is hindered.

available for those that want it. Including prosodic cues typographically does not necessitate removing the option for other people to have captions presented in a single, blank typeface.

The overall perception of the typefaces was also affected by the fact that salience requires context. A particular way of speaking can only be significant if it is recognizable as distinct from a person's usual way of speaking. For a shout to be loud or a whisper to be quiet, they must be contrasted against the volume of regular speech. After the first focus group, I used less of my prosodic typefaces in the captioned study video. I believe that this may have had two effects. The first is that marking fewer lines of dialogue prosodically makes the captions easier to follow. A question on the ease of readability (question seven on the questionnaire) returned much better results after the first focus group. The second effect this has is providing a stronger context for the prosodic typefaces. A typeface can only look especially thick, thin, wide, oblique, curved, or angular in the context of other typefaces. Using the regular weight of *Open Sans* more frequently than I had initially thought was needed certainly provides this contrast and allows viewers to be more attentive to the use of prosodic typefaces.

Focusing on prosody rather than emotion was pivotal to this project's relevance and success. Past work that erroneously removed from the viewer's power the act of interpreting the emotional meaning of an actor's speech cannot be as relevant to pushing for equal access as my approach based in prosody can. This is because prosody is closer to what a hearing viewer experiences. We do not hear emotion, we hear speech prosody and then interpret emotion from that.

Prosody is also more easily quantifiable in distinct categories that when mixed together in specific ways will produce a specific prosodic styles. The fact that prosody is simpler than emotion on this level means that it is a better place to begin building typography from. Emotion

is so tied up in the psychological realm that happiness, for example, can manifest itself physiologically in a number of different ways. How does one typographically reconcile tears of joy with other forms of happiness that do not share physiological effects with sadness? Prosody, on the other hand, is produced by the body in distinct and measurable ways, regardless of its meaning. That is not to say that all elements of prosody are as straightforward to show as others. In this study, I found that some aspects of prosody lend themselves more easily to typographic expression than others.

The Success of Showing Different Prosodic Elements

The loudness of “Shouting” and the quietness of “Whispering” were perhaps the two most easily discernable prosodic qualities of all the typefaces. The relation between these prosodic styles and my visual design choices that give these typefaces their meaning is based on the materiality of the communicative acts. It seems there is a clear iconic connection between the material thickness of stroke in a letter and the material loudness of a person’s speech.

Expressing tension was somewhat successful in this study. While the specific term “tension” was not always employed by participants to describe “Tense,” there was still a great deal of description given by participants that fit with both the prosodic element I was trying to convey and ideas related to tension (such as harshness or assertiveness).

“Trembling” was quite well regarded as conveying the shakiness I tried to show. A lot of this could have to do with the very emotionally charged (salient) experiences when our bodies shake. I feel that the most notable experience where our bodies tremble—and this was strongly reflected in the responses of focus group participants—are those when we are scared. While I was explicitly trying to show prosody not emotion in my work, I could not disregard these emotional responses because there is such a strong connection between the physical sensation of

a fearful experience and the prosody that is produced from fear. I do not doubt that nervousness, which I also see associated with a shaky body, could just as easily be read into this typeface, given the proper video image and dialogue context.

A relaxed and casual style was occasionally pointed out for the “Quick” typeface. I believe that this is because it has many smooth and rounded features established in the almost cursive style I tried to create. These round and smooth elements, as I explained above, are indexical of a relaxed arm writing. I tried to highlight this aspect the most in “Relaxed,” but that typeface was not identified as relaxed quite as strongly as I would have preferred. It is possible that this style of voice could be too close to the speaking voice people usually use—when prosody is less meaningful. Marking relaxation typographically may therefore not have a great deal of value in captioning. A relaxed voice may be important if all prosody is going to be shown, but with the finding that too many changes to the typeface has the effect of confusing and distracting the viewer, it may not need to be included in a set of typefaces for captioning prosody at the moment. That is not to say that my designs for “Relaxed” were a waste. I believe that they could still be useful—with a bit more design work and study—for a related way of speaking, such as with a laughing or lilting tone.

Giving the impression of quick and slow speech was perhaps the most complex aspect of prosody I worked to convey in letterforms. A great deal of being able to show this aspect of prosody in type came from slanting the letters. The success of this method is important because it gives further credence to other research discussed above that pins a connection between oblique letter shapes and the impression of speed. However, even with the strongest associations between slanting and speed, expressing the rate a person speaks at is always indisputably tied up in the rate that a viewer reads at because some people read slower than others talk. There is therefore

an incongruity whenever a fast talker meets a slow reader. Timing in captioning is a tricky factor to deal with. In fact, the timing of contemporary captioning was a common concern for nine of the 13 participants in this study. But, what makes the relationship between quick speech and reading most complex is that when a typeface hopes to show quick speech there is an increased importance on ensuring that words are easily readable. This is further complicated when viewers takes the legibility to mean clarity, which is an entirely understandable association to draw. This complication points squarely in the direction that my work can be expanded.

Considerations for Future Work

Peirce's third type of sign, the symbol, has a great deal of potential for expressing prosody. I see value in augmenting my work rooted in bodily experience, with a deeper understanding of the social and cultural contexts of type and signs. There are many symbols already widely in use that could help lend an understanding of loudness, rhythm, tempo, pitch, or timbre to typography. For instance, there is an abundance of symbolic meaning found in roadside speed limit signs that could be employed to indicate how quickly a person is talking without compromising on readability. My work is part of what I see as a groundwork (but certainly not the only groundwork) for captioning prosody that uses handwriting and the experience of the body to show prosody in as universal and as inherent a way as possible. The approach I have started with aims to be less culturally dependant, but starting from the body does not prohibit symbolic meaning. Just as social context connects the Blackletter typeface Gutenberg developed to the religious authority of the time, so too can a typeface employ a symbolic form of meaning to convey prosody. For instance, an institutional typeface, such as one resembling *Helvetica*, could perhaps show a firm and authoritative voice without requiring a connection to feeling firm

and authoritative. As advantageous as it might be to swap the iconic and indexical meaning I suggest for symbolic meaning, this is not the best approach.

I see a potentially successful route to better prosodic captioning through augmenting the iconic and indexical meaning of a typeface with a symbolic form of meaning would magnify the successes of this work. If a typeface both looks loud through the mechanism I have explored here *and* symbolically means loud to a viewer, then the viewer will be that much better equipped to understand the captioning. Longer-term exposure than was possible in my study, where viewers watch at least a few hours of video captioned with prosodic typefaces, would undoubtedly help make people become more comfortable with the approach I take in this project. Once viewers have moved past the novelty of the typefaces, they would be able to bring their past experiences of when a particular typeface has been used, to bear on how a typeface is presently being used. If enough shouting people are presented along with a ‘shouting’ typeface, then when an off-screen narrator’s voice is presented in that same typeface, it will be read symbolically as a shout regardless of whether or not the viewer understands the indexical/iconic connection between the typeface and a shout. This learning process could happen with any typeface, but it would be made especially successful by starting from my standpoint because a system based on embodied experience has a built-in way of understanding the meaning of a typeface. I believe that even if symbolic meaning is used to effectively show prosody in typography, if that symbolic meaning coincides with an underlying framework of indexical and iconic meaning that is as understandable as embodied experience is, then the whole system will be that much easier to develop, learn, and extend.

With the doors open to symbolic meaning, there is good reason to re-evaluate the use of diacritics and icons in future study. Although I have said that learning to read for the first time is

a different experience than learning to read in a new way, I do not think that automatically rules out the use of signs beyond the letters. The human mind is a powerful thing. We already have a number of prosodic marks in English: commas, periods, exclamation marks, and question marks are just a few of these. Other languages use accent marks to give meaning to words without any extra difficulties forced on the reader. It is therefore reasonable to think that captioning users can learn to associate diacritics with specific meaning, especially if those new marks ride on meanings already present in the human experience. In future study it would be possible to develop diacritics and icons from the standpoint that bodily experience, especially handwriting, is a useful way to form a connection between the written and spoken word.

I wish to present one final possibility of how the iconic and indexical meanings I have explored in this thesis can be extended to an approach based on diacritics. It involves a line that sits above the captioned text, and employs different curves, angles, thicknesses, and hatching to convey different types of prosody. To show how such a system might function, I have designed five styles of lines based on five²³ of the typefaces created for this study (see Appendix A, Example 26) and applied those to four sample lines of text (see Appendix A, Example 27).

This approach offers possible advantages over the typefaces that incorporate the prosodic information. Firstly, the legibility of the typeface will remain consistent, regardless of the prosody. This will be a benefit for captioning viewers who might have trouble reading more complex letterforms, such as viewers with visual impairments or lower English literacy skills. Since prosody is no longer marked directly within the text (and is perhaps easier to ignore), this approach might also be preferred by individuals that need less prosody marked, such as the viewers who use captions to clarify the words spoken, but still have access to prosody through

²³ The “Quick” and “Slow” prosodic styles were not included because I could not find a sufficient way to show them here.

their own hearing. I think that another advantage of this system is that it can piggyback on pre-existing prosodic cues in text to reinforce the connection between the overline and how the words are said, such as in Appendix A, Example 27 where the curved ends of the first two lines indicate questions. One aspect of the prosodic features of a spoken question in English is that the pitch of the speaker's voice is higher at the end of the sentence. I have marked this in the line, even though it is technically not needed—the prosody is already indicated by the question mark—because it allows the question mark to act as a point of reference. That is, if an element of the text that marks prosody (such as a question mark) is associated with a co-occurring change to the line, it should be easier to understand the *whole* line as marking prosody.

The final area I wish to point out as ripe for future work falls more on the technological side of my work. It could be interesting to try algorithmically manipulating letterforms. That is, a computer could be used to automatically generate prosodic typefaces based on different shades of loudness, tension, shakiness, or breathiness. Developing typefaces in this way would allow for subtleties between the different ways that a person speaks to be reflected in the exact designs of a typeface. Whereas both Dialogue A and Dialogue B are shouted, Dialogue A could show more loudness with thicker strokes than Dialogue B, which in turn shows more roughness through a greater degree of distress applied to the strokes. A fine-grained algorithmic approach that is as nuanced as I foresee would be a less heavy-handed approach to showing prosody than I have tried. It could also easily allow for more complex cases than my work has, such as those where prosody is both especially loud and especially tense. More nuances would mean a steeper learning curve for the viewer, but this could come sometime down the line, when a simple form of prosodic captioning is common practice.

Final Summation

Psychological life is always embedded in the corporeal, embodied life that predates it evolutionarily; it is a growth to support corporeal life (Hoffermeier 169). That is, the body is all-pervasive. In “Voice” van Leeuwen sketches “a semiotics in which the voice is what it means and means what it is, and in which meaning is made with the body, and understood on the basis of bodily experience” (432). Keeping my designs rooted in this basic way of understanding speech prosody, as a bodily experience for both the speaker *and* receiver appears to have been a useful tool for framing a translation between the auditory medium of speech and the visual medium of typography. The groundwork that I have attempted to lay with my typefaces, for a system of captioning prosody, is a useful one and fertile for future explorations.

Appendix A: Examples

Example 1



Line 21 captioning. Image from Line 21 Media Services ltd.
Mustard Pancakes online video.

Example 2

Ngang is an unmarked and even mid-pitch tone.

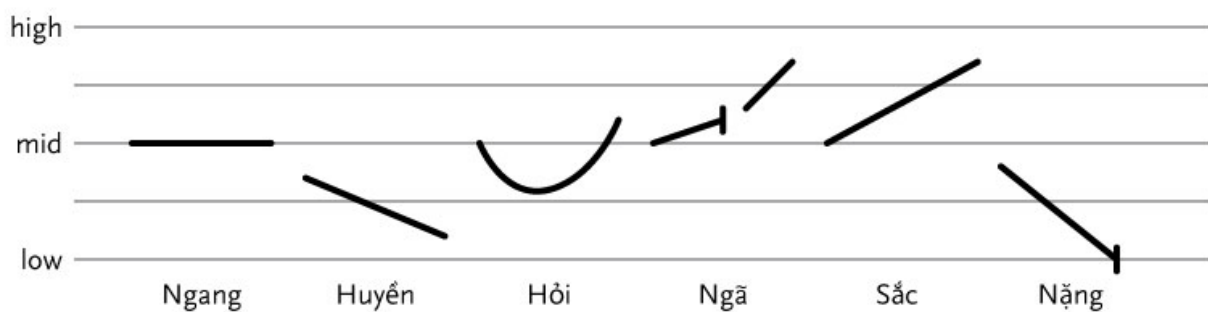
Huyền is marked with an accent grave (à). The pronunciation of a syllable featuring this tone starts with a low pitch, moves lower, and becomes increasingly breathy.

Hỏi, which is a similar inflection to the one used to ask a question in English, starts a middle-pitch, falls, and then rises. It is marked with a hook (ã) over the vowel in question.

Ngã features a glottal stop midway through. A glottal stop is the act of cutting off the flow of air during speech with the throat. In English, a glottal stop is commonly used in the expression “uh oh.” Ngã begins with a mid-range pitch, rises slightly to a glottal stop, and then rises sharply following the glottal stop. This is marked with a tilde (ã).

Sắc is marked with an acute accent. Words pronounced with this tone start at mid-pitch and rise sharply (á).

Nặng, marked by a dot below the vowel it effects (ạ), starts with a low pitch, and falls lower while becoming increasingly creaky. The sound finally ends in a glottal stop.



Pitch ranges of Quốc Ngữ tones

Example 3

Andantino con moto
_HAMLET.

dolce. *cresc.*

Dou - te de la lu - miè - re, Dou - te du soleil et du
jour, Dou - te des cieux et de la ter - re, Mais ne dou - te jamais de mon a -
mour! Ah! ne doute jamais, jamais de mon a - mour!

Theme of Hamlet's Love for Ophelia
Thomas, Ambroise, M. Carré, and J. Barbier (n.d.)

Example 4

Blackletter

“Cloister Black BT,” 72 pt.

Example 5

Humanist

“Arno Pro Regular,” 72 pt.

Example 6

Geometric Sans Serif

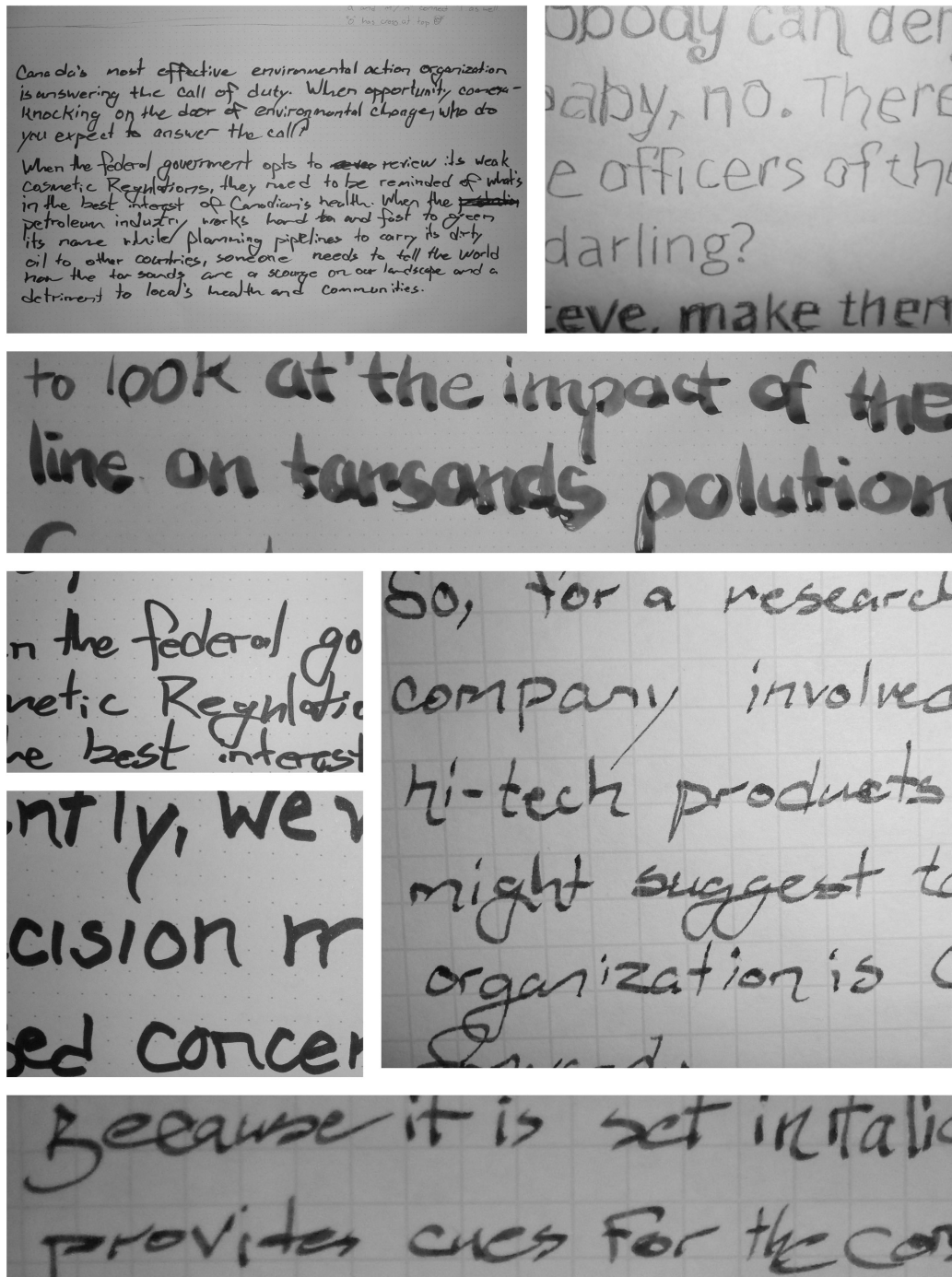
“Futura Medium,” 72 pt.

Example 7

Humanist Sans Serif

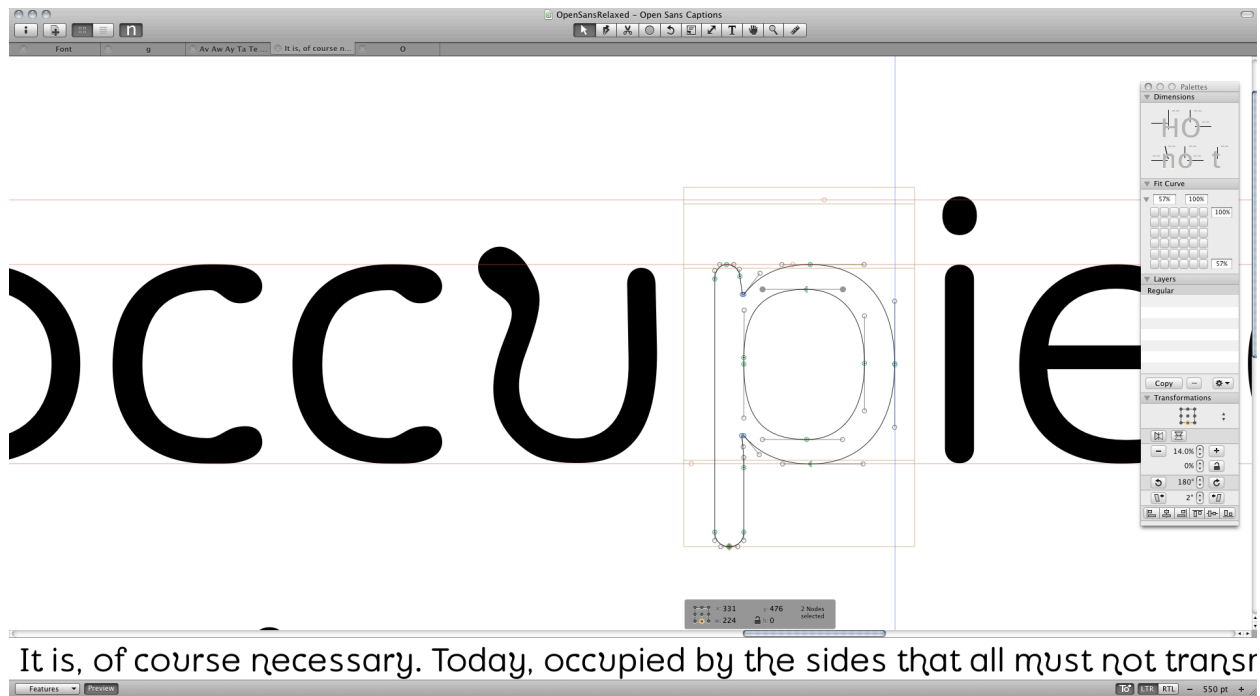
“Gill Sans Regular,” 72 pt.

Example 8



Samples of various handwritten prosodic styles.

Example 9



Glyphs, editing type with preview window sized to resemble captions

Example 10



Still Image from study video, ©Walt Disney Pictures, 1999

Example 11

Uneven Metrics

Example 12



Gross word shapes for “Open Sans Regular” (above) and “Tense” (below).
These are the gross shapes for the words “Different Shapes.”

Example 13

Before Tea houses for sale
After Tea house for sale

“Tense” before and after the changes made based on the results of focus group one.

Example 14

Before Hours are Monday to Friday
After Hours are Monday to Friday

“Trembling” before and after the changes made based on the results of focus group one.

Example 15

Before Hawker Horsley
After Hawker Horsley

“Slow” before and after the changes made based on the results of focus group one.

Example 16

Before *Early football in Thurrock*
After *Early football in Thurrock*

“Quick” before and after the changes made based on the results of focus group one.

Example 17

Before A hot pruse
After A hot pruse

“Tense” before and after the changes made based on the results of focus group two.

Example 18

Before

Madame Willow

After

Madame Willow

“Whispering” before and after the changes made based on the results of focus group two.

Example 19

Before

The bad quips

After

The bad quips

“Relaxed” before and after the changes made based on the results of focus group two.

Example 20

Before

The last name

After

The last name

“Quick” before and after the changes made based on the results of focus group two.

Example 21

Before Indexation of contracts

After Indexation of contracts

“Slow” before and after the changes made based on the results of focus group two.

Example 22

Before Thought about it, and...

After Thought about it, and...

“Slow” before and after the changes made based on the results of focus group three.

Example 23

Before ;Whoa!

After ;Whoa!

“Shouting” before and after the changes made based on the results of focus group three.

Example 24

Before Bill and the Sweeter Stuff

After Bill and the Sweeter Stuff

“Trembling” before and after the changes made based on the results of focus group three.

Example 25

Focus Group One

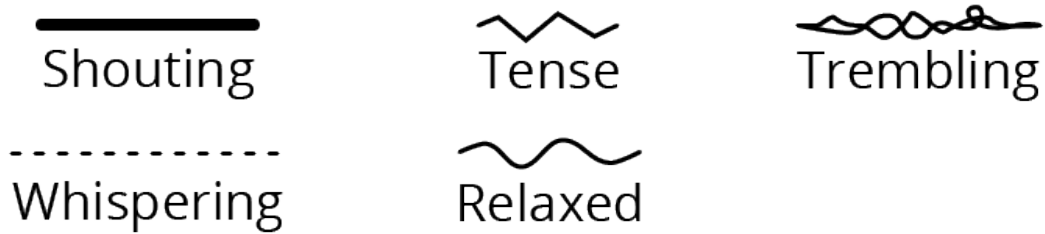


Focus Groups Two & Three



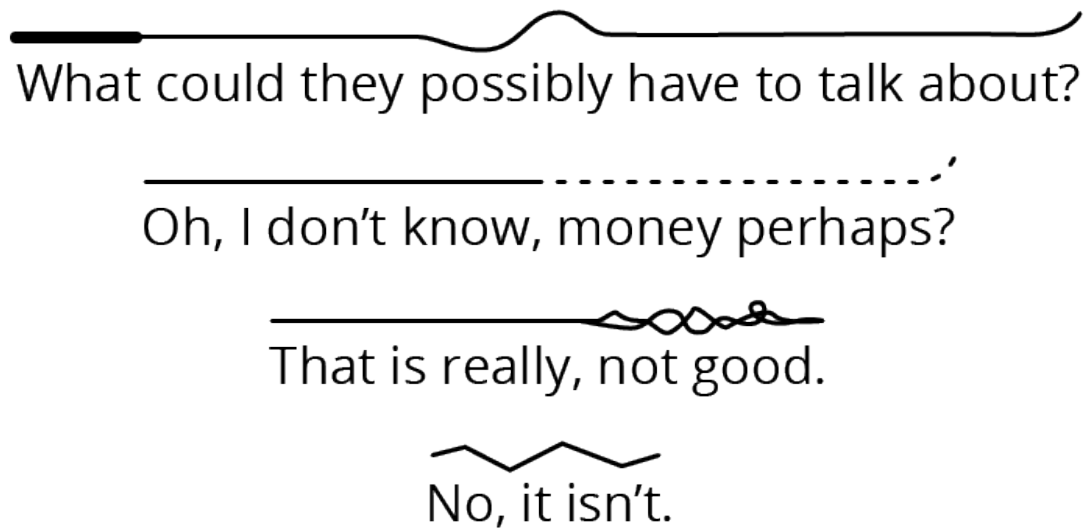
Representative frames showing a change to the amount of dialogue marked prosodically
between focus group one and the following two focus groups.
©Walt Disney Pictures, 1999

Example 26



Five potential styles of overlines for marking prosody

Example 27



Sample text marked with prosodic overlines.

Appendix B: Glossary of Typographic Terms

This is an extremely simplified and condensed list of typographic terms and really only covers the terms used in this paper. These definitions may not be suitable for non-Latin typefaces. For a more complete list of typographic terms, please see www.typographydeconstructed.com or www.adobe.com/type/topics/glossary.html.

Baseline

The line that runs below most letters. The line that letters appear to rest on.

Bowl

The curved part of a letter that encloses a counter. For example the rounded sections of *B*.

Counter

The negative space enclosed or partially enclosed in a letter. For example, the centre of *o* or the triangle inside an uppercase letter *A*.

Ligature

A single character combining two or more letters.

Metrics

General measurements used in a typeface, such as the height of the ascender line, descender line, x-height, baseline, and cap height.

Stroke

Any line used to draw a character.

Terminal

The end of any stroke without a serif.

x-Height

The horizontal metric that runs along the top of most lowercase letters, usually the height of the letter *x*.

Appendix C: Study Questionnaire



Date:
Participant Number:

What is your age: _____ Gender: _____

1. Which term do you feel best identifies your hearing? Please circle one answer.

Hearing Hard of Hearing Deaf

2. Approximately how many hours of captioned video (TV, movies, online) do you usually watch per week?
Please circle one answer.

0 1-5 6-10 10-15 15-20 20+

3. Approximately how many hours of uncaptioned video (TV, movies, online) do you usually watch per week?
Please circle one answer.

0 1-5 6-10 10-15 15-20 20+

4. What is your level of satisfaction with the way videos are usually captioned? Please mark a point on the scale provided to indicate your satisfaction.

Very Neither satisfied Very
unsatisfied nor unsatisfied satisfied
1 2 3 4 5

5. Please mark any of the categories below that you feel are not adequately conveyed in captioning:

- ☐ Music
- ☐ Sound effects
- ☐ Emotion in dialogue
- ☐ Speaker identification
- ☐ Timing
- ☐ Other: _____

6. Are there any other comments you would like to make about captioning?



Date:
Participant Number:

7. How readable did you find the captions in the video just watched?

Very difficult to read		Not especially difficult or easy to read		Very easy to read
1	2	3	4	5

8. Did the use of these fonts help your understanding of this scene?

Not at all		Somewhat		Very much
1	2	3	4	5

9. How would you describe the experience of watching a video captioned with these fonts?

Very disagreeable		Neutral		Very enjoyable
1	2	3	4	5

10. Did the use of these fonts make understanding of this scene difficult?

Not at all		Somewhat		Very much
1	2	3	4	5



Date:
Participant Number:

11. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

Num id ridets bon-
rum eloquentian
probo safenys

If there are any other words you associate with this typeface, please list them:

12. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

Num id ridets bonrum
eloquentian probo
safenys

If there are any other words you associate with this typeface, please list them:



Date:
Participant Number:

13. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

Num id ridets bonrum
eloquentian proba
safenyys

If there are any other words you associate with this typeface, please list them:

14. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

Num id ridets bonrum
eloquentian proba
safenyys

If there are any other words you associate with this typeface, please list them:



Date:
Participant Number:

15. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

Num id ridets bonrum
eloquentian probo
safenys

If there are any other words you associate with this typeface, please list them:

16. Please circle any words you would associate with the typeface on the right:

Quiet	Tentative	Slow	Quick
Tense	Relaxed	Raspy	Intense
Harsh	Loud	Gentle	Breathy

*Num id ridets bonrum
eloquentian probo
safenys*

If there are any other words you associate with this typeface, please list them:

Date:

Participant Number:

17. In this image, which phrase fits best with what the typeface describes? Please check only one box.

- ☐ The speaker is singing.
- ☐ The speaker is bored.
- ☐ The speaker is shouting.
- ☐ The volume of the speaker's voice is changing.
- ☐ The speaker is happy.
- ☐ The pitch of the speaker's voice is changing gradually.

18. In this image, which phrase fits best with what the typeface describes? Please check only one box.

- ☐ The speaker is whispering.
- ☐ The speaker is nervous.
- ☐ The pitch of the speaker's voice is shaking.
- ☐ The speaker is singing.
- ☐ The speaker's voice is quiet.
- ☐ The speaker is angry.

19. In this image, which phrase fits best with what the typeface describes? Please check only one box.

- ☐ The speaker is angry.
 - ☐ The speaker is nervous.
 - ☐ The speaker is speaking normally.
 - ☐ The speaker's voice is rough.
 - ☐ The speaker is singing.
 - ☐ The speaker's voice is quiet.
-



Date:
Participant Number:

20. In this image, which phrase fits best with what the typeface describes? Please check only one box.

- ☐ The speaker is speaking normally.
 - ☐ The speaker is nervous.
 - ☐ The speaker is sad.
 - ☐ The speaker's voice is smooth.
 - ☐ The speaker is yelling.
 - ☐ The speaker's voice is breathy.
-

21. Do you have any further comments about how this video was captioned?

Appendix D: Prosodic Type Samples

Typeface at
23 pt. in black
captioning box

Regular prosodic typography,
black captioning box.
The regular brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had locked the
hen-houses for the night, but was too drunk
to remember to shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses for the
night, but was too drunk to remember to shut the popholes. With
the ring of light from his lantern dancing from side to side, he
lurched across the yard, kicked off his boots at the back door, drew
himself a last glass of beer from the barrel in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ¿ ' ; / _ ()
— — — “ ” ‘ ’ \$ = - + & ^

Typeface at
23 pt. in black
captioning box

**Shouting prosodic typography,
black captioning box.
The shouting brown fox
jumps over the lazy dog.**

Typeface
at 18 pt.

**Mr. Jones, of the Manor Farm, had locked
the hen-houses for the night, but was too
drunk to remember to shut the popholes.**

Typeface
at 12 pt.

**Mr. Jones, of the Manor Farm, had locked the hen-houses for the
night, but was too drunk to remember to shut the popholes. With
the ring of light from his lantern dancing from side to side, he
lurched across the yard, kicked off his boots at the back door, drew
himself a last glass of beer from the barrel in the scullery**

Full repertoire
of typeface
at 24 pt.

**A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ¿ ' ; / _ ()
_ - . " " ' ' \$ = - + & ^**

Typeface at
23 pt. in black
captioning box

Whispering prosodic typography,
black captioning box.
The whispering brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had locked the
hen-houses for the night, but was too drunk
to remember to shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses for the
night, but was too drunk to remember to shut the popholes. With
the ring of light from his lantern dancing from side to side, he
lurched across the yard, kicked off his boots at the back door, drew
himself a last glass of beer from the barrel in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ¿ ' ; / _ ()
— _ - " " ' ' ' \$ = - + & ^

Typeface at
23 pt. in black
captioning box

*Quick prosodic typography,
black captioning box.
The quick brown fox
jumps over the lazy dog.*

Typeface
at 18 pt.

*Mr. Jones, of the Manor Farm, had locked the
hen-houses for the night, but was too drunk to
remember to shut the popholes.*

Typeface
at 12 pt.

*Mr. Jones, of the Manor Farm, had locked the hen-houses for the
night, but was too drunk to remember to shut the popholes. With the
ring of light from his lantern dancing from side to side, he lurched
across the yard, kicked off his boots at the back door, drew himself a
last glass of beer from the barrel in the scullery*

Full repertoire
of typeface
at 24 pt.

*A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ' ; / _ ()
— — — “ ” ‘ ’ \$ = - + & ^
ae as ce cs ee es ge gs he hs
je js me ms ne ns te ts tt ue us*

Typeface at
23 pt. in black
captioning box

Slow prosodic typography,
black captioning box.
The slow brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had
locked the hen-houses for the night,
but was too drunk to remember to
shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses
for the night, but was too drunk to remember to shut the
popholes. With the ring of light from his lantern dancing
from side to side, he lurched across the yard, kicked off his
boots at the back door, drew himself a last glass of beer
from the barrel in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ' ; / _ ()
— — — “ ” ‘ ’ \$ = — + & ^

Typeface at
23 pt. in black
captioning box

Slow prosodic typography,
black captioning box.
The slow brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had locked
the hen-houses for the night, but was too
drunk to remember to shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses for the
night, but was too drunk to remember to shut the popholes. With
the ring of light from his lantern dancing from side to side, he
lurched across the yard, kicked off his boots at the back door, drew
himself a last glass of beer from the barrel in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! j # . ? ; ' ; / _ ()
— — — “ ” ‘ ’ \$ = - + & ^

Typeface at
23 pt. in black
captioning box

Relaxed prosodic typography,
black captioning box.
The relaxed brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had locked
the hen-houses for the night, but was too
drunk to remember to shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses
for the night, but was too drunk to remember to shut the
popholes. With the ring of light from his lantern dancing from
side to side, he lurched across the yard, kicked off his boots at
the back door, drew himself a last glass of beer from the barrel
in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ' ; / _ ()
— — — “ ” ‘ ’ \$ = - + & ^

Typeface at
23 pt. in black
captioning box

Trembling prosodic typography,
black captioning box.
The trembling brown fox
jumps over the lazy dog.

Typeface
at 18 pt.

Mr. Jones, of the Manor Farm, had locked
the hen-houses for the night, but was too
drunk to remember to shut the popholes.

Typeface
at 12 pt.

Mr. Jones, of the Manor Farm, had locked the hen-houses for
the night, but was too drunk to remember to shut the popholes.
With the ring of light from his lantern dancing from side to side,
he lurched across the yard, kicked off his boots at the back door,
drew himself a last glass of beer from the barrel in the scullery

Full repertoire
of typeface
at 24 pt.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m
n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9
* \ : , ! ; # . ? ' ; / _ ()
— — — " " ' ' \$ = - + & ^
ae an ee er ff in ll nd
oo re st the tt

Works Cited

- Abrams, M. H. and Geoffrey Galt Harpham. "Prosody." *A Glossary of Literary Terms*. Boston: Thomson Wadsworth, 2005. 256. Print.
- "About CHS." Canadian Hearing Society. Web. 1 May 2012.
- Alves, Mark J.. "Sino-Vietnamese Grammatical Vocabulary and Sociolinguistic Conditions for Borrowing." *Journal of the Southeast Asian Linguistics Society* 1 (2009): 1–9. Electronic.
- Baines, Phil, and Andrew Haslam. *Type & Typography*. 2nd ed. London: Laurence King Publishing Ltd., 2005. Print.
- Bodine, Kerry, and Mathelide Pignol. "Kinetic Typography-Based Instant Messaging." *CHI 2003 Poster Session Ft. Lauderdale, Florida, April 5–10, 2003*. 914–915. Electronic.
- Brown, Steven E.. *Film–Video Terms and Concepts*. Stoneham: Butterworth-Heinemann, 1992. Print.
- "CEA-608 and CEA-708 Feature Comparison Chart." *Closed Captioning and Subtitles Software and Services*. CPC, n.d. Web. 23 Apr 2012.
- Canada. Canadian Radio-television Telecommunications Commission. Broadcasting and Telecom Regulatory Policy CRTC 2009-430: Accessibility of Telecommunications and Broadcasting Services. Gatineau: CRTC, 2009. Electronic.
- Canada. Canadian Radio-television Telecommunications Commission. Broadcasting Notice of Consultation CRTC 2011-488: Call for comments on proposed English-language Closed Captioning Quality Standards. Gatineau: CRTC, 2011. Electronic.
- Canada. Canadian Radio-television Telecommunications Commission. *Broadcasting Regulatory Policy CRTC 2009-741: Quality standards for French-language closed captioning*. Gatineau: CRTC, 2011. Electronic.
- Carlyle, Thomas. *The French Revolution: A History*. London: Robson, Levey, and Franklyn, 1858. *Google Books*. Print.
- Chang, Karen. *Designing Type*. London: Laurence King Publishing Ltd., 2005. Print.
- Chappell, Warren. *A Short History of the Printed Word*. New York: The New York Times, 1970. Print.
- Civera, Clara, and Pilar Orero. "Introducing Icons in Subtitles for the Deaf and Hard of Hearing: Optimising Reception?" *Listening to Subtitles: Subtitles for the Deaf and Hard of Hearing*. Eds. Anna Matamala & Pilar Orero. Bern: Peter Lang, 2010. Print.
- de Angulo, Jamie. "Gramatical Processes: Incremental vs. Autonomic." *Language* 5.2 (1929): 117–118. *JSTOR*. Print.
- Dachkovsky, Svetlana, and Wendy Sandler. "Visual Intonation in the Prosody of a Sign Language." *Language and Speech* 52.2/3 (2009): 287–314. Print.
- Factor, R. Lance. "Peirce's Definition of Metaphor and its Consequences." *Peirce's Doctrin of Signs: Theory, Applications, and Connections*. Eds. Vincent M. Colapietro, and Thomas M. Olshewsky

- Fels, Deborah, et al. "Emotive Captioning and Access to Television." Proceedings of the Eleventh Americas Conference on Information Systems. Omaha, NE, USA August 11–14 2005. Web. 21 Aug. 2011.
- Forlizzi, Jodi, Johnny Lee, and Scott E. Hudson. "The Kinedit System: Affective Messages Using Dynamic Texts." *CHI 2003 Ft. Lauderdale, Florida, USA April 5-10, 2003*. 377–384. Electronic.
- Fourney, David, and Deborah Fels. "'Thanks for Pointing that Out.' Making Sarcasm Accessible for All." Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting. 2008. 571–5. Web. 23 Sep. 2010.
- Füssel, Stephan. *Gutenberg and the Impact of Printing*. Trans. Douglas Martin. Aldershot: Ashgate Publishing Limited, 2005. Print.
- Horace. "On the Art of Poetry." *Classical Literary Criticism*. Ed. Betty Radice and Robert Baldick. Trans. T. S. Dorsch. Harmondsworth: Penguin Books. 1965. Print.
- Hutchings, R. S.. *The Western Heritage of Type Design*. London: Cory, Adams & MacKay LTD, 1963. Print.
- Jones, Phil. "Italicization and Understanding Texts Through Metaphoric Projections of Movement," *Visible Language* 41.3 (2007): 246–265. Print.
- Knowles, Murray, and Rosamund Moon. *Introducing Metaphor*. Florence: Routledge, 2005. Electronic.
- Lakoff, George and Mark Johnson. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980. Print.
- Laver, John. *The Phonetic Description of Voice Quality*. New York: Cambridge University Press, 1980.
- Lee, Daniel G., Deborah I. Fels, and John Patrick Udo. "Emotive Captioning," *ACM Computers in Entertainment* 5.2 (2007): 1–15. Electronic.
- Lupton, Ellen. *Thinking With Type*. 2nd ed. New York: Princeton Architectural Press, 2010. Print.
- Malik, Sabrina, Jonathan Aitken, and Kelly Waalen. "Communicating Emotion with Animated Text." *Visual Communication* 8.4 (2009): 469–79. Web. 21 Aug. 2011.
- Marr, David G.. *Vietnamese Tradition on Trial, 1920 to 1945*. Berkeley: University of California Press 1981. *ACLS Humanities E-Books*. Electronic.
- Mayo Clinic Staff. "Essential Tremor." *Mayo Clinic Health Information*. May Clinic, 10 Aug 2010. Web. 8 Jul 2012.
- Nguyễn, Đình Hòa. "Vietnamese." *International Encyclopedia of Linguistics*. Ed. William J. Frawley. Oxford University Press 2005. Oxford Reference Online. Web. 25 Apr. 2012
- "On August 31, 2011, Canadian Local Over-the-Air Television Stations Converted to Digital Television." *Canadian Radio-Television Telecommunications Commission*. Government of Canada, 22 Sep 2011. Web. 23 Apr 2012.
- "Open Sans." *Google Web Fonts*. Google, n.d. Web. 11 May 2012.

- Nørgaard, Nina. "The Semiotics of Typography in Literary Texts: A Multimodal Approach." *Orbis Litterarum* 64.2, 2009. 141–160. Print.
- Peirce, Charles Sanders. *The Collected Papers of Charles Sanders Peirce*. Ed. Charles Hartshorne And Paul Weiss. Cambridge: Harvard University Press. 1932. *Intelix Past Masters*. Electronic.
- Peirce, Charles Sanders. *The Essential Peirce*. Ed. the Peirce Edition Project. Bloomington: Indiana University Press. 1998. Print.
- Pelli, Denis G., And Katherine A. Tillman. "Parts, Wholes, and Context in Reading: A Triple Dissociation." *PLoS ONE* 2.8 (2007): 1–8. Electronic.
- Perea, Manuel, and Eva Rosa. "Does 'Whole-Word Shape' Play a Role in Visual Word Recognition." *Perception & Psychophysics* 64.5 (2002): 785–794. Print.
- Perron, Paul, and Marcel Danesi, eds. *Classical Readings in Semiotics*. New York: LEGAS. 2003. Print.
- Pham, Andrea Hoa. *Vietnamese Tone: A New Analysis*. Florence: Routledge, 2003. Electronic.
- "prosody, n.." *Oxford English Dictionary*. 3rd ed. Oxford University Press. March 2012. Web. 3 April 2012.
- Rashid, Raisa, *et al.* "Dancing with Words: Using Animated Text for Captioning." *International Journal Of Human–Computer Interaction* 24.5 (2008): 505–19. Web. 21 Aug. 2011.
- Rosenberger, Tara, and Ronald L. MacNeil. "Prosodic Font: Translating Speech into Graphics." *Computer-Human Interaction Conference of the ACM*. Pittsburg, PA. 1999. Web. 21 Aug. 2011.
- The Straight Story*. Dir. David Lynch. Walt Disney Pictures, 1999. DVD.
- Strauss, Karen Peltz. *A New Civil Right: Telecommunications Equality for Deaf and Hard of Hearing Americans*. Washington: Gallaudet University Press, 2002. Print.
- Swanson, Gunnar, ed. *Graphic Design and Reading*. New York: Allworth Press, 2000. Print.
- Thomas, Ambroise, M. Carré, and J. Barbier. *Hamlet: Opéra en Cinq Actes*. PDF.
- van Leeuwen, Theo. *Introducing Social Semiotics*. New York: Routledge, 2005. Print.
- . "Towards a Semiotics of Typography." *Information Design Journal + Document Design* 14.2 (2006): 139–55. Print.
- . "Voice: A Semiotics of the Voice." *Sound and Music in Film and Visual Media: A Critical Overview*. Ed. Graeme Harper. New York: Continuum, 2009. 425–36. Print.
- Wang, Hua, Helmut Prendinger, and Takeo Igarashi. "Communicating Emotion in Online Chat using Physiological Sensors and Animated Text." *CHI 2004 Vienna, Austria, April 24–29, 2004*. 1171–1174. Print.
- Warde, Beatrice. "The Crystal Goblet, or Printing Should Be Invisible." *Graphic Design & Reading*. Ed. Gunnar Swanson. New York: Allworth Press, 2000. 91–95. Print.
- Willen, Bruce, and Nolen Strals. *Lettering & Type: Creating Letters and Designing Typefaces*. New York: Princeton Architectural Press, 2009. Print.