MELODY MATCHER: A MUSIC-LINGUISTIC APPROACH TO ANALYZING THE INTELLIGIBILITY OF SONG LYRICS

A Thesis

Presented to

the Faculty of California Polytechnic State University

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Jennifer "Jenee" Gayle Hughes

June 2012

COMMITTEE MEMBERSHIP

TITLE:	Melody Matcher: A Music-Linguistic Approach to Analyzing the Intelligibility of Song Lyrics
AUTHOR:	Jennifer "Jenee" Gayle Hughes
DATE SUBMITTED:	June 2012
COMMITTEE CHAIR:	Zoë Wood, Ph.D.
COMMITTEE MEMBER:	Franz Kurfess, Ph.D.
COMMITTEE MEMBER:	John Clements, Ph.D.

Abstract

Melody Matcher: A Music-Linguistic Approach to Analyzing the Intelligibility of Song Lyrics

Jennifer "Jenee" Gayle Hughes

Melody Matcher is a semi-automated music composition support program. It analyzes English lyrics along with a melody, and alerts the composer of the locations in the song where the lyrics are not deterministically understandable. Basically, it's grammar- and spell-check for songs. This is significant, because very little research has been done specifically on the quantifiable measurement of English-language lyric intelligibility, other than our project.

Melody Matcher aims to replicate the human ability to identify lyrics in a song that are easily misheard. We started on this project, thinking that there would be carefully-specified research on how lyrics match melodies, mathematically. As it turned out, there was very little objective literature on the subject. Because of the lack of objective information of the subject, we had to develop our method from scratch. As we progressed through our work, we went from thinking that understandability depended only on emphasis-matching, to realizing that syllable length played a huge part as well, to realizing that there are many other musical, harmonic, and linguistic factors.

Melody Matcher analyzes the intelligibility of song lyrics by investigating several root causes:

- Lyric/Music emphasis mismatch, due to:
  - Note intervals

  - Phrase emphases
  - Word emphases

- Word "cramming", due to:
  - Syllable lengths that exceed that of note length
  - Mouth movement delta time intervals

- Word misidentification, due to:
  - Altered pronunciation of words
  - Phone similarity
    * Voicing (voiced vs. voiceless)
    * Beginning/end mouth positions
    * Type (Plosive, Fricative, affricate, nasal, lateral, approximant, semivowel)
  - Phone sequences with multiple syntactically-correct interpretations

The fully-implemented Melody Matcher program will eventually take into account all of these causes of unintelligibility. In this abstract, we will focus on lyric/emphasis mismatch, which has already been implemented and is fully functional in primary testing. The other sections have been implemented, but are not fully tested and/or integrated into the main program.

0.1 Target Audience and Goals

This program is to be used as a compositional aid by anyone who wants to write songs and make them sound good, technically. It should allow the song

writer to focus on more subjective criteria of what makes a song "good", because it will make the structural rules of lyric composition immediately apparent.

Our hope for this project is that it will be useful to burgeoning songwriters, who have the creative spark to make wonderfully poetic lyrics, but lack the "ear" to match their lyrics successfully to music. It should be particularly helpful to songwriters who place a high emphasis on understandability of lyrics (such as parody song writers, or lyricists for musical theater).

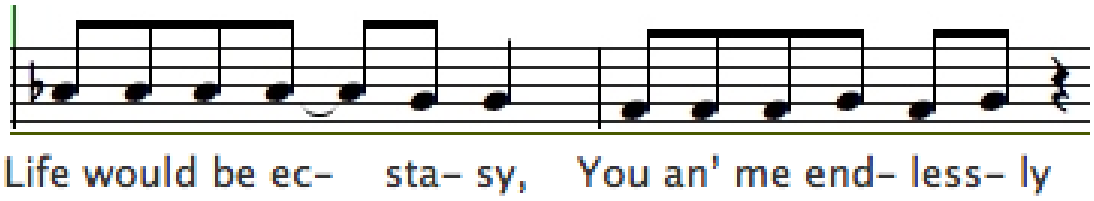
Additionally, Melody Matcher will be useful for songwriters for whom English is a second language. While they may be a master lyricist in their native language, writing lyrics in English can be a particular challenge, since so much of lyric-writing is dependent upon knowing the cadence of the language you're writing lyrics in, and since English has no easily-discernible rules for emphasis placement in words.

0.2 PRACTICAL EXAMPLE OF UNDERLYING THEORY

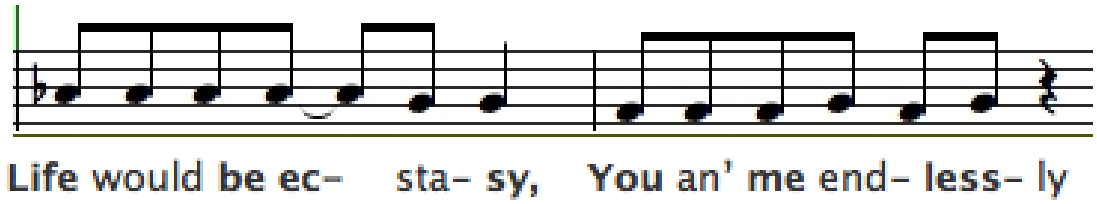
The structural rules of lyric placement are important, because without them, lyrics can become muddled and/or unin- telligible. For example, in the song *"Groovin' (on a Sunday Afternoon)"*, by the Young Rascals, there's a part in the bridge that many people hear as *"Life would be ecstasy, you an' me an' Leslie"*. In fact, the line is *"Life would be ecstasy, you and me endlessly"*. The confusion lies with the last three syllables of the phrase. The pronunciation of each version, if spoken normally, is as follows:

Alphabetic:	and Les- lie	end- less- ly
SAMPA:	@nd “lEs li	“End l@s li

So, in the first phrase, we see that the emphasis pattern can be simplified to “dum DUM-dum”, where the first syllable of “Leslie” is emphasized. The second phrase’s emphasis pattern is “DUM-dum-dum”, so the first syllable of “endlessly” is emphasized. When words are put to music, however, the musical emphasis overrides the textual emphasis. Sometimes, the meaning of the phrase can change, if a previously un-emphasized syllable becomes emphasized, or a previously emphasized syllable loses its emphasis. For “Groovin”’, the lyrics match up to the music in the song as follows:



In this musical phrase, the emphasis always goes on the first part of a beat (for the purposes of this example, a “beat” is defined as a quarter note). In this case, the first measure is emphasized for the notes that correspond to the lyrics, “Life”, “be”, “ec-”(as in ec-sta- sy) and “sy”(again, as in ec-sta-sy) (This is a vast oversimplification, but it works for now). So, the lyrics would be emphasized as such:



Or, more simply:

1. `Life would be ec-sta-sy`
2. Life would be ec-sta-sy
3. LIFE would BE EC-sta-SY
4. *Life* would *be ec*-sta-*sy*
5. LIFE WOULD BE EC-STA-SY
6. LIFE would BE EC-sta-SY
7. LIFE would BE EC-sta-SY
8. LIFE would BE EC-sta-SY
9. LIFE would BE EC-sta-SY
10. *Life* WOULD *be ec*-STA-*sy*
11. LIFE *would* BE EC-*sta*-SY
12. **Life** *would* **be ec**-*sta*-**sy**
13. **Life** WOULD **be ec**-STA-**sy**
14. **Life would be ec-sta-sy**
15. **Life** would **be ec**-sta-**sy**
16. *Life* would *be ec*-sta-*sy*
17. Life would be ec-sta-sy
18. *Life would be ec-sta-sy*
19. Life *would* be ec-*sta*-sy

20. *Life* would *be ec*-sta-*sy*

21. LIFE would BE EC-sta-SY

22. *Life* would *be ec*-sta-*sy*

23. LIFE would BE EC-sta-SY

24. LIFE would BE EC-sta-SY

This musical emphasis matches the spoken emphasis of the phrase, so it is intelligible as a lyric. (Though ecstasy's first syllable doesn't start on the first part of beat three, it is still on the first part of beat three, and therefore still emphasized. Alternatively, since the first part of beat two didn't have a hard stop to it, the emphasis could have rolled over to the second part, "ec", which does have a hard stop.)

In contrast, take the second measure: the syllables "You", "me", and "less" are emphasized in the music. This leads to conflicting musical and spoken phrasing:

Musical Phrasing: **You** and **me** end**less**ly

Spoken Phrasing: **You** and **me end**lessly

The singer is now singing the phrase, syllable by syllable, which they think of as syllable-note combinations:

YOU and ME end LESS lee

The singer, for his part, is doing what many singers are taught to do, to make it easier to sustain the singing of words that end with unsingable consonants: the

unsingable consonant is displaced onto the front of the next word. In this case, the consonant "d" is not singable, so he displaces it onto the next syllable, when he can: "and ME" becomes "an dME", and "end LESS" becomes "en dLESS". So, the singer can effectively think of the sung phrase as:

YOU an dME en dLESS lee

This doesn't cause confusion for listeners, because they're used to hearing it. This does mean, however, that lyric placement does not provide an accurate barometer to a listener of where a word actually ends.

In addition, the singer is singing fudging his vowels, like singers are taught to do, so "and" and "end" sound almost indistinguishable. So, really, what listeners are hearing is this:

YOU en dME en dLESS lee

Now, the listener's brain has to take this syllabic gobbledy-gook, and parse it into something useful. They've currently got this mess to deal with (represented in SAMPA syllables):

ju *En* ***dmi*** *En* ***dl@s*** *li*

They parse the first part just fine, because the emphases match:

you and **me** *En* ***dl@s*** *li*

But no one says endLESSly. People say ENDlessly. So, the listeners don't recognize it. They have to work with what they have. They already turned one "En d" into an "and", so they do it again:

you and **me** and ***l@s*** *li*

Now, they're just left with LESS lee. And that fits Leslie, a proper noun that fits in context and in emphasis placement. So, the final heard lyric is:

you and **me** and **Les-** lie

The misunderstanding can be traced back to improper emphasis placement. The songwriter probably didn't even think of that, and now he's stuck: a one-hit-wonder with a misunderstood song. We bet that in interview after interview, someone asks him who Leslie is. It's probably very frustrating — especially since he could have just moved the word an eight note later, and it would have been understood perfectly.

That's the sort of situation this program is going to help avoid.

III. FUTURE WORK

We plan to continue developing and refining the methods through which Melody Matcher makes its determinations. Eventually, we plan to use this as an underlying framework for an interactive virtual environment where your surroundings are affected and created via musical and lyrical input. This should be completed in March 2012, with incremental updates discussed on www.melodymatcher.com.

IV. CONCLUSION

In this paper, we have discussed at a high level parts of the music-lingustic approach that Melody Matcher takes to measure the intelligibility of lyrics. We covered some of the major reasons that lyrics get misheard, along with a few examples. Melody Matcher's specific implementation details, while fully specified

elsewhere, were outside the scope of this abstract, and we hope to cover them in a later paper.

sflashjdgkfjshfdgvksjhfga

Or, more simply: Life would be ec-sta-sy This musical emphasis matches the spoken emphasis of the phrase, so it is intelligible as a lyric. (Though ecstasy's first syllable doesn't start on the first part of beat three, it is still on the first part of beat three, and therefore still emphasized. Alternatively, since the first part of beat two didn't have a hard stop to it, the emphasis could have rolled over to the second part, "ec", which does have a hard stop.) In contrast, take the second measure: the syllables "You", "me", and "less" are emphasized in the music. This leads to conflicting musical and spoken phrasing: Musical Phrasing: You and me endlessly Spoken Phrasing: You and me endlessly The singer is now singing the phrase, syllable by syllable, which they think of as syllable-note combinations: YOU and ME end LESS lee The singer, for his part, is doing what many singers are taught to do, to make it easier to sustain the singing of words that end with unsingable consonants: the unsingable consonant is displaced onto the front of the next word. In this case, the consonant "d" is not singable, so he displaces it onto the next syllable, when he can: "and ME" becomes "an dME", and "end LESS" becomes "en dLESS". So, the singer can effectively think of the sung phrase as: YOU an dME en dLESS lee

This doesn't cause confusion for listeners, because they're used to hearing it. This does mean, however, that lyric placement does not provide an accurate barometer to a listener of where a word actually ends. In addition, the singer is singing fudging his vowels, like singers are taught to do, so "and" and "end" sound almost indistinguishable. So, really, what listeners are hearing is this: YOU en dME en dLESS lee

Now, the listener's brain has to take this syllabic gobbledy- gook, and parse it into something useful. They've currently got this mess to deal with (represented in SAMPA syllables): ju En dmi En dl@s li They parse the first part just fine, because the emphases you and me En dl@s li But no one says endLESSly. People say ENDlessly. So, the listeners don't recognize it. They have to work with what they have. They already turned one "En d" into an "and", so they do it again: you and me and l@s li Now, they're just left with LESS lee. And that fits Leslie, a proper noun that fits in context and in emphasis placement. So, the final heard lyric is: you and me and Les- lie The misunderstanding can be traced back to improper emphasis placement. The songwriter probably didn't even think of that, and now he's stuck: a one-hit-wonder with a misunderstood song. We bet that in interview after interview, someone asks him who Leslie is. It's probably very frustrating — especially since he could have just moved the word an eight note later, and it would have been understood perfectly. That's the sort of situation this program is going to help avoid. III. FUTURE WORK We plan to continue developing and refining the methods through which Melody Matcher makes its determinations. Eventually, we plan to use this as an underlying framework for an interactive virtual environment where your surroundings are affected and created via musical and lyrical input. This should be completed in March 2012, with incremental updates discussed on www.melodymatcher.com. IV. CONCLUSION In this paper, we have discussed at a high level parts of the music-lingustic approach that Melody Matcher takes to measure the intelligibility of lyrics. We covered some of the major reasons that lyrics get misheard, along with a few examples. Melody Matcher's specific implementation details, while fully specified elsewhere, were outside the scope of this abstract, and we hope to cover them in a later paper.

This is where the abstract goes. Hopefully this document will serve as an example for preparing a Cal Poly Master's thesis. It was thrown together pretty quickly. A lot more neat LaTeX features, help, and examples can be found on the web. Here is one: http://en.wikibooks.org/wiki/LaTeX

For developing LaTeX documents in the Windows environment, I use TeXnic-Center (http://www.toolscenter.org/). A simple WYSIWYG LaTeX editor (though I had problems getting it to work with this thesis format) is **LyX** (http://www.lyx.org/).

I use **InkScape** (http://www.inkscape.org/) to create any drawings/figures needed. It is a free vector graphics editor that is very powerful and popular. There is an example figure produced with InkScape in Figure 1.1. InkScape can export images in many different formats. Export your images as PDF or EPS and put into your LaTeX document. If you're creating a PDF document with **pdflatex**, then export as a PDF image. If you're creating PostScript then export as EPS. Rasterized images such as JPEG can also be easily included in LaTeX.

LaTeX can also produce nice equations. Did you know that $\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots = \frac{\pi}{4}$? A non-inline equation can be found in Figure 2.1. I treated my equations as figures but they can be treated specially as Equations.

An example of a table can be found in Table 3.1.

The bibliography section is very easy to create. When gathering references, I used the ACM digital library (http://portal.acm.org/portal.cfm) to grab the Bibtex entries. Papers in the digital library have Bibtex entries ready to be copied and pasted into your bibliography. Create a separate file called something like "bibliography.bib" and paste in your Bibtex entries. LaTeX (and Bibtex) generate your bibliography section for you – very easy! I can cite references very easily. Here is a paper called *Dual contouring of hermite data* [3]. Here is a

paper called *Surface simplification using quadric error metrics* [2]. I've also cited software located at some websites [1] [4].

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This is the introduction.

LaTeX is a document markup language and document preparation system for the TeX typesetting program.

It is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, and by others as a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF) because of the quality of typesetting achievable by TeX. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language to access the power of TeX. LaTeX essentially comprises a collection of TeX macros, and a program to process LaTeX documents. Since TeX's formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in 1984 by Leslie Lamport at SRI International

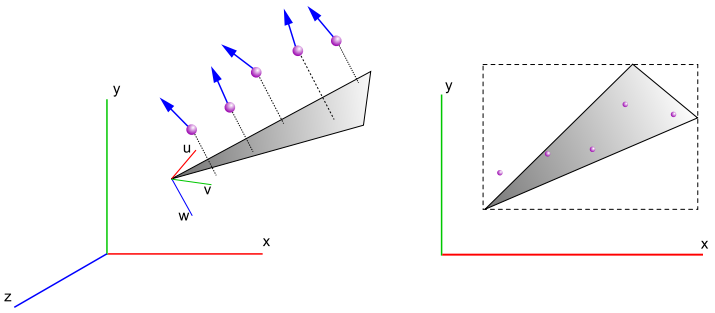


Figure 1.1: This is an example figure. This graphic was drawn in InkScape (http://www.inkscape.org/), a free vector drawing application. Figures can be exported from InkScape as PDF images or as EPS (depending upon if you want LaTeX to generate a PDF or a PostScript document.)

and has become the dominant method for using TeXfew people write in plain TeX anymore.

LaTeX is based on the idea that authors should be able to focus on the meaning of what they are writing, without being distracted by the visual presentation of the information. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures. It therefore encourages the separation of layout from content, while still allowing manual typesetting adjustments where needed. This is similar to the mechanism by which many word processors allow styles to be defined globally for an entire document, or the CSS mechanism used by HTML.

Chapter 2

Previous Work

LaTeX is a document markup language and document preparation system for the TeX typesetting program.

It is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, and by others as a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF) because of the quality of typesetting achievable by TeX. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language to access the power of TeX. LaTeX essentially comprises a collection of TeX macros, and a program to process LaTeX documents. Since TeX's formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in 1984 by Leslie Lamport at SRI International and has become the dominant method for using TeXfew people write in plain

TeX anymore.

LaTeX is based on the idea that authors should be able to focus on the meaning of what they are writing, without being distracted by the visual presentation of the information. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures. It therefore encourages the separation of layout from content, while still allowing manual typesetting adjustments where needed. This is similar to the mechanism by which many word processors allow styles to be defined globally for an entire document, or the CSS mechanism used by HTML.

2.1 This is a new section

LaTeX is a document markup language and document preparation system for the TeX typesetting program.

It is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, and by others as a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF) because of the quality of typesetting achievable by TeX. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language to access the power of TeX. LaTeX essentially comprises a collection of TeX macros, and a program to process LaTeX documents. Since TeX's formatting commands are very low-level,

$$\begin{bmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} p'_x \\ p'_y \\ p'_z \end{bmatrix}$$

Figure 2.1: This is a sample matrix equation.

it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in 1984 by Leslie Lamport at SRI International and has become the dominant method for using TeXfew people write in plain TeX anymore.

LaTeX is based on the idea that authors should be able to focus on the meaning of what they are writing, without being distracted by the visual presentation of the information. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures. It therefore encourages the separation of layout from content, while still allowing manual typesetting adjustments where needed. This is similar to the mechanism by which many word processors allow styles to be defined globally for an entire document, or the CSS mechanism used by HTML.

1. Here is a list item.

    (a) Here is a sub list item.

        i. Here is a sub sub list item.

Chapter 3

Results

Here is the results section.

LaTeX is a document markup language and document preparation system for the TeX typesetting program.

It is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, and by others as a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF) because of the quality of typesetting achievable by TeX. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language to access the power of TeX. LaTeX essentially comprises a collection of TeX macros, and a program to process LaTeX documents. Since TeX's formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in 1984 by Leslie Lamport at SRI International

and has become the dominant method for using TeXfew people write in plain TeX anymore.

LaTeX is based on the idea that authors should be able to focus on the meaning of what they are writing, without being distracted by the visual presentation of the information. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures. It therefore encourages the separation of layout from content, while still allowing manual typesetting adjustments where needed. This is similar to the mechanism by which many word processors allow styles to be defined globally for an entire document, or the CSS mechanism used by HTML.

LaTeX can be arbitrarily extended by using the underlying macro language to develop custom formats. Such macros are often collected into packages which are available to address special formatting issues such as complicated mathematical content or graphics. In addition, there are numerous commercial implementations of the entire TeX system, including LaTeX, to which vendors may add extra features like additional typefaces and telephone support. LyX is a free visual document processor that uses LaTeX for a back-end. TeXmacs is a free, WYSIWYG editor with similar functionalities as LaTeX, but a different typesetting engine.

A number of popular commercial desktop publishing systems use modified versions of the original TeX typesetting engine. The recent rise in popularity of XML systems and the demand for large-scale batch production of publication-quality typesetting from such sources has seen a steady increase in the use of LaTeX.

LaTeX is a document markup language and document preparation system for

	Some Data			Some More Data		
	Hi-Res	Lo-Res	Reduction	Hi-Res	Lo-Res	Speedup
Row Data A	225,467	43,850	80.6%	360	90	4.0
Row Data B	225,467	16,388	92.7%	360	26	13.8

Table 3.1: Here is some performance data for the system.

the TeX typesetting program.

It is widely used by mathematicians, scientists, philosophers, engineers, and scholars in academia and the commercial world, and by others as a primary or intermediate format (e.g. translating DocBook and other XML-based formats to PDF) because of the quality of typesetting achievable by TeX. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language to access the power of TeX. LaTeX essentially comprises a collection of TeX macros, and a program to process LaTeX documents. Since TeX's formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

LaTeX was originally written in 1984 by Leslie Lamport at SRI International and has become the dominant method for using TeXfew people write in plain TeX anymore.

LaTeX is based on the idea that authors should be able to focus on the meaning of what they are writing, without being distracted by the visual presentation of the information. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures.

It therefore encourages the separation of layout from content, while still allowing manual typesetting adjustments where needed. This is similar to the mechanism by which many word processors allow styles to be defined globally for an entire document, or the CSS mechanism used by HTML.

LaTeX can be arbitrarily extended by using the underlying macro language to develop custom formats. Such macros are often collected into packages which are available to address special formatting issues such as complicated mathematical content or graphics. In addition, there are numerous commercial implementations of the entire TeX system, including LaTeX, to which vendors may add extra features like additional typefaces and telephone support. LyX is a free visual document processor that uses LaTeX for a back-end. TeXmacs is a free, WYSIWYG editor with similar functionalities as LaTeX, but a different typesetting engine.

A number of popular commercial desktop publishing systems use modified versions of the original TeX typesetting engine. The recent rise in popularity of XML systems and the demand for large-scale batch production of publication-quality typesetting from such sources has seen a steady increase in the use of LaTeX.

Bibliography

[1] ATI NormalMapper. http://ati.amd.com/developer/tools.html.

[2] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[3] T. Ju, F. Losasso, S. Schaefer, and J. Warren. Dual contouring of hermite data. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 339–346, New York, NY, USA, 2002. ACM Press.

[4] nVidia Melody. http://developer.nvidia.com/object/melody_home.html.