Abstract—Minimal research has been done on how letter repetition affects readers’ perception of expressed sentiment within a text. To the best of researchers’ knowledge, no studies have tested samples of text with letter repetition using sentiment tools. The main aim of this paper is to investigate whether letter repetition in product reviews are perceived to have any sentiment value, based on ratings by individual participants and analyses using sentiment tools. This study collected and analysed 1,041 consumer reviews in the form of online comments using the UCREL Wmatrix system, and simulated emotional words within the comments to contain repeated letters. A group of 500 participants rated 15 positive comments and 15 negative comments and their respective simulated counterparts, while 32 sentiment tools are used to analyse a pair of positive comment and its simulated counterpart and a pair of negative comment and its simulated counterpart. Results indicate that readers perceive letter repetition to amplify a comment’s sentiment value, in which the effect was found more strongly in negative comments than positive comments. On the other hand, analyses using sentiment tools show that a majority of these tools are unable to detect letter repetition within a word and instead, treats the word as a spelling mistake. As consumers or online users, in general, have been found to use letter repetition to intensify and express their sentiments in their comments, this study’s findings suggest that letter repetition processing in any text-based mechanism needs to be enhanced. The outcome of this paper is useful for improving the measurement of sentiment analysis for the use of marketing applications.

Index Terms—Computer-Mediated Communication (CMC); Letter Repetition; Online Reviews; Product Reviews; Sentiment Tools; Text-Based Cue.

I. INTRODUCTION

Social media text, such as Twitter posts and product reviews, often contains a variety of non-verbal and non-grammatical codes and symbols including exclamation marks, emoticons, and letter repetition. Such symbols are usually used to express mood, intonation, and emphasis that are ignored or difficult to convey in the text [1]. Past researchers found that letter repetition defined as a paralinguistic cue in relaying non-verbal communication via computer-mediated channels [2][3]. For example, Carey [2] observed that paralinguistic features and concluded that people find it important to outline tonal and expressive information even if such information is difficult to convey. Carey [2] categorised the usage of repeated letters as vocal spelling (e.g., “weeeell” and “breakkk”), lexical and vocal surrogates (e.g., “Boo, boo Horror of horrors!!..”, “uh huh” and “hmmmm”).

Another study by Darics [4] also examined the specific use of letter repetition in conveying socio-emotional messages and evoking auditory cues through a single letter repetition. This is a common phenomenon in social media platforms like Twitter [5], identifying the real expressed meaning of the letter repetition accurately have a significant contribution to the understanding of sentiment in the text. As shown in the findings of the aforementioned research, letter repetition usage is prevalent and may play a role sentiment analysis of online product or service reviews.

This study examines letter repetition usage in product reviews and how it affects readers’ perception of positive and negative sentiment in online comments on commercial products. The study’s primary goal is to investigate whether there is a significant difference in sentiment expression when letter repetition is used. The focus is particularly on letter repetition within the most sentimentally expressive word in the statement. Additionally, the study also examines the accuracy of available online sentiment tools in letter repetition detection. A sample of reviews is tested on 32 sentiment tools are used to explore how these tools reflect letter repetition in their sentiment scores. Findings from this study contribute to the accurate detection and measurement of consumers’ preferences and attitudes toward commercial products, which is key to understanding online consumers’ behaviour.

II. LITERATURE REVIEW

Several studies on online language found that the usage of letter repetition increases when emotionally-laden interjections are employed. Kalman and Gergle [5]–[7] suggested that repeated letters and punctuations indicate the stretching of a word, emulating a stretched-out syllable of how words are articulated in a spoken conversation (e.g., “It is sweeeeeeet” and “Whaaaaassssuppppp”). It was found that vowels are repeated more often than consonants on average, and that letter repetition functions to denote a change in pitch (e.g., “Veeeeeeehaaaaw!!!!!!!”), decrease in voice volume (e.g., “sshhhhhhhh....”), a pause (e.g. “Hmmm”), or sounds (e.g. “vvvvrrrrrroooooommnnmm,” “pffffff,” “Heeeeheeee!,” “ugggghhhhh!!!”, and “Happy birthday to youuuu”) [7]. Besides communicating pitch, tempo, and prosody, letter repetitions also feature other paralinguistic elements that focus on achieving visual emphasis (e.g., “lllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllllll
they represented words in other languages, slang, abbreviations, or acronyms not found in the dictionary (e.g., "gonna"); (c) whether they were onomatopoeic words — words that imitate sounds (e.g., "boom" or "grrrr"); and (d) whether they can be attributed to the name of the message’s sender name or e-mail address. These initial studies emphasise the prevalent usage of letter repetition and how it may play a strong pragmatic role in online product or service reviews. For instance, letter repetition in messages are often found to be heavy with emotionally-laden interjections (e.g., “ooops”) [7] and may imitate phoneme extension found in a spoken conversation (e.g., “soooo”) [5]. These cues are used to express information beyond the literal meaning of the message, suggesting a pragmatic intention not present in the words themselves.

III. METHODS

A. Survey Set-Up

This study collected a total of 1,041 online review comments from different social media platforms, including Amazon, e-Bay, Facebook, and GSM Arena. These reviews are taken from the following product categories [1]: (a) Beauty and Health; (b) Camera; (c) Computer; (d) Consumer Electronics; (e) Fashion; (f) Home Appliances; (g) Jewelry and Watches; (h) Mobile and Tablets; (i) Sporting Goods; and (j) Toys and Kids. Other studies have used this same dataset but for different research purposes, such as finding the most accurate machine learning classifier [8], processing emoticons [9], and exploring how emoticons and punctuations are used in online reviews [10]. The current study focuses on the usage of letter repetition and understanding the changes in polarity after simulation of letter repetition.

The collected reviews are analysed using the UCREL Wmatrix system [11] to extract emotional words that appeared most frequently. This resulted in the selection of 30 comments comprising 15 positive comments and 15 negative comments. These comments are then simulated with letter repetition, whereby a vowel within one keyword for each comment is randomly selected and repeated in patterns that frequently occur in social media messages. The original comments and simulated comments of a positive nature are shown in Table 1 while those of a negative nature are shown in Table 2. The simulation was performed to test how it affects polarity when letter repetition is used in the text. Following that, five hundred participants were requested to rate simulated text scaling from 1—“Strongly Dislike” to 7—“Strongly Like”. Participants were chosen based on random sampling within Subang Jaya district, Malaysia, with a various age range. All of the participants have experience in reading and writing online reviews.

B. Sentiment Tools’ Set-Up

Numerous sentiment tools are available online for various research analyses. For instance, SentiStrength analyses short informal text [12] while TensiStrength is used to detect relaxation magnitude in social media text [13]. This study selected 32 freely available online sentiment tools (Table 3) to explore how they detect and reflect letter repetition in their sentiment scores.

<table>
<thead>
<tr>
<th>Text</th>
<th>Simulated text with letter repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I love it</td>
<td>I looooooove it</td>
</tr>
<tr>
<td>I like it</td>
<td>I liiiiiiiike it</td>
</tr>
<tr>
<td>I am very happy</td>
<td>I am very haaaaaappy</td>
</tr>
<tr>
<td>I am glad</td>
<td>I am glaaaaaaaad</td>
</tr>
<tr>
<td>I am big fan</td>
<td>I am big faaaaaaan</td>
</tr>
<tr>
<td>My favorite</td>
<td>My faaaaaaaavorite</td>
</tr>
<tr>
<td>Hours of fun</td>
<td>Hours of funuuuuun</td>
</tr>
<tr>
<td>Very satisfied</td>
<td>Very saaaaaaaftisfied</td>
</tr>
<tr>
<td>I prefer it</td>
<td>I preeeeeeefer it</td>
</tr>
<tr>
<td>Really enjoy</td>
<td>Really eeeeeeennjoy</td>
</tr>
<tr>
<td>I recommend it</td>
<td>I recommenneeeed it</td>
</tr>
<tr>
<td>Exceed expectations</td>
<td>Exceeeeeeeeeed expectations</td>
</tr>
<tr>
<td>I will continue taking</td>
<td>I will continue taking this</td>
</tr>
<tr>
<td>this brand</td>
<td>braaaaaaan</td>
</tr>
<tr>
<td>Are you kidding me?</td>
<td>Are you kidding meeeeee?</td>
</tr>
<tr>
<td>No need to say more</td>
<td>Nooooo need to say more</td>
</tr>
</tbody>
</table>

From the 1,041 online review comments, a positive comment and a negative comment are selected to be analysed by the sentiment tools. Letter repetition is again simulated in one keyword for each sentence to check for differences in scores between the original comment and the simulated comment with repeated letters. These comments analysed by the sentiment tools are depicted in Table 4.

IV. RESULTS AND DISCUSSION

A. Survey Analysis

To examine the impact of letter repetition in sentiment analysis of online product reviews, the researchers invited 500 participants to rate the intensity and polarity of the sentiment of the 30 comments and their simulated counterparts. A 7-point Likert Scale [15] ranging from “Strongly Dislike” to “Strongly Like” is used to rate the comments. The difference in sentiment rating between each original comment and its simulated counterpart is also recorded. The results of the ratings for positive comments and negative comments are shown in Table 5 and Table 6 respectively. The term “increase” means that the ratings shifted towards “Strongly Like” while the term “decrease” means that the ratings shifted towards “Strongly Dislike”.

<table>
<thead>
<tr>
<th>Text</th>
<th>Simulated text with letter repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some serious abuse</td>
<td>Some serious abuuuuuuuuuse</td>
</tr>
<tr>
<td>Very disappointed</td>
<td>Very disappointed</td>
</tr>
<tr>
<td>I don’t care</td>
<td>I don’t caaaaaaare</td>
</tr>
<tr>
<td>I did hit it well</td>
<td>I did hiiiiininiiit well</td>
</tr>
<tr>
<td>I hate it</td>
<td>I haaaaaaaate it</td>
</tr>
<tr>
<td>It is really annoying</td>
<td>It is really annooooooooying</td>
</tr>
<tr>
<td>I boot it</td>
<td>I hoooooooso0000000000000000</td>
</tr>
<tr>
<td>Too much trouble</td>
<td>Too much trouble</td>
</tr>
<tr>
<td>Totally fierce</td>
<td>Totally fierceeeeemmm</td>
</tr>
<tr>
<td>I have to worry</td>
<td>I have to wooooooooooorry</td>
</tr>
<tr>
<td>I can afford it</td>
<td>I caaaaaaaaafford it</td>
</tr>
<tr>
<td>What a lie</td>
<td>Whaaaaaaat a lie</td>
</tr>
<tr>
<td>Don’t come here to shop</td>
<td>Don’t come here to shop</td>
</tr>
<tr>
<td>Fine until it breaks</td>
<td>Fiiiiiiniiiiiii until it breaks</td>
</tr>
<tr>
<td>Never, ever, never</td>
<td>Never, ever, neeeeeeever</td>
</tr>
</tbody>
</table>

Table 1

Table 2

Table 3

Table 4

Table 5

Table 6
Table 3 Sample for Sentiment Tools Testing

<table>
<thead>
<tr>
<th>No*</th>
<th>Name of Sentiment Tool</th>
<th>Web Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Selasdia Intelligent Sales Assistant</td>
<td><a href="http://www.aiasio.com:8080/annotat">http://www.aiasio.com:8080/annotat</a> or-0.1/automation/demoView/1</td>
</tr>
<tr>
<td>2</td>
<td>Sentaero</td>
<td><a href="http://www.sentaero.com/textsearch.php">http://www.sentaero.com/textsearch.php</a></td>
</tr>
<tr>
<td>3</td>
<td>Meaning cloud</td>
<td><a href="http://www.meaningcloud.com/demo">http://www.meaningcloud.com/demo</a></td>
</tr>
<tr>
<td>4</td>
<td>TheySay</td>
<td><a href="http://apidemo.theysay.io/">http://apidemo.theysay.io/</a></td>
</tr>
<tr>
<td>5</td>
<td>Repustate</td>
<td><a href="https://www.reputstate.com/api-demo/">https://www.reputstate.com/api-demo/</a></td>
</tr>
<tr>
<td>6</td>
<td>Text sentiment analyzer</td>
<td><a href="http://werfamous.com/sentimentanal">http://werfamous.com/sentimentanal</a> yzer</td>
</tr>
<tr>
<td>7</td>
<td>MIOPIA Supervised Model</td>
<td><a href="http://miopia.grupolys.org/demo">http://miopia.grupolys.org/demo</a></td>
</tr>
<tr>
<td>8</td>
<td>SentiStrength</td>
<td><a href="http://sentistrength.wlv.ac.uk/">http://sentistrength.wlv.ac.uk/</a></td>
</tr>
<tr>
<td>9</td>
<td>Python NLTK Demos for Natural Language Text Processing</td>
<td><a href="http://text-processing.com/demo/">http://text-processing.com/demo/</a></td>
</tr>
<tr>
<td>10</td>
<td>Text scoring:</td>
<td><a href="http://sentiment.christopherpotts.net/lexicon/textscores_results/">http://sentiment.christopherpotts.net/lexicon/textscores_results/</a></td>
</tr>
<tr>
<td>11</td>
<td>Text scoring:</td>
<td><a href="http://sentiment.christopherpotts.net/lexicon/textscores_results/">http://sentiment.christopherpotts.net/lexicon/textscores_results/</a></td>
</tr>
<tr>
<td>12</td>
<td>Text scoring:</td>
<td><a href="http://sentiment.christopherpotts.net/lexicon/textscores_results/">http://sentiment.christopherpotts.net/lexicon/textscores_results/</a></td>
</tr>
<tr>
<td>13</td>
<td>Text scoring:</td>
<td><a href="http://sentiment.christopherpotts.net/lexicon/textscores_results/">http://sentiment.christopherpotts.net/lexicon/textscores_results/</a></td>
</tr>
<tr>
<td>14</td>
<td>Text scoring:</td>
<td><a href="http://sentiment.christopherpotts.net/lexicon/textscores_results/">http://sentiment.christopherpotts.net/lexicon/textscores_results/</a></td>
</tr>
<tr>
<td>15</td>
<td>LIWC</td>
<td><a href="http://lwc.wpengine.com/">http://lwc.wpengine.com/</a></td>
</tr>
<tr>
<td>16</td>
<td>Sentiment Analyzer</td>
<td><a href="http://www.danielsoper.com/sentimentanalysis/">http://www.danielsoper.com/sentimentanalysis/</a></td>
</tr>
<tr>
<td>17</td>
<td>Sentiment Analysis:</td>
<td><a href="http://text2data.org/Demo">http://text2data.org/Demo</a></td>
</tr>
<tr>
<td>18</td>
<td>Pattern Sentiment Analysis</td>
<td><a href="http://textanalysisonline.com/pattern-sentiment-analysis/">http://textanalysisonline.com/pattern-sentiment-analysis/</a></td>
</tr>
<tr>
<td>19</td>
<td>Sentiment Vivekn</td>
<td><a href="http://sentiment.vivekn.com/">http://sentiment.vivekn.com/</a></td>
</tr>
<tr>
<td>20</td>
<td>Alchemy Language Document Sentiment Analysis</td>
<td><a href="https://alchemy-language-demo.mybluemix.net/">https://alchemy-language-demo.mybluemix.net/</a></td>
</tr>
<tr>
<td>21</td>
<td>Alchemy Language Targeted Emotion</td>
<td><a href="https://alchemy-language-demo.mybluemix.net/">https://alchemy-language-demo.mybluemix.net/</a></td>
</tr>
<tr>
<td>22</td>
<td>Intellixer</td>
<td><a href="http://demo.intellixer.com/">http://demo.intellixer.com/</a></td>
</tr>
<tr>
<td>23</td>
<td>ParallelDots</td>
<td><a href="http://www.paralleldots.com/sentiment-analysis/">http://www.paralleldots.com/sentiment-analysis/</a></td>
</tr>
<tr>
<td>26</td>
<td>uClassify</td>
<td><a href="https://www.uclassify.com/browse/uclassify/sentiment?inputs=Text">https://www.uclassify.com/browse/uclassify/sentiment?inputs=Text</a></td>
</tr>
<tr>
<td>27</td>
<td>Tone Analyzer</td>
<td><a href="https://tone-analyzer-demo.mybluemix.net/">https://tone-analyzer-demo.mybluemix.net/</a></td>
</tr>
<tr>
<td>28</td>
<td>Pythia Semantic Features</td>
<td><a href="http://omiotis.hua.gr/pythia/#">http://omiotis.hua.gr/pythia/#</a></td>
</tr>
<tr>
<td>29</td>
<td>Pythia Term n-grams</td>
<td><a href="http://omiotis.hua.gr/pythia/#">http://omiotis.hua.gr/pythia/#</a></td>
</tr>
<tr>
<td>30</td>
<td>Pythia Character n-grams</td>
<td><a href="http://omiotis.hua.gr/pythia/#">http://omiotis.hua.gr/pythia/#</a></td>
</tr>
<tr>
<td>31</td>
<td>Pythia All n-grams</td>
<td><a href="http://omiotis.hua.gr/pythia/#">http://omiotis.hua.gr/pythia/#</a></td>
</tr>
<tr>
<td>32</td>
<td>Pythia All Features</td>
<td><a href="http://omiotis.hua.gr/pythia/#">http://omiotis.hua.gr/pythia/#</a></td>
</tr>
</tbody>
</table>

*Numbers of the tools are same for the Table 3, Table 8 and Table 9.

As shown in Table 5, there is an average of 50.68% increment in ratings between the original comments and their simulated version. In our case, the term “increase” means that the rating shifts towards "Strongly Like” value and term “decrease” means that the rating shifts towards "Strongly Dislike” value. This indicates that participants found the simulated comments to have a higher intensity in positive sentiment than their original text.

Table 4 Sample for Sentiment Tools’ Testing

<table>
<thead>
<tr>
<th>Format of Text</th>
<th>Positive text</th>
<th>Example used in experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text without letter repetition</td>
<td>love our new tv. the tv is so light and thin it has a great picture and the colors are true very happy customer</td>
<td>love our new tv. the tv is so light and thin it has a great picture and the colors are true very happy customer</td>
</tr>
<tr>
<td>Text with letter repetition</td>
<td>love our new tv. the tv is so light and thin it has a great picture and the colors are true very happy customer</td>
<td>love our new tv. the tv is so light and thin it has a great picture and the colors are true very happy customer</td>
</tr>
</tbody>
</table>

For example, some participants rated “I am very happy” as only “Slightly Like” but rated “I am very haaaaaappy” as “Like” or “Strongly Like”. The pair of positive comments that underwent the largest increase in ratings is “Really enjoy” and “Really eeeeeeeenjoy”, of which 62.2% of participants increased their ratings for the latter comment towards “Like”. Overall, approximately 31% to 62% of participants increased their “Like” rating for the simulated version.

Table 5 Rating Changes in Positive Comments

<table>
<thead>
<tr>
<th>Positive Comment</th>
<th>Rating increases from Original Comment to Simulated Comment</th>
<th>Rating decreases from Original Comment to Simulated Comment</th>
<th>Rating maintains from Original Comment to Simulated Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Really enjoy</td>
<td>62.2%</td>
<td>23.8%</td>
<td>14.0%</td>
</tr>
<tr>
<td>I love it</td>
<td>61.2%</td>
<td>18.0%</td>
<td>20.8%</td>
</tr>
<tr>
<td>My favorite</td>
<td>59.6%</td>
<td>23.4%</td>
<td>17.0%</td>
</tr>
<tr>
<td>I prefer it</td>
<td>58.4%</td>
<td>24.8%</td>
<td>16.8%</td>
</tr>
<tr>
<td>I am very happy</td>
<td>58.0%</td>
<td>26.4%</td>
<td>15.6%</td>
</tr>
<tr>
<td>I am glad</td>
<td>56.4%</td>
<td>21.0%</td>
<td>22.6%</td>
</tr>
<tr>
<td>I recommend it</td>
<td>54.8%</td>
<td>26.0%</td>
<td>19.2%</td>
</tr>
<tr>
<td>I will continue</td>
<td>54.4%</td>
<td>24.0%</td>
<td>21.6%</td>
</tr>
<tr>
<td>taking this</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>brand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of fun</td>
<td>51.8%</td>
<td>16.8%</td>
<td>31.4%</td>
</tr>
<tr>
<td>I can afford it</td>
<td>49.8%</td>
<td>24.0%</td>
<td>26.2%</td>
</tr>
<tr>
<td>Exceed expectations</td>
<td>44.8%</td>
<td>30.8%</td>
<td>24.4%</td>
</tr>
<tr>
<td>Very satisfied</td>
<td>43.4%</td>
<td>26.6%</td>
<td>30.0%</td>
</tr>
<tr>
<td>I like it</td>
<td>39.8%</td>
<td>32.0%</td>
<td>28.2%</td>
</tr>
<tr>
<td>No need to say</td>
<td>34.6%</td>
<td>36.0%</td>
<td>29.4%</td>
</tr>
<tr>
<td>more</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am big fan</td>
<td>31.0%</td>
<td>23.6%</td>
<td>45.4%</td>
</tr>
<tr>
<td>Average</td>
<td>50.68%</td>
<td>25.15%</td>
<td>24.17%</td>
</tr>
</tbody>
</table>

For negative comments, Table 6 shows that there is an average of 60.23% decrement in ratings between the original comments and their simulated versions. This indicates that participants found the simulated comments to have a higher intensity in negative sentiment than their original versions. For example, some participants rated “What a lie” as “Slightly Dislike” but rated “Whaaaaaaat a lie” as “Dislike” or “Strongly Dislike”. The pair of negative comments that had the largest decrease in ratings is “I don’t care” and “I don’t caaaaaaaare”, of which 69.6% of participants decreased their ratings for the simulated comments towards “Slightly Dislike”.

For example, some participants rated “I am very happy” as only “Slightly Like” but rated “I am very haaaaaappy” as “Like” or “Strongly Like”. The pair of positive comments that underwent the largest increase in ratings is “Really enjoy” and “Really eeeeeeeenjoy”, of which 62.2% of participants increased their ratings for the latter comment towards “Like”. Overall, approximately 31% to 62% of participants increased their “Like” rating for the simulated version.
These tools gave different scores due to their inability to identify the word with repeated letters. For instance, Sales Assistant is not able to detect the word “loooooove”. As shown in Table 7, there is a consistent and noticeable shift towards “Dislike” tendency in all three measures of central tendency for negative comments as compared to the shift towards “Like” tendency for positive comments. Among the three measures, the median is found to be the most reliable measurement because it measures the middle score for a set of data that has been sorted by magnitude, such as the ordinal Likert scale ranging from 1 to 7. Furthermore, the median is also less affected by outliers and skewed data. Therefore, when all three measures of central tendency are compared, the median displays the largest difference between some positive and negative comments that are significantly affected by repeated letter simulation. All of the negative comments experienced stronger “Dislike” tendency whereas only two-thirds of the positive comments experienced stronger “Like” tendency. This observation further confirms the earlier finding that letter repetition has a greater augmenting effect on negative comments compared to positive ones.

Table 6 - Rating Changes in Negative Comments

<table>
<thead>
<tr>
<th>Negative Comment</th>
<th>Rating increases from Original Comment to Simulated Comment</th>
<th>Rating decreases from Original Comment to Simulated Comment</th>
<th>Rating maintains from Original Comment to Simulated Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t care</td>
<td>11.4%</td>
<td>69.6%</td>
<td>19.0%</td>
</tr>
<tr>
<td>Some serious abuse</td>
<td>10.8%</td>
<td>67.2%</td>
<td>22.0%</td>
</tr>
<tr>
<td>Fine until it breaks</td>
<td>10.6%</td>
<td>66.2%</td>
<td>23.2%</td>
</tr>
<tr>
<td>Too much trouble</td>
<td>18.2%</td>
<td>65.6%</td>
<td>16.2%</td>
</tr>
<tr>
<td>Don’t come here to shop</td>
<td>11.2%</td>
<td>65.6%</td>
<td>23.2%</td>
</tr>
<tr>
<td>It is really annoying</td>
<td>19.0%</td>
<td>64.0%</td>
<td>17.0%</td>
</tr>
<tr>
<td>I have to worry</td>
<td>17.8%</td>
<td>62.6%</td>
<td>19.6%</td>
</tr>
<tr>
<td>Are you kidding me?</td>
<td>17.0%</td>
<td>59.8%</td>
<td>23.2%</td>
</tr>
<tr>
<td>What a lie</td>
<td>15.6%</td>
<td>58.8%</td>
<td>25.6%</td>
</tr>
<tr>
<td>Very disappointed</td>
<td>16.8%</td>
<td>57.2%</td>
<td>26.0%</td>
</tr>
<tr>
<td>Never, ever, never</td>
<td>20.0%</td>
<td>56.2%</td>
<td>23.8%</td>
</tr>
<tr>
<td>I did hit it well</td>
<td>19.8%</td>
<td>55.0%</td>
<td>25.2%</td>
</tr>
<tr>
<td>I hate it</td>
<td>13.4%</td>
<td>52.8%</td>
<td>33.8%</td>
</tr>
<tr>
<td>Totally fierce</td>
<td>16.0%</td>
<td>51.6%</td>
<td>32.4%</td>
</tr>
<tr>
<td>I boot it</td>
<td>21.8%</td>
<td>51.2%</td>
<td>27.0%</td>
</tr>
<tr>
<td>Average</td>
<td>15.96%</td>
<td>60.23%</td>
<td>23.81%</td>
</tr>
</tbody>
</table>

Overall, 50% to 70% of participants decreased their ratings for the simulated comments. To sum up, letter repetition has a stronger amplifying effect on the sentiment value of negative comments compared to positive ones.

Table 7 - Rating Changes in Positive Comments

<table>
<thead>
<tr>
<th>Positive changes (Higher “like” tendency)</th>
<th>Mode ratings</th>
<th>Positive</th>
<th>Median ratings</th>
<th>Mean ratings</th>
<th>Mode ratings</th>
<th>Negative</th>
<th>Median ratings</th>
<th>Mean ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Higher “dislike” tendency)</td>
<td></td>
<td>13</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No changes (No higher “like” or “dislike” tendency)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Total no. of comments significantly affected by its higher “like” or “dislike” tendency</td>
<td></td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

B. Sentiment Tools Analysis

Table 8 presents the results of the sentiment analysis of positive comments using sentiment tools. Overall, the results suggest that 41% of tested sentiment tools showed no difference in scores between the original comment and its simulated counterpart. Such results indicate that these tools do not detect any change in sentiment value for comments with repeated letters. In other words, 41% of these tested tools do not consider letter repetition as an indication of a change in their sentiment score. For instance, Sentaero (Tool 2) gave a 100% positive result for both comments. Similarly, Meaning cloud (Tool 3) and Repustate (Tool 5) respectively showed positive results.

The remaining 59% of the tools showed different sentiment scores between the original comment and its counterpart. However, many of these tools gave different scores due to their inability to identify the word with repeated letters. For example, Selasdia Intelligent Sales Assistant (Tool 1) gave an overall polarity to the comment by breaking down sentences and words. The comment “love our new tv” is marked as positive with the word “love” given a positive polarity. However, “looooooove our new tv” is given a neutral polarity with zero sentiments, indicating that Selasdia Intelligent Sales Assistant is not able to detect the word “looooooove”. Hence, the sentiment for this word changed from positive to neutral. Another example is Text sentiment analyser (Tool 6), which gives score breakdowns for each word. The word “love” has a sentiment score of 0.5, but when letter repetition is added, the word “looooooove” is not on its list of sentiment-by-word.

There is no sentiment value assigned to this word. Another tool, Twinword (Tool 25), assigned a positive score of 0.917220858 to “love” but zero score to “looooooove”. To sum up, although a majority of these tools gave different scores for the original comment and its simulated counterpart, the score difference is mainly due to the tools’ lack of ability to detect the word with repeated letters. Changes in sentiment score are due to fewer words in the text (when tools are unable to detect the word with repeated letters) and not because letter repetition carries a unique sentiment value. The only tool that is an exception to this is IMDB (Tool 14), which can spot the difference in a word between the original comment and simulated comment and increased the sentiment score for the word with repeated letters.
Table 8
Results of Sentiment Tools Comparison for Positive Text

<table>
<thead>
<tr>
<th>Tool</th>
<th>Scores for original comment</th>
<th>Scores for simulated comment</th>
<th>Tool</th>
<th>Scores for original comment</th>
<th>Scores for simulated comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P</td>
<td>P/N</td>
<td>17</td>
<td>P: +0.881</td>
<td>P: +0.877</td>
</tr>
<tr>
<td>2</td>
<td>P 100%</td>
<td>P 100%</td>
<td>18</td>
<td>P: 0.398052</td>
<td>P: 0.381061</td>
</tr>
<tr>
<td>3</td>
<td>P 98%</td>
<td>P 98%</td>
<td>19</td>
<td>P: 99.9658</td>
<td>P: 99.9376</td>
</tr>
<tr>
<td>4</td>
<td>P:0.922 NU:0.078</td>
<td>P:0.948 NU:0.052</td>
<td>20</td>
<td>P: 0.994583</td>
<td>P: 0.994583</td>
</tr>
<tr>
<td></td>
<td>Anger 0.002273</td>
<td>Disgust 0.00381</td>
<td></td>
<td>Anger 0.004542</td>
<td>Disgust 0.008128</td>
</tr>
<tr>
<td></td>
<td>Fear 0.00381</td>
<td>Fear 0.006955</td>
<td></td>
<td>Joy 0.954265</td>
<td>Joy 0.914345</td>
</tr>
<tr>
<td></td>
<td>Joy 0.954265</td>
<td>Sadness 0.02287</td>
<td></td>
<td>Sadness 0.031392</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>0.95</td>
<td>21</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>P:40%</td>
<td>P:38%</td>
<td>22</td>
<td>P:100%</td>
<td>0.67%</td>
</tr>
<tr>
<td>7</td>
<td>P:9</td>
<td>P:9</td>
<td>23</td>
<td>P:0.8</td>
<td>P:0.22898919857143</td>
</tr>
<tr>
<td>8</td>
<td>P:3</td>
<td>P:4</td>
<td>24</td>
<td>Don't care:0.931</td>
<td>Don't care:0.94</td>
</tr>
<tr>
<td></td>
<td>N: -1</td>
<td>N: -1</td>
<td></td>
<td>Happy:0.645</td>
<td>Happy:0.637</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inspired:0.878</td>
<td>Inspired:0.805</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sad:0.28</td>
<td>Sad:0.274</td>
</tr>
<tr>
<td>9</td>
<td>Overall:P</td>
<td>Overall:N</td>
<td>25</td>
<td>P: 0.4261340582143</td>
<td>P: 0.22808919857143</td>
</tr>
<tr>
<td></td>
<td>N:0.9</td>
<td>N:0.1</td>
<td></td>
<td>N:0.9%</td>
<td>N:0.9%</td>
</tr>
<tr>
<td>10</td>
<td>Overall:3.29</td>
<td>Overall:3.29</td>
<td>26</td>
<td>P:92%</td>
<td>N:8%</td>
</tr>
<tr>
<td>11</td>
<td>Overall: 1.75</td>
<td>Overall: 1.5</td>
<td>27</td>
<td>P:91%</td>
<td>N:9%</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2</td>
<td>28</td>
<td>Joy 0.95</td>
<td>Joy 0.91</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>6</td>
<td>29</td>
<td>Whole text is P</td>
<td>Whole text is P</td>
</tr>
<tr>
<td>14</td>
<td>Overall: 0.764</td>
<td>Overall: 0.905</td>
<td>30</td>
<td>Whole text is P</td>
<td>Whole text is P</td>
</tr>
<tr>
<td></td>
<td>“loo000ooove our new tv” is P</td>
<td>“loo000ooove our new tv” is</td>
<td></td>
<td></td>
<td>“loo000ooove our new tv” is</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“the tv is so light”</td>
<td>“the tv is so light”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“and thin it has a great picture and the”</td>
<td>“and thin it has a great picture and the”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“colors are true very very happy customer”</td>
<td>“colors are true very very happy customer”</td>
</tr>
<tr>
<td>15</td>
<td>P: 16.7</td>
<td>P: 12.5</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>P 100</td>
<td>P 100</td>
<td>32</td>
<td>Whole text is N</td>
<td>Whole text is P</td>
</tr>
</tbody>
</table>

Table 9 presents the results of the sentiment analysis of negative comments using sentiment tools. Overall, the results suggest that 53% of sentiment tools showed no difference in score between the original comment and simulated comment. Sentiment tools such as Selasdia Intelligent Sales Assistant (Tool 1), Sentaero (Tool 2), Meaning cloud (Tool 3), Repustate (Tool 5), ParallelDots (Tool 23), and others showed similar results for both original and simulated comments. For instance, Repustate and Alchemy Language Document Sentiment (Tool 20) gave negative sentiment scores of -0.95 and -0.894785 respectively to the original comment and the simulated comment. Other tools such as Alchemy Language Targeted Emotion (Tool 21) and DepecheMood (Tool 24) treated the word with repeated letters as a spelling mistake.

The remaining 47% of the tools showed different scores in the comments. However, this score difference is due to the tools’ inability to recognise the word with repeated letters. Interestingly, IMDB (Tool 14) differentiated “love” and “loooooove” in the positive comment, but it could not detect “хаааааааа” in the negative comment. Twinword (Tool 25) gave the word “hate” a negative sentiment score of -0.918459669 but no score for the word “хааааааа”. Additionally, some tools like Sentiment Viveka (Tool 19), Tone Analyzer (Tool 27), and Pythia Semantic Features (Tool 28) showed different scores for both comments without giving a breakdown or detailed analysis of the score. However, as the score became less negative for the simulated comment, it can be assumed that these tools are also unable to detect the word with repeated letters.

V. CONCLUSION

The current study examines the impact of letter repetition on perceived sentiment expression in online product reviews, as assessed by both individual participants and sentiment tools. Based on a collection of 30 online comments that were manually classified into positive or negative sentiments by 500 individual participants results revealed that letter repetition indeed affects readers’ perceived sentiment connotation of the comments. Letter repetition has a notably greater augmenting effect on negative comments than positive comments.

On the other hand, the results of sentiment tools suggest that many of them are unable to detect words with repeated letters. This indicates that developers should pay more attention to fine-tuning these tools in analysing the sentiment value of repeated letters. This study’s findings imply that automated social media analysis systems, such as sentiment analysis tools, should take into account letter repetition in social media messages for a more accurate and efficient analysis and extraction of opinions of consumers and other users in general. The study’s human-rated dataset will be made publically available with the paper under a creative commons license.

ACKNOWLEDGMENT

The authors would like to thank Paul Rayson for giving access to Wmatrix. We would like to give credits to the online freely available sentiment tools used in the analysis of this study. This study is partially funded by Universiti Sains Malaysia Research University Grant 1001/PKOMP/801402 and Sunway-Lancaster Grant SGSSL-FST-DCIS-0115-11.
Table 9

Results of Sentiment Tools Comparison for Negative Text

<table>
<thead>
<tr>
<th>Tool</th>
<th>Scores for original comment</th>
<th>Scores for simulated comment</th>
<th>Tool</th>
<th>Scores for original comment</th>
<th>Scores for simulated comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 N</td>
<td>N</td>
<td>N</td>
<td>17</td>
<td>NU: +0.411</td>
<td>N: -0.153</td>
</tr>
<tr>
<td>2 N</td>
<td>N 100%</td>
<td>N 100%</td>
<td>18</td>
<td>N: -0.7</td>
<td>N: -0.5</td>
</tr>
<tr>
<td>3 N</td>
<td>N 100%</td>
<td>N 100%</td>
<td>19</td>
<td>N: 99.9104</td>
<td>N: 73.0657</td>
</tr>
<tr>
<td>4 N</td>
<td>N:0.938</td>
<td>N:0.941</td>
<td>20</td>
<td>N: -0.894785</td>
<td>N: -0.894785</td>
</tr>
<tr>
<td></td>
<td>NU:0.062</td>
<td>NU:0.059</td>
<td></td>
<td>Anger 0.639612</td>
<td>Anger 0.480416</td>
</tr>
<tr>
<td>5 N</td>
<td>N: -0.95</td>
<td>N: -0.95</td>
<td>21</td>
<td>Fear 0.217572</td>
<td>Fear 0.254317</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Joy 0.186902</td>
<td>Joy 0.041167</td>
</tr>
<tr>
<td>6 N</td>
<td>N: 70%</td>
<td>N: 50%</td>
<td>22</td>
<td>N: 50%</td>
<td>N: 50%</td>
</tr>
<tr>
<td>7 N</td>
<td>N: 10</td>
<td>N: 10</td>
<td>23</td>
<td>N:5</td>
<td>N:5</td>
</tr>
<tr>
<td>8 P</td>
<td>P:1</td>
<td>N:4</td>
<td>24</td>
<td>Dots: 0.386</td>
<td>Dots: 0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Happy: 0.649</td>
<td>Happy: 0.762</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inspired: 0.605</td>
<td>Inspired: 0.629</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sad: 0.548</td>
<td>Sad: 0.661</td>
</tr>
<tr>
<td>9 Overall:N</td>
<td>Overall:P</td>
<td>P:0.1</td>
<td>25</td>
<td>N: -0.20522543125</td>
<td>N: -0.1381325554</td>
</tr>
<tr>
<td></td>
<td>P:0.2</td>
<td>N:0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Overall: -4.976</td>
<td>Overall: -4.976</td>
<td>26</td>
<td>P:3%</td>
<td>P:2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N:97%</td>
<td>N:0.8</td>
<td></td>
<td>N:98%</td>
<td></td>
</tr>
<tr>
<td>11 Overall: -0.125</td>
<td>Overall: -0.125</td>
<td>27</td>
<td>Anger 0.64</td>
<td>Analytical 0.6</td>
<td></td>
</tr>
<tr>
<td>12 -2</td>
<td>-1</td>
<td>28</td>
<td>Whole text is N</td>
<td>Whole text is N</td>
<td></td>
</tr>
<tr>
<td>13 -1</td>
<td>-1</td>
<td>29</td>
<td>“i hated this iron because the steam comes out in all the wrong places” is N</td>
<td>“i hated this iron because the steam comes out in all the wrong places” is N</td>
<td></td>
</tr>
<tr>
<td>14 Overall: -0.0302</td>
<td>Overall: 0.0918</td>
<td>30</td>
<td>Whole text is N</td>
<td>Whole text is N</td>
<td></td>
</tr>
<tr>
<td>15 N</td>
<td>N:10.0</td>
<td>N:5</td>
<td>31</td>
<td>Whole text is N</td>
<td>Whole text is N</td>
</tr>
<tr>
<td>16 N</td>
<td>N:100</td>
<td>N:100</td>
<td>32</td>
<td>Whole text is N</td>
<td>Whole text is N</td>
</tr>
</tbody>
</table>

REFERENCES


