How is tinnitus represented in the United States newspaper media and in Facebook pages?

*Vinaya Manchaiah¹,²,³, *Pierre Ratinaud⁴ &*Gerhard Andersson⁵,⁶

1. Department of Speech and Hearing Sciences, Lamar University, Beaumont, Texas, USA
2. Department of Speech and Hearing, School of Allied Health Sciences, Manipal University, Manipal, Karnataka, India
3. Audiology India, Mysore, Karnataka, India
4. LERASS Laboratory, University of Toulouse, Toulouse, France
5. Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden
6. Department of Clinical Neuroscience, Division of Psychiatry, Karolinska Institute, Stockholm, Sweden

*Corresponding author

Email: vinaya.manchaiah@lamar.edu (VM)

*These authors contributed equally to this work.

**Corresponding author:** Dr Vinaya Manchaiah

**Communication address:** Department of Speech and Hearing Sciences,
Lamar University, Beaumont, Texas 77710, USA

**Email:** vinaya.manchaiah@lamar.edu

**Tel:** +1 (409) 880 8927
Abstract

Background: When people with health conditions begin to manage their health issues, one important issue that emerges is the question as to what exactly do they do with the information that they have obtained through various sources (e.g., news media, social media, health professionals, friends and family). The information they gather helps form their opinions, and will also to some degree influence their attitudes towards managing their condition.

Objective: The current study was aimed at understanding how tinnitus is represented in the United States newspaper media and in Facebook pages (i.e., social media) using text pattern analysis.

Method: This was a cross-sectional study based upon secondary analyses of publicly available data. The two data sets (i.e., text corpuses) analyzed in this study were generated from: (1) U.S. newspaper media during 1980-2017 (downloaded from the database U.S. Major Dailies by ProQuest); and (2) Facebook pages during 2010-2016. The text corpuses were analyzed using the Iramuteq software using cluster analysis and Chi square tests.

Results: The newspaper data set had 432 articles. The cluster analysis resulted in five clusters, which were named as follows: (1) brain stimulation (26.2%); (2) symptoms (13.5%); (3) coping (19.8%); (4) social support (24.2%); and (5) treatment innovation (16.4%). A time series analysis of clusters indicated a change in the pattern of information presented in newspaper media during 1980-2017 (e.g., more emphasis on cluster 5, focusing on treatment inventions). The Facebook data set had 1,569 texts. The cluster analysis resulted in five clusters, which were named as: (1)
diagnosis (21.9%); (2) cause (4.1%); (3) research and development (13.6%); (4) social support (18.8%); (5) challenges (11.1%); (6) symptoms (21.4%); and (7) coping (9.2%). A time series analysis of clusters indicated no change in information presented in Facebook pages on tinnitus during 2011-2016.

**Conclusions:** The study highlights the specific aspects about tinnitus that the U.S. newspaper media and Facebook pages focus on, and also how these aspects change over time. These findings can help healthcare providers better understand the presuppositions that tinnitus patients may have. More importantly, the findings can help public health experts and health communication experts in tailoring health information about tinnitus in order to promote self-management, as well as assisting in appropriate choices of treatment for tinnitus sufferers.

**Key Words**

Tinnitus, Chronic condition, Health communication, Health information, Media, News media, Social media

**Introduction**

Tinnitus can be defined as the conscious perception of an auditory sensation in the absence of a corresponding external stimulus [1]. It is a common medical symptom, which can be debilitating to some individuals. Risk factors for tinnitus include: hearing loss, use of ototoxic drugs, exposure to loud noise, and head injury. Various kinds of treatments are being offered to help tinnitus sufferers, although no effective curative drug treatment exists. Some main treatment and management approaches for tinnitus include: sound therapy (i.e., external noise to counteract the perception of tinnitus), hearing aids, and counseling [1,2]. However, Cognitive Behavioural
Therapy (CBT) has the most evidence-base for the management of the distress associated with tinnitus [2].

Medical therapy alone is not sufficient to address the distress caused by various chronic symptoms and conditions [3], which can also be true for tinnitus. For this reason, there is a move towards developing self-management initiatives in which individuals with a health condition take responsibility and ownership in managing their health conditions and symptoms [4]. Self-management can facilitate acceptance, coping, learning to live well with chronic health conditions despite the adverse consequences of such conditions and symptoms, and more importantly can result in improved health outcomes [5]. However, when people with health conditions start to manage their health issues, a significant factor that determines how effectively they manage their issues involves the type and quality of information that they obtain through various sources (e.g., news media, social media, health professionals, friends and family). The information they gather helps form their opinions, and will also to some degree influence their attitudes towards managing their condition. Literature from health psychology, particularly the Self-Regulation Model of Illness (SRMI) which discusses the “common-sense model of illness representation,” provides a framework for understanding how individual symptoms and emotions during a health threat or diagnosis influence perception of illness and hence guide subsequent behavior [6,7]. In addition to symptoms and emotions experienced, individual opinions are also formed as a result of exposure to information from external sources. For this reason, health communication is an important element of facilitating self-management in chronic disease management from the public health viewpoint [8].
The ubiquitous nature of mass media, in particular news media, makes it one of the most powerful sources of influence on specific issues (e.g., climate change, political party), including health issues [9]. News media plays an important role in determining the content of our thoughts and opinions [10]. For example, the more coverage a specific topic receives on a particular news media, the more likely people will take it into serious consideration. On the other hand, a topic that does not get much attention in the news media, although important, may not receive much public attention. Although traditionally society has relied heavily on the news media, in recent years social media has radically transformed the way in which people communicate and gather information around the world [11]. Social media has changed the news business on a fundamental level [12]. For example, in social media content, providers are the first ones to report any given particular news, whereas editors and journalists are involved in content development and decision making of what appears in the news media. People use social media to share and exchange perspectives and opinions about social, cultural, economic, and religious aspects; furthermore, there is a growing trend to also exchange information about health issues and concerns [13]. In keeping with this recent development, there is a critical need to better understand the influence of social media on health behavior and outcomes [14].

While the media can help place specific health issues on a local and national agenda, there may be a discrepancy in what issues are covered and in their social and economic impact on society. For instance, we can assume that the issues more frequently raised in the media are those which are more likely to draw more attention and readership. Moreover, Hartz et al. reports that there is a growing tension between media reporters, scientists, and health professionals [15]. This may be
because of discrepancies in understanding specific health issues, and also because of the importance placed on specific health issues by these key stakeholders [15].

There is some literature on how the deaf population is represented in the media [16, 17]. Some recent literature also focuses on newspapers’ representation of workers with hearing loss [18], and the use of social media in the hearing aid community [19]. We recognize that both mass media and social media can have much influence on people's knowledge and in the shaping of individual attitudes [14]. While this is of importance, we were unable to find any studies focusing specifically on what information is presented about tinnitus in the news media, and also in social media.

The current study was aimed at understanding how tinnitus is represented in the United States newspaper media and in Facebook pages (i.e., social media) using text pattern analysis.

**Method**

*Study Design*

The study used a cross-sectional design based upon secondary analyses of publicly available data. The two data sets analyzed in this study were generated from: (1) U.S. newspaper media; and (2) Facebook pages. The study did not require ethical approval as the data was gathered from publicly available sources.

*Data Extraction*
To develop the newspaper media text corpus (i.e., large and structured set of texts) we first explored the databases with the newspaper collection available at Lamar University, U.S. Major Dailies by ProQuest (http://www.proquest.com/products-services/US-Major-Dailies.html) was the database that had the largest newspaper collection. We then searched for articles related to “tinnitus” in this database between the years 1980 till 2017 and downloaded the results as text corpus. A python script was written to convert the text corpus to a format that was needed for data analysis and to preserve the meta-data (i.e., newspaper name, year of publication).

A different python script was written to extract posts with Facebook pages dealing with tinnitus (during 2010 to 2016). Twenty Facebook pages with more than 100 likes were identified, and the postings were downloaded as a text corpus. Only the publicly available Facebook pages (not personal pages) were included in the data extraction, thus maintaining the anonymity of the responses and no personally identifiable information was included. Also, no individual data set was discussed in the paper, again maintaining the anonymity of the data. Considering the minimal or no potential risk to individual participants, no IRB approval was required [20].

It is important to note that the data extraction was limited to what data was available publicly (i.e., newspaper data during 1980-2017; and Facebook posts during 2010-2016).

**Data Analysis**

The text corpuses were analyzed using the Iramuteq software [21], which is open source software. This software can perform various types of analyses on text data.
The text corpus is composed of multiple newspaper articles. The software treats each of these articles as text (i.e., it’s the first unit). The first step of the analysis is to segment each article (i.e., text) into smaller units called text segments (i.e., each text is split into multiple text segments based on criteria of size and punctuation). The split of the text into segments decreases the granularity of the units and thus makes it possible to increase the precision of certain analysis, in particular the search of themes within the text corpus. The goal is to create segments of consistent size while trying to maintain the natural segmentation of texts marked by punctuation. The segmentation process is based on a cut-off criteria that weighs the segment size by punctuation. The procedure does not fully respect the parameterized segment size if a strong punctuation (e.g., the point, the question mark and the exclamation point) is present within a 15% margin of the planned cutoff. In the next step, the text corpus is lemmatized (i.e., words are sorted by grouping inflected or variant forms of the same word) to their simplest forms, which are called lemmas (i.e., group of words in a single form). Also, the software makes distinctions between “full words” (e.g., verbs, noun, adjectives, and adverbs) and “tool words” (e.g., pronoun, determent, and useful verbs such as – to be and to have). This distinction is done so that only full words are included in the main analysis. These steps are necessary in order to convert the large corpus to easily a more manageable data set that is essential for further analysis. In order to specifically analyze the text that is closely related to “tinnitus,” the text segments related to this object were extracted and new corpuses were formed. It is important to note that the expression “tinnitus” was inside each of the text segments extracted. The sub-corpus with more directed text segments were used for all further analysis.

This was followed by cluster analysis made with the Reinert's method used for textual data
This method has existed since 1983, and has been used in automated text pattern analysis [22-24]. The Reinert method is older and it is significant that it has been used more than topic models in French media studies. The method has the advantage to facilitate the study of links between topics and the metadata associated with the text (e.g., dates to study chronology). Also, a recent comparison between the Reinert method and LDA show that topics from the Iramuteq software which uses Reinert method (i.e., a divisive hierarchical clustering) are more accurate [25]. While the traditional qualitative methods (e.g., content analysis, thematic analysis) can provide in-depth understanding of the data, the automated text pattern analysis using the cluster analysis can provide a broader understanding of the data [22]. However, the automated analysis has the advantage of analyzing large amounts of data (i.e., big data) as used in this study, whereas using the traditional qualitative methods for analyzing large quantities of data is not feasible.

In this analysis initially, the program builds a binary matrix with text segments in rows and full words in columns. Afterwards, it performs hierarchical divisive clustering based on a series of bi-partitions made with correspondent analysis. At each step of the classification, the bigger remaining cluster is cut into two parts by computing the information extracted while cutting after each line of the matrix along the first factor of the correspondent analysis. The remaining cut is the one that maximizes the information extracted. In a second step, each line is reversed from one cluster to the other. If this reversion increases the information extracted, it is kept. This step loops until no inversion increases the information extracted for the whole table. This cluster analysis groups the text segments based on co-occurrence of lemmas. Each of the clusters aims to be homogeneous (regrouping text segments with the common pattern of lemmas). Also,
clusters have to be as heterogeneous as possible between them (pattern of lemmas between
groups should be as different as possible). The results are presented in a dendrogram (see Figures
1 and 4) that characterizes the clustering. For each cluster, the program computes profiles of
lemmas, which are overrepresented (i.e., significantly in a higher proportion within the cluster
when compared to the rest of the text corpus based on chi square analysis). Finally, the same text
corpus was subjected to a time-series analysis using the meta-data. For example, in this corpus,
we analyzed how the patterns of clusters change over time. Figures 2 and 5 present proportions
of cluster for each year. In this width of the bar, it is proportional to the number of text segments
each year (i.e., the higher the width number of text segments) and the height of the clusters
(indicated in different colors) represents the frequency of text segments within clusters. Figures 3
and 6 present what clusters are significantly over represented (i.e., based on chi square analysis)
in each year during 1990 to 2016. In this width of the bar, it is proportional to the number of text
segments each year, and the height of the bar represents the size of clusters. For example, Figure
3 shows that cluster 5 is over represented during the years 2013-17.

Results

Newspaper data

Context analysis: Initial analysis text corpus had 433 texts (i.e., each article from a newspaper is
considered a text), 9,176 text segments (i.e., each text is split into multiple text segments based
on criteria of size and punctuation) and 309,524 occurrences or tokens (i.e., number of words).
After extraction of a text segment related to “tinnitus,” the text corps was reduced to 432 texts,
2,173 text segments, and 79,684 occurrences or tokens.
Table 1 provides information about the newspaper source and the frequency of texts from each source. Most of the texts are from the U.S. Fed News Services (30.79%), followed by The Washington Post (16.67%), Targeted News Service (13.19%), and the Chicago Tribune (10.19%). Table 2 provides information about the frequency of texts related to tinnitus over time. The frequency of texts published about tinnitus increases over time. The reason for this can be due to an increase in attention and publication of articles on tinnitus and/or the number of newspapers publishing articles about tinnitus. However, it is hard to distinguish the relative contribution regarding these two reasons due to the increase in the frequency of texts related to tinnitus over time.

<Table 1 near here>
<Table 2 near here>

Content analysis: Figure 1 provides results of cluster analysis on the newspaper data. The cluster analysis resulted in 5 main clusters that were named based on their characteristics. Cluster 1 was the largest cluster, which included 26.2% of texts, and mainly focused on aspects related to auditory and electrical stimulation in the brain as a cure for tinnitus, which was named as a brain stimulation cluster. Cluster 2 was the smallest cluster that included 13.5% of texts, mainly focusing on the symptoms of tinnitus such as buzzing, ringing, hiss, and was named as a symptoms cluster. Cluster 3 included 19.8% of texts and included aspects related to psychological problems (e.g., anxiety), relaxation, and coping, and hence was named as a coping cluster. Cluster 4 was the second largest cluster and included 24.2% of the texts. This cluster had elements related to health and social support from government and charity organizations, and
hence was named as a *social support* cluster. Cluster 5 included 16.4% of texts, and was focused on new and innovative treatment methods, and hence was named as a *treatment innovation* cluster. Table 3 provides an example of text segments that typically represent each of these five clusters.

*Analysis of trends over time:* Figure 2 provides a chronological bar, and Figure 3 a chronological bar with chi-square for tinnitus newspaper data. These two figures provide information on how the information presented in newspaper media changes over time, and helps us understand media trends. For example in Figure 2, it is evident that the cluster 5 (i.e., treatment innovation) was not discussed in newspapers before the year 2006, but the number of texts increased substantially from 2006 to 2017. The chi square analysis (see Figure 3) shows that some clusters are more likely to occur on some timescale. For example, cluster 5 (i.e., treatment innovation) was statistically significant and more frequent during the years 2013 to 2017, whereas the cluster 3 (i.e., coping) was statistically significant and more frequent during the years 2000 to 2005. Overall, the time-series analysis of clusters indicated changes in the pattern of information presented in newspaper media during the years 1980 to 2017.
**Facebook pages data**

**Context analysis:** Facebook pages text corpus had 1,569 texts, 2,747 text segments, and 78,218 occurrences or tokens. We have removed the URL, and also texts that contained only the URL before analysis.

**Content analysis:** Figure 4 provides results of cluster analysis on the Facebook pages data, which identifies seven main clusters. Cluster 1 (21.9%) was the largest and had words related to ear disorders and its diagnosis, and hence was named as *diagnosis* cluster. Cluster 2 (4.1%) was the smallest and was focused on causes and trigger factors of tinnitus (e.g., tooth problems, neuropathy) and was named as the *cause* cluster. Cluster 3 included 13.6% texts focusing on studies that focused on tinnitus and was named as the *research and development* cluster. Cluster 4 (18.8%) had elements related to social support from various sources, and was named as the *social support* cluster. Cluster 5 was named as challenges and had 11.1% of texts. Cluster 6 (21.4%) focused on symptoms of tinnitus such as buzzing, ringing, hiss, and was named as the *symptoms* cluster. Cluster 7 (9.2%) was focused on relaxation and coping, and hence was named as the *coping* cluster. Table 4 provides an example of text segments that typically represent each of these seven clusters.

<Figure 4 near here>

<Table 4 near here>

**Analysis of trends over time:** Figure 5 provides a chronological bar and Figure 6 a chronological bar with chi-square for tinnitus Facebook pages data. As indicated in Figure 5, there is not much
variation in what clusters are discussed in each year over time. The chi square analysis (see Figure 6) shows that cluster 1 (i.e., diagnosis), cluster 4 (i.e., social support), and cluster 5 (i.e., challenges) occurred more significantly during the years of 2013, 2014 and 2015 respectively.

Discussion

The media has played a prominent role in the dissemination of health information [9]. The current study explores the way tinnitus is represented in the media, specifically in the United States newspaper media and in Facebook pages using text pattern analysis. Overall, the study results suggest that newspaper media and Facebook pages offering information on tinnitus focus on various elements including: symptoms, coping, diagnosis, social support, research and development, and treatment innovations.

The analysis of text corpus extracted from the U.S. newspaper media suggested that the information disseminated via newspaper media focuses mainly on two elements: (1) new developments in treatments (i.e., brain stimulation, treatment innovations); and (2) disease information (i.e., symptoms, coping and social support). It is interesting to note that there is a fairly equal spread of information among all of these elements. Also, the analysis trends over time regarding information indicated a change in patterns of information presented in the newspaper media during 1980-2017. For example, there is an increasing emphasis of information focusing on treatment inventions (cluster 5 in Figure 2), whereas the information on coping
appears to be steady over the years (cluster 3 in Figure 2). We were unable to find previous studies on newspaper media on tinnitus in order to compare current study results. However, a recent study explored the Canadian newspaper representation of hearing impaired workers [18]. The study results reveal that the media focuses on a wide range of issues (e.g., hearing impaired people taking action to find employment, and being determined to find success in sought after positions of employment).

The analysis of content in Facebook pages related to tinnitus in the current study suggests that tinnitus sufferers use social media for various purposes, including gaining symptoms and diagnostic information, social support, learning to cope, and also to obtain information about research in this area. It is important to note that nearly half of the discussions in Facebook pages were related to diagnosis and symptoms (i.e., 43.3%), suggesting that this platform is used by tinnitus sufferers for self-assessment of their condition. Facebook pages are fairly recent when compared to news media, and we explored Facebook pages information only during the years 2011 to 2016. Time series analysis showed no change in patterns of information during this time, unlike newspaper media. The current study results relate well to previous studies on social media and other health conditions; for example, a recent study was done on the hearing aid community who used social media sources for advice and support, information sharing, and service-related information [19]. Analysis of information related to the level of usefulness of diabetes foot care-related Facebook groups showed a significant association with related posts [26]. Social media is also reported to be useful in creating indicators of the food environment that are associated with area-level mortality, health behaviors, and chronic conditions [27].
One would expect that the media would provide publicity for an intervention that is evidence-based. While there are various treatments and/or management strategies available for tinnitus sufferers, psychological interventions such as cognitive behavioral therapy have the best evidence base for alleviating tinnitus distress [2]. However, it is surprising to see that there is a lack of publicity regarding psychological intervention, which becomes clear from the analysis of the information in the media. This is perhaps one of the reasons contributing to a relatively low uptake of evidence-based counseling interventions by tinnitus sufferers in the U.S. population [28].

Overall, there is growing literature about the importance of health communication in chronic condition management, particularly concerning the role of mass media and social media in forming individuals' opinions, and its bearing upon health behaviors. For this reason, it is important for healthcare professionals to be aware of what kind of information is being provided by the media on specific health conditions such as hearing loss and/or tinnitus.

**Study Implications**

The current study has several practical implications. As the media plays an important role in formulating people’s knowledge and opinions [14], it is important for healthcare professionals to be aware of what kind of information is being disseminated via the media regarding specific health conditions. This, we believe, may assist healthcare professionals in better understanding the presuppositions patients may have about various health aspects. It is suggested that when used wisely and prudently, social media can help promote individual and public health [29]. Also, the stakeholders which include healthcare professionals, patient organizations, and public health specialists can use the media to ensure that appropriate and accurate information is being
disseminated in order to promote positive health behavior and health outcomes [30]. However, the first step in this process is to understand the trends or patterns of information being presented in the media as explored in this study.

Study Limitations and Future Directions

The current study is the first of its kind in the area of tinnitus. However, it has a few limitations. First, the text pattern analysis using software helps analyze big data, and although the analysis provides more of a macro view of the data, it may only provide a superficial understanding of its content. The automated data analysis may help us understand 'what' aspects of tinnitus are represented in the media, rather than ‘how’ tinnitus and these aspects are represented. For example, questions such as, “Is brain stimulation portrayed positively or negatively in the media?”, Or, “Is tinnitus portrayed as a curable condition?” And finally “How are the causes of tinnitus portrayed?” cannot be answered through this kind of analysis. Hence, detailed qualitative analysis may be useful in future studies. Second, the newspaper data extracted from the ProQuest database may not be a true representation of all the U.S. newspapers. While ProQuest has some of the most widely circulated newspapers in the U.S., using a different search engine may have potentially resulted in a different corpus and perhaps different results patterns. Furthermore, some selection bias may have occurred while choosing Facebook pages as we only included 20 pages with 100 "likes". Different criteria may have resulted in a different text corpus. The current study mainly focuses on the patterns of information found in newspapers and in Facebook pages, but not about how appropriate and/or accurate the information actually is. Hence, future studies should explore the appropriateness and accuracy of content. Also, future studies on automated text pattern analysis should also aim to validate the results using predictive
validity methods which validates the data looking at the identified trends over time [31]. It would be interesting to study how the change over time may have occurred as a result of other metadata (e.g., Newspaper or Facebook group), which was not possible in the current analysis due to limitations with software. Future studies can also consider studying the commonalities and differences among different types of media (e.g., newspaper media, social media) using automated text analysis through more powerful software. Moreover, there is a great need to explore the role of the media on health behaviors, specifically in the tinnitus population.

Conclusions

The current study explored how tinnitus is represented in the media, specifically in the United States newspaper media and in Facebook pages. The information in the newspapers regarding tinnitus was mainly about: brain stimulation; symptoms; coping; social support; and treatment innovations. Time series analysis showed that there were some changes in the patterns of information presented in newspaper media during 1990-2017. The information in Facebook pages about tinnitus was mainly related to: diagnosis; cause; research and development; social support; challenges; symptoms; and coping. However, no changes in the patterns of information presented in Facebook pages during 2011-2016 were noted. These findings can help clinicians to better understand presuppositions tinnitus patients may have in regards to their condition. In addition, the findings can also be of interest to public health and health communication experts, allowing them to tailor health information about tinnitus in order to promote self-management and appropriate treatment choices for tinnitus sufferers.

Acknowledgements
The authors would like to acknowledge Gayatri Bhatta and Amit Kumar for help with data
eextraction from the U.S Newspaper database and Facebook pages and Prof. William Harn for
initial thoughts about text pattern analysis.

References

analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus
there time for management of patients with chronic diseases in primary care? *Ann Fam
illness. *Am Fam Physician*, 72(8), 1503-1510.
representation: Theoretical and practical considerations. *Journal of Social Distress in the
Homeless*, 5, 11–38.
8. Rimal RN, Lapinski MK. (2009). Why health communication is important in public
are the roles and responsibilities of the media in disseminating health information? *PLoS
Med* 2(7): e215. https://doi.org/10.1371/journal.pmed.0020215


research: Common concerns and practical considerations. *CyberpsycholBehavSocNetw*, 16(9), 708-713.


## Tables

**Table 1**: Frequency and percentage of articles containing at least one text-segment related to tinnitus in the sub-corpus from different newspapers

<table>
<thead>
<tr>
<th>Newspaper</th>
<th>Frequency of articles</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago Tribune</td>
<td></td>
<td>46</td>
<td>10.65</td>
</tr>
<tr>
<td>Farm Weekly</td>
<td></td>
<td>3</td>
<td>0.69</td>
</tr>
<tr>
<td>Investors Business daily</td>
<td></td>
<td>8</td>
<td>1.85</td>
</tr>
<tr>
<td>Journal of Commerce</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>Journal Record</td>
<td></td>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>Los Angeles Times</td>
<td></td>
<td>25</td>
<td>5.79</td>
</tr>
<tr>
<td>Marine Corps Times</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>Miami Daily Business Review</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>Missouri Lawyers Media</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>NASDAQ OMX’s News Release Distribution Channel</td>
<td></td>
<td>30</td>
<td>6.94</td>
</tr>
<tr>
<td>New York Times</td>
<td></td>
<td>27</td>
<td>6.25</td>
</tr>
<tr>
<td>Roll Call</td>
<td></td>
<td>2</td>
<td>0.36</td>
</tr>
<tr>
<td>Targeted News Service</td>
<td></td>
<td>57</td>
<td>13.19</td>
</tr>
<tr>
<td>The Village Voice</td>
<td></td>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>The Washington Post</td>
<td></td>
<td>72</td>
<td>16.67</td>
</tr>
<tr>
<td>The Weekly Times</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>US Fed News Services</td>
<td></td>
<td>133</td>
<td>30.79</td>
</tr>
<tr>
<td>Wall Street Journal</td>
<td></td>
<td>20</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>432</td>
<td>100</td>
</tr>
</tbody>
</table>

**Table 2**: Frequency and percentage of articles containing at least one text-segment related to tinnitus in the sub-corpus based on timescales from different newspapers

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency of articles</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td>3</td>
<td>0.69</td>
</tr>
<tr>
<td>Year</td>
<td>Count</td>
<td>TF-IDF</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>3</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>3</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>4</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>1</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>2</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>3</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>4</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>3</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>1</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>10</td>
<td>2.31</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>7</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>7</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>3</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>6</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>7</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>14</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>43</td>
<td>9.95</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>6</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>21</td>
<td>4.86</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>18</td>
<td>4.17</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>14</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>24</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>46</td>
<td>10.65</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>28</td>
<td>6.48</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>20</td>
<td>4.63</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>42</td>
<td>9.72</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>41</td>
<td>9.49</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>33</td>
<td>7.64</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>8</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>432</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: Example of a text segment for each cluster in newspaper media text corpus**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Example of a text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1: Brain</td>
<td>Past research has shown that the severity of chronic pain and tinnitus is tied to the degree of plasticity in the brain's cortex as a previous study showed that repeatedly pairing</td>
</tr>
</tbody>
</table>


stimulation with electrical stimulation of
a brain structure called nucleus basalis generates powerful and long lasting changes in cortical organization

Cluster 2: commonly called ringing in the ears but for millions of people who suffer tinnitus
the sound that plagues them actually may be a buzz hiss chirp or high pitched whistle

Cluster 3: playing soft background music or other white noise can be quite effective
in masking the noise and short term use
of antidepressant medications sometimes helps and often a hearing aid helps suppress tinnitus

Cluster 4: a senior member of the senate armed services committee sasc
recently received the senate legislative champion award by
the american tinnitus association ATA for his dedicated efforts to include
tinnitus research and funding to the annual national defense authorization act

Cluster 5: as of today neither a universal standard of care for acute inner ear tinnitus nor a
truly proven and effective treatment method is available about keyzilen tm am
101 keyzilen tm is
a small molecule n methyl d aspartate nmda receptor antagonist formulated in
a biocompatible gel for intratympanic injection

Table 4: Example of a text segment for each cluster in Facebook pages text corpus

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Example of a text segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1: Diagnosis</td>
<td>went to an ent doctor 2 weeks ago he checked my eustachian tubes because my ears were feeling a bit full that's all no ringing no pain he placed a probe up my nose but first sprayed 2 pumps of lidocaine anesthetic</td>
</tr>
<tr>
<td>Cluster 2: Causes</td>
<td>i have tmj temporo mandibular joint which has left me with very limited jaw movement only open mouth to less than an inch between teeth,</td>
</tr>
<tr>
<td>Cluster 3: Research and development</td>
<td>have been referred to hospital but up to 2 year waiting time. Tinnitus can be caused by tmj</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cluster 4: Social support</td>
<td>Gohearing we provide tinnitus and hearing therapy and treatment for bay area and canada providing tinnituslab pax100 and tinnitusreater sound conditioning programs which does not cost thousands. Affordable effective programs and treatment, check us out at www go hearing com</td>
</tr>
<tr>
<td>Cluster 5: Challenges</td>
<td>greeting please share our group page in your page its some time not about money we want to share and spread love to all those who are suffer from tinnitus. Please do support us, we need your help</td>
</tr>
<tr>
<td>Cluster 6: Symptoms</td>
<td>it was torture and a lot of bad thoughts were running through my mind i used sleep machines we think it was from coming off a drug serequel since jan and feb this year i have the ringing 3 to 4 times a week</td>
</tr>
<tr>
<td>Cluster 7: Coping</td>
<td>complete cure to tinnitus i had tinnitus in both ears for fifteen years with a high pitched two tone sound the noises are constant and have learned to ignore the ringing later another sound was added</td>
</tr>
<tr>
<td></td>
<td>can a thyroid effect your tinnitus n my thyroid is coming out soon i am hoping the hissing in my ear will subside a bit i had to sleep last night listening to white noise for the first time in years usually i listen to the radio playing music</td>
</tr>
</tbody>
</table>

**Figure Legends**

**Figure 1:** Dendrogram (i.e., classification of clusters), size of clusters as percentage of the text segments, and over-represented words in each cluster in tinnitus newspaper corpus.

(Note: The words are ordered by chi square value with words at the bottom having a lower value).
Figure 2: Chronological bar showing proportion of each cluster for each year in tinnitus U.S. newspaper media corpus (Note: Width of the bar is proportional to the number of text segments each year, and the height of the clusters represents the frequency of text segments within clusters).

Figure 3: Chronological bar based on chi square analysis showing proportion of each cluster for each year in tinnitus U.S. newspaper media corpus (Note: Width of the bar is proportional to the number of text segments each year, and the height of the bar represents the size of clusters).

Figure 4: Dendrogram (i.e., classification of clusters), size of clusters as percentage of the text segments, and over-represented words in each cluster in tinnitus on Facebook pages corpus. (Note: The words are ordered by chi square value with words at the bottom having a lower value).

Figure 5: Chronological bar showing proportion of each cluster for each year on tinnitus Facebook pages corpus (Note: Width of the bar is proportional to the number of text segments each year, and the height of the clusters represents the frequency of text segments within clusters).

Figure 6: Chronological bar based on chi square analysis showing the proportion of each cluster for each year on tinnitus Facebook pages corpus (Note: Width of the bar is
proportional to the number of text segments each year, and the height of the bar represents the size of clusters).