Acta Acustica 2024, **8**, 2

https://doi.org/10.1051/aacus/2023067

Available online at:
https://acta-acustica.edpsciences.org

LETTER TO THE EDITOR OPEN ACCESS

An experiment on an automated literature survey of data-driven speech enhancement methods

Arthur dos Santos[1,*], Jayr Pereira[2], Rodrigo Nogueira[2], Bruno Masiero[1], Shiva Sander Tavallaey[3,4], and Elias Zea[4]

[1] Communication Acoustics Lab, School of Electrical and Computer Engineering, Universidade Estadual de Campinas, Campinas – SP 13083-970, Brazil
[2] NeuralMind, Campinas – SP, 13083-898, Brazil
[3] ABB Corporate Research, SE-72226 Västerås, Sweden
[4] The Marcus Wallenberg Laboratory for Sound and Vibration Research, Department of Engineering Mechanics, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden

Received 1 November 2023, Accepted 8 December 2023

Abstract – The increasing number of scientific publications in acoustics, in general, presents difficulties in conducting traditional literature surveys. This work explores the use of a generative pre-trained transformer (GPT) model to automate a literature survey of 117 articles on data-driven speech enhancement methods. The main objective is to evaluate the capabilities and limitations of the model in providing accurate responses to specific queries about the papers selected from a reference human-based survey. While we see great potential to automate literature surveys in acoustics, improvements are needed to address technical questions more clearly and accurately.

Keywords: Speech enhancement methods, Data-driven acoustics, Literature survey, Natural language processing, Large language models

1 Introduction

A recent study has shown an increasing publication rate after analyzing 45 million scientific articles produced in the past six decades [1]. In the context of applications of data-driven methods in acoustics alone, as shown in the Scopus[1] search in Figure 1, the number of articles in the first half of 2023 had exceeded the total number of articles in the entire year of 2019. Given this growth in the literature, the acoustics community faces the limitations of traditional survey methods. At the same time, the remarkable advancements in the field of natural language processing (NLP) and large language models (LLMs) in recent years – which lead to the "boom" of the generative pre-trained transformer (GPT) [2] – offers a unique opportunity to guide and advance knowledge in acoustics through automated large-scale text processing. This can provide more accessible information for researchers, practitioners and engineers interested in data-driven methods for acoustics and vibration in the broader sense.

Recent literature surveys in acoustics have reviewed the theory and applications of machine learning (ML) in acoustics [3], sound source localization (SSL) using deep learning (DL) methods [4], as well as noise-induced hearing loss in several contexts [5–7]. The survey by Gannot et al. [8] analyzed 393 papers on speech enhancement and source separation through four queries: What is the acoustic impulse response model? What is the spatial filter design? What is the parameter estimation algorithm? and What is the post-filtering technique? Other related review papers have covered more specific applications of acoustics, such as source-range estimation for underwater acoustics [9], SSL for wireless acoustic sensor networks [10], the LOCATA challenge for source localization and tracking [11] and 15 years of SSL in robotics applications [12].

After it transitioned to Open Access (OA) in 2020 [13], Acta Acustica featured three review articles on the state-of-the-art (SOTA) of the acoustics-related topics: in 2021, McLachlan et al. [14] reviewed the use of Bayesian inference in recent developments on sound localization and the relevance of dynamic cues as a way of explaining behavioral human data; in 2022, Rafaely et al. [15] provided an updated account of emerging methods in spatial audio signal processing for binaural sound reproduction, including perceptually motivated approaches, beamforming, ML related approaches, among others; and, in 2023, Döllinger et al. [16] surveyed current trends and progress in numerical

*Corresponding author: a264372@dac.unicamp.br
[1] https://www.scopus.com/

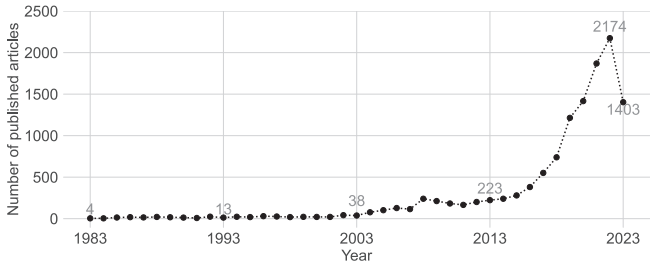

Figure 1. Four decades of articles on applications of data-driven methods in acoustics. The results have been obtained from a Scopus search on August 2, 2023.

modeling of the human phonatory process, including reduced order models, fluid mechanics approaches and models combined with ML methods.

Writing a literature survey can be viewed as the art of *making a long story short*, which can be pretty laborious. Typically, it starts by selecting a topic of interest and elaborating a list of questions. Then, a search for relevant literature items must be fulfilled, which, nowadays, can be facilitated by search engines and databases that assess the credibility and reliability of sources (e.g., Scopus, Google Scholar[2], etc.). This is followed by processing the selected literature, organizing items into categories based on their similarities and differences, analyzing them, and noting essential trends, patterns, knowledge gaps, etc. To do so, several tools exist to provide researchers with ways to document the whole process, with mechanisms to build quality assessment checklists, data extraction forms, among others (e.g., Covidence[3], Parsif.al[4], Rayyan[5], etc.). However, until now, one has to *read through* all the literature.

Reading a scientific paper typically involves scanning the text for the research problem, assumptions, methods, evaluations and main findings; interpreting relevant mathematical terminology; understanding the structure and organization of the text; and synthesizing information to form a coherent understanding of it as a whole [17]. Thus, the time taken to read an academic paper varies depending on various factors, such as its length, the complexity of the topic and the reader's familiarity with the subject matter. Assuming that a familiar reader has a typical reading speed of approximately 200–300 words per minute [18], it would take roughly 1–2 h to read a 10-page academic paper. Math-intensive documents might take even longer. Therefore, scanning 100 articles would take approximately one month of uninterrupted work to read through the literature.

The usage of LLMs for automated text summarization and generation is relatively new and has had applications in medicine and news enterprises. A relevant study to this work was published recently by Tang et al. [19], who performed zero-shot medical evidence summarization generated with GPT-3.5 and ChatGPT and compared them to human-generated summarization. Similar methodologies have been applied to, for example, compare abstracts generated by ChatGPT to real abstracts from medical journals [20], identify and assess key research questions in gastroenterology [21] and answer multiple-choice questions about human genetics [22]. LLMs have also been used for automatic news summarization [23, 24]. A common element in these studies is that LLM-based methodologies have substantial potential in medical and news applications, but more work is needed to increase the accuracy and fidelity.

In this paper, we employ a GPT model to query a literature corpus comprising 117 texts on data-driven speech enhancement methods. The main goal is to speed up literature surveys. The structure of this paper is as follows: Section 2 describes the methodology, including the literature corpus, a short description of the GPT model and the queries posed to the model. Section 3 presents the results of the GPT model and a comparison with a reference (human-based) survey [25]. Lastly, conclusions are drawn in Section 4.

2 Methodology

2.1 Text corpus

In this study, the corpus consists of 117 articles published in the English language between January and December 2021, matching the search strings "*audio enhancement*" OR "*dereverberation*" AND in the context of "*machine learning*" OR "*deep learning*", from various databases, including the AES E-Library[6], ACM Digital Library[7], Google Scholar, IEEE Digital Library[8], JASA[9], MDPI[10], ResearchGate[11], Research Square[12], Science Direct[13], Springer[14], arXiv[15] and some repositories of higher education institutions and subsidiary research departments of corporations.

Conference, journal and challenge papers, book series and chapters, extended abstracts, technical notes, M.Sc. theses and Ph.D. dissertations were included in the search. The average number of pages per article was 8, varying from 2 to 30 (except for the M.Sc. and Ph.D. monographies, which varied from 31 to 118). For the complete list of texts reviewed, please refer to this external link[16].

2.2 Generative pre-trained transformer model

First released in 2018 [26] and then continuously updated, the generative pre-trained transformer (GPT) is

[2] https://scholar.google.com/
[3] https://www.covidence.org/
[4] https://parsif.al/
[5] https://www.rayyan.ai/
[6] https://www.aes.org/e-lib/
[7] https://dl.acm.org/
[8] https://ieeexplore.ieee.org/
[9] https://asa.scitation.org/journal/jas
[10] https://www.mdpi.com/
[11] https://www.researchgate.net/
[12] https://www.researchsquare.com/
[13] https://www.sciencedirect.com/
[14] https://link.springer.com/
[15] https://arxiv.org/
[16] https://drive.google.com/file/d/1rpRiSjyNpHIF9GzNzy8qTQEmKHLMzDkN/

a large autoregressive language model designed to generate human-like responses to natural language input. It can be used for various tasks, including chatbots, language translation and text summarization. Its ability to generate coherent text has made it a valuable tool for researchers and developers in NLP and ML applications. For example, it is possible today to ask ChatGPT or Bard to summarize a scientific paper or generate a list of sources for a literature survey on a specific topic. However, it has been seen that generated responses are often partially (sometimes entirely) fake [27] and the answering accuracy can deteriorate when the answer to the query lies in the middle of the context [28]. Therefore, we have focused on applying the underlying GPT model, not on the direct usage of chatbots.

In this study, we used the large language model of Open AI `GPT3.5-turbo-16k`[17] to process the research papers and extract relevant information. This allows us to explore the model's ability to handle long contexts (i.e., 16k tokens or up to about 50 pages of pure text, assuming an average of 300 tokens/page), enabling a comprehensive analysis of an entire scientific paper. This is in contrast to previous studies on automatic literature summarization [19], which examined scientific abstracts. It should be stressed that articles in PDF in acoustics most often translate into fewer pages of pure text due to figures, tables, etc. We utilize the GPT model's ability to answer questions to help us address specific inquiries about the papers. First, we convert the PDF files into text using the PyPDF2 library[18]. Next, we prompt the GPT model with each paper's full text and specific questions to obtain comprehensive answers. This iterative process is performed for every paper to address the four queries presented in the following section. Compared to the human pace, this methodology requires much less time to analyze academic papers and provide an answer to the question posed.

2.3 Queries

Four questions are considered in this study: two relatively "simple" and two relatively "hard":

- *Query 1 (Q1)*: What country were the authors based in? The output of this question is a list of the authors' countries of affiliation.
- *Query 2 (Q2)*: Was it a single- or multi-channel scenario? The output of this question is either one of the two classes (single or multi) and we are interested in determining the probability of the GPT model in obtaining the right class.
- *Query 3 (Q3)*: What type of architecture was used? This question is relatively more difficult than the previous one, requiring domain knowledge for proper comprehension. This question relates to determining the data-driven model used in the studies. Thus, the output of this question is a string and we are interested in knowing the probability that the GPT will produce the string as accurately as possible.
- *Query 4 (Q4)*: In what context were these applications used (e.g., hearing aids, communication, speech enhancement)? This is the most challenging question posed to the GPT in this study, which involves determining the application area of speech enhancement considered in previous studies. Thus, the output of this question is a string and we are interested in determining the probability that the GPT will produce the string as accurately as possible.

[17] https://platform.openai.com/docs/models/gpt-3-5
[18] https://pypi.org/project/PyPDF2/

3 Results

3.1 Outputs from questions

These four questions were selected from our reference literature survey [25], whose answers are taken as ground truth. Section 3.1.1 summarizes the answers presented in [25], whereas Section 3.1.2 compares the answers produced by the GPT model with the answers in the reference survey.

3.1.1 Human-based survey

Authors' affiliations included higher education institutions, subsidiary research departments of corporations (e.g., Adobe, Facebook, Google, Microsoft), and semi-private and fully financed government research institutions. The main contributors were the United States of America (USA), China and Japan, as illustrated in Figure 2 (Q1), with 28 countries represented. Other contributing countries include South Korea, Germany, the United Kingdom (UK), India, Switzerland, France, Denmark, the Netherlands, Canada, Ireland, Italy, Norway, Spain, Taiwan, Vietnam, Austria, Brazil, Chile, Greece, Hong Kong, Israel, Malaysia, Pakistan, Poland and Singapore.

Not all corpora account for multi-channel scenarios. Among the reviewed articles, only 23% explicitly addressed multi-channel scenarios, whereas 24% focused on single-channel scenarios, as illustrated in Figure 2 (Q2.1). Other scenarios include binaural, Ambisonics and stereo signals. However, most articles did not specify this information. For articles with a complete system format or configuration details, most are single-input-single-output (SISO) systems, followed by multiple-input-multiple-output (MIMO) and multiple-input-single-output (MISO) systems, as shown in Figure 2 (Q2.2). Other formats include multiple-input systems without a specified output format (MIXX), single-input systems without a specified output format (SIXX) and systems with completely unspecified input-output formats (XXXX).

The most commonly used model architectures are 1- and 2-D convolutional neural networks (CNN), uni- or bi-directional long short-term memory (LSTM) blocks, U-nets, fully connected (FC) architectures, attention networks, recurrent neural networks (RNN) and temporal convolutional networks (TCN), as illustrated in Figure 2 (Q3). Other architectures include adversarial, convolutional, encoder/decoder, feedforward, geometrical, neural beamformer, recurrent, reinforcement learning, Seq2Seq and statistical/probabilistic models.

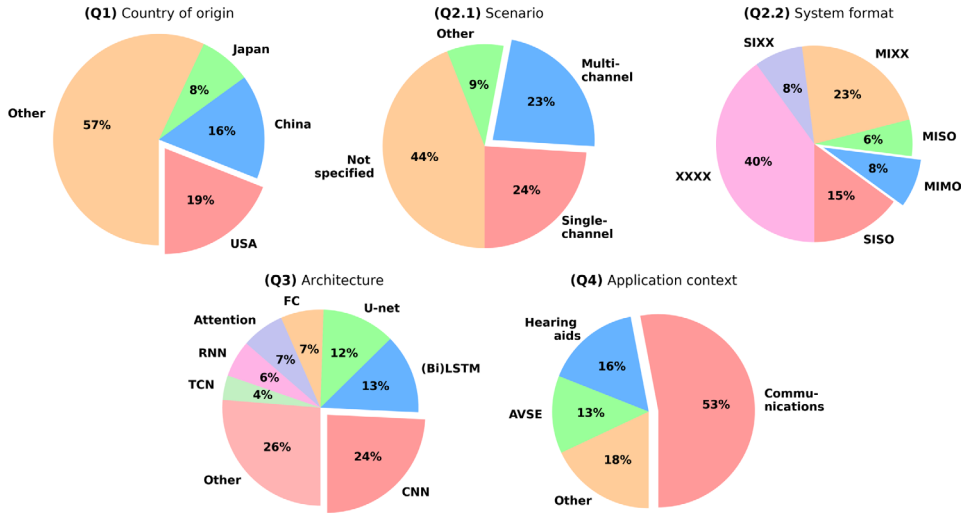

Figure 2. Simplified pie-charts for the human-based survey [25].

Applications are often joint, including speech enhancement, dereverberation, noise suppression, speech recognition and source separation. These applications focus mainly on communication, hearing aids and audio-visual speech enhancement (AVSE), as illustrated in Figure 2 (Q4). Additional applications include suppressing nonlinear distortions, enhancing heavily compressed signals in speech and musical domains, audio inpainting applied to both speech and music signals, law enforcement and forensic scenarios, acoustic-to-articulatory inversion, input-to-output mapping of auditory models, studio recordings and selective noise suppression.

3.1.2 Machine-based survey

Because the GPT model is designed to generate human-like responses to natural language input, even if prompted with the same questions posed by humans, its answers are expected to vary from those of humans. To quantify the extent to which these variations differ from the desired responses, a tier list was elaborated as follows to compare the machine-based results with the human ground truth:

- *Tier 1* – No answer/Completely wrong/Not a pertinent answer: the model fails to provide any response or provides a completely incorrect or irrelevant answer (e.g., the author failed to mention, yet GPT prompts a specific answer);
- *Tier 2* – Marginally correct: the model provides a response that contains at least some correct information;
- *Tier 3* – Mostly correct with minor errors or omissions: the model produces the majority of the information correctly but might miss a few details or make minor mistakes;
- *Tier 4* – Perfectly correct: the model produces completely accurate and correct responses.

The first author, who also conducted the reference human-based survey, performed the tier-based assessment of responses to the survey questions (Q1–Q4 in Sect. 2.3). This choice has been taken to prevent the need for an analysis of the subjective interpretation of the machine-generated responses, which is beyond the scope of this paper. It is worth noting that evaluating more technical questions, such as Q3 and Q4, requires domain knowledge of speech enhancement and data-driven methods. However, assessing responses to more straightforward questions like Q1 and Q2 requires little to no domain knowledge.

Figure 3 illustrates stacked bar charts containing the tier distribution for each question after comparing the machine-based responses with the human-based responses. For full results of the raw human-based survey in comparison with machine outputs for Q1–Q4, please refer to this external link[19]. In what follows, the results are analyzed in more detail.

3.2 Analysis of results

From Figure 3, most answers are either perfectly correct or have minor errors for Q1 ("*What country were the authors based in?*"), which is a simple question that can be answered based on authors' affiliations. In this case, the errors could be related to the fact that the country of affiliation was not included or correctly linked to their names in the provided metadata.

Table 1 illustrates examples of human-based ground-truth versus machine-based responses for Q1. It can be seen that some machine-based answers are more concise than others, specifically, stating only the country versus starting a sentence with "The authors were based in..." This reflects the coherent and diverse capacity of the GPT model to respond to such a question. We strongly suspect that the model's accuracy can be improved by providing only the article's metadata as context instead of the complete text, thus minimizing potential issues due to the length of the context [28].

[19] https://drive.google.com/drive/folders/1jfud4LVkwBQd8KhWKHxXUAkaxZCO9igu?usp=sharing

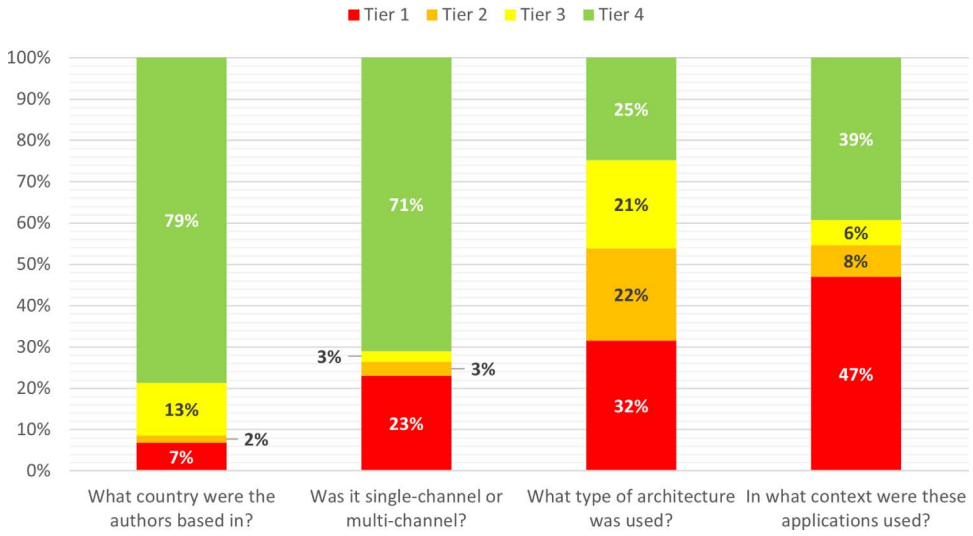

Figure 3. Stacked bar charts for the machine-based answers to the four questions using the corpus of 117 papers.

Table 1. Examples of human-based ground truth versus machine-based responses for Q1.

	Human-based ground truth	Machine-based responses
Tier 1	Germany, UK	The authors were based in the USA and China.
Tier 2	China, USA, Denmark	China
Tier 3	Germany, Canada	Germany
Tier 4	UK, Switzerland	The authors were based in the UK (specifically, the University of Edinburgh) and Switzerland (specifically, ETH Zurich).

Table 2. Examples of human-based ground truth versus machine-based responses for Q2.

	Human-based ground truth	Machine-based responses
Tier 1	Not specified	The approach described in the paper was a single-channel approach.
Tier 2	Single-channel	The proposed system includes both single-channel and multi-channel approaches.
Tier 3	Binaural, Monaural	The proposed method is a multi-channel approach.
Tier 4	Multi-channel	The approach described in the paper is a multi-channel approach.

Regarding Q2 ("*Was it a single- or multi-channel scenario?*"), most predictions are perfectly correct; however, there is an increase in completely inaccurate answers compared to Q1. This is partially due to cases with no specified response (the authors failed to mention) and the GPT model prompting a response [27] instead of stating that it could not find that information in the text provided. In addition, there are cases where the speech enhancement method could be used for multi-channel purposes, yet the authors chose not to. Interestingly, the GPT model may assign these cases as single- and multi-channel scenarios.

Table 2 illustrates examples of human-based ground-truth versus machine-based responses for Q2. The examples show that the GPT model provides nuanced answers for the different tiers, including "The proposed system includes..." or "The approach described in the paper is..." This reflects the complexity of the text generation capabilities of the GPT used in this experiment. As pointed out in Section 2.3, the prompt is a classification question: whether the approach is of a single-channel or multi-channel class. However, when looking at the Tier 2 example in Table 2, the authors discussed a single-channel approach (see Ref. 107 in [25]); however, it can also be applied in multi-channel scenarios, which is the machine response. Similarly, for Tier 3, the GPT model does not explicitly produce the result "Binaural" or "Monoaural". However, it responds that the approach belongs to the multi-channel class. Making the question more precise may help to improve the model's accuracy further.

For question Q3 ("*What type of architecture was used?*"), there is an observable balance between all tiers. One of the most common reasons for completely wrong answers is that the GPT model identifies the "trade" name of the architecture as the type of architecture (e.g., "VGGNet" instead of "fully connected, CNN"). We suspect this can be improved by fine-tuning the GPT model to determine the underlying architecture instead of its variant name. Another common error is simply outputting the

Table 3. Examples of human-based ground truth versus machine-based responses for Q3.

	Human-based ground truth	Machine-based responses
Tier 1	Fully connected, LSTM	The network architecture used in the proposed method is a deep neural network (DNN).
Tier 2	Attention Network, Autoencoder, Conformer Network, CNN, LSTM	The network architecture used in the study is a conformer-based architecture.
Tier 3	CNN, RNN	The network architecture used in the study is a convolutional neural network (CNN) with variable dilation factors.
Tier 4	CNN	The network architecture used in the study was Convolutional Neural Networks (CNNs).

Table 4. Examples of human-based ground truth versus machine-based responses for Q4.

	Human-based ground truth	Machine-based responses
Tier 1	Not specified	The applications of the Neural Beamforming Front-End described in the paper include real-time audio signal enhancement for human-computer interaction, telecommunication, web conferencing, and pro-audio applications.
Tier 2	Voice-Over-IP applications	The applications of high fidelity speech regeneration with application to speech enhancement include speech enhancement for better communication and collaboration, particularly in the context of physical distancing during the COVID-19 pandemic.
Tier 3	Podcasts, video voice-overs, and audiobooks that require studio-quality audio with full bandwidth and balanced equalization (EQ)	The applications of signal processing to audio and acoustics discussed in the paper are focused on speech enhancement, specifically improving the quality of real-world consumer-grade recordings to sound like studio recordings. This can be beneficial for various content creation tasks such as podcasts, video voice-overs, and audiobooks.
Tier 4	Law-enforcement related audio recordings where it's difficult to understand the speech of one specific speaker amongst all the others	The applications mentioned in the text are related to law enforcement and investigations involving audio recordings.

answer "DNN" (e.g., "the network architecture is a DNN") instead of detailing its type. Once again, we strongly suspect that providing the GPT model with the necessary context would prevent these mistakes. At any rate, most answers are partially correct, i.e., either it got something or almost everything right, which, together with the wrong answers, reduces the quantity of perfectly correct answers.

Table 3 presents examples of human-based ground-truth predictions versus machine-based predictions for Q3. Interestingly, for Tier 3, it can be seen that the GPT model not only (nearly) produces the right architecture type of CNN, but it also adds "variable dilation factors." Based on our observations with other papers on the survey, this extracted additional information, if accurate, holds significant value and analysis depth in the context of large-scale surveys.

Finally, for question Q4 ("*In what context were these applications used?*"), most answers were perfectly correct or entirely incoherent. This is because, in most cases, the authors do not mention the context of their applications in the full texts. This adds a higher degree of complexity to the GPT model to infer the application from partially incomplete information, something a human with domain knowledge might infer more accurately at this point. Still, it is interesting that the GPT model considers the broader field of study (e.g., dereverberation and speech enhancement) as an application context and attempts to answer the query nonetheless. Table 4 illustrates examples of human-based ground-truth versus machine-based predictions for Q4. As can be seen, the Tier 4 response example is remarkably similar to the human response. However, for Tier 1, the GPT model answers even though the human has found that it is not specified in the text. Further examination and understanding must rely on the GPT model to answer these queries more accurately.

4 Conclusions

Considering the results unearthed in our study, GPTs are promising tools for processing large amounts of text, but they are not yet a substitute for the nuanced intuition and critical thinking inherent in human researchers. The human factor plays a pivotal role in interpreting complex data and drawing nuanced conclusions, a skill that GPTs have yet to master fully. This does not diminish the overall usefulness of such language models in handling straightforward tasks. We suspect improvements in prompt engineering will enhance these results further, such that the context provides the right cues from which the models can

generate more accurate responses. Similar conclusions can be drawn concerning the length of the prompt and the proper focus on the subject (e.g., data-driven speech enhancement methods in this study) to fine-tune the GPT.

Our study opens the door to a new era in academic research where human intelligence and artificial intelligence (AI) work in tandem. We foresee a collaborative future where AI tools are employed to *augment*, rather than automate human efforts in surveying large amounts of text. While still in its nascent stage, this tandem promises to transform how we conduct literature surveys, write theses and engage in academic discourse. The remark also extends to teaching and learning in higher education, which will re-frame how students are examined today. In the long run, we hope this paper stimulates the adoption of artificially intelligent systems to aid humans in surveying larger corpora (e.g., thousands of articles) in acoustics and beyond.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This study was partially sponsored by the São Paulo Research Foundation (FAPESP) under grants #2017/08120-6, #2019/22795-1, and #2022/16168-7. We also thank Prof. Roberto Lotufo and Prof. Renato Lopes for their valuable discussions and suggestions.

Data availability statement

The data are available from the corresponding author on request.

References

1. M. Park, E. Leahey, R.J. Funk: Papers and patents are becoming less disruptive over time, Nature 613, 7942 (2023) 138–144.
2. C. Stokel-Walker, R. Van Noorden: What ChatGPT and generative AI mean for science. Nature 614, 7947 (2023) 214–216.
3. M.J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M.A. Roch, S. Gannot, C.A. Deledalle: Machine learning in acoustics: Theory and applications. The Journal of the Acoustical Society of America 146, 5 (2019) 3590–3628.
4. P.A. Grumiaux, S. Kitić, L. Girin, A. Guérin: A survey of sound source localization with deep learning methods. The Journal of the Acoustical Society of America 152, 1 (2022) 107–151.
5. R.L. Neitzel, B.J. Fligor: Risk of noise-induced hearing loss due to recreational sound: Review and recommendations. The Journal of the Acoustical Society of America 146, 5 (2019) 3911–3921.
6. K.E. Radziwon, A. Sheppard, R.J. Salvi: Psychophysical changes in temporal processing in chinchillas with noise-induced hearing loss: A literature review, The Journal of the Acoustical Society of America 146, 5 (2019) 3733–3742.
7. K. Sonstrom Malowski, L.H. Gollihugh, H. Malyuk, C.G. Le Prell: Auditory changes following firearm noise exposure, a review. The Journal of the Acoustical Society of America 151, 3 (2022) 1769–1791.
8. S. Gannot, E. Vincent, S. Markovich-Golan, A. Ozerov: A consolidated perspective on multimicrophone speech enhancement and source separation. IEEE/ACM Transactions on Audio, Speech, and Language Processing 25, 4 (2017) 692–730.
9. H.C. Song, G. Byun: An overview of array invariant for source-range estimation in shallow water. The Journal of the Acoustical Society of America 151, 4 (2022) 2336–2352.
10. M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, B. Lee: A survey of sound source localization methods in wireless acoustic sensor networks. Wireless Communications and Mobile Computing. 2017.
11. C. Evers, H.W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P.A. Naylor, W. Kellermann: The LOCATA challenge: Acoustic source localization and tracking. IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020) 1620–1643. https://doi.org/10.1109/TASLP.2020.2990485.
12. S. Argentieri, P. Danes, P. Souères: A survey on sound source localization in robotics: From binaural to array processing methods. Computer Speech & Language 34, 1 (2015) 87–112.
13. M. Kaltenbacher, J. Kergomard, M. Gaborit, T. Scotti, A. Ruimy: Acta Acustica: State of art and achievements after 3 years. Acta Acustica 7 (2023) E1. https://doi.org/10.1051/aacus/2023003.
14. G. McLachlan, P. Majdak, J. Reijniers, H. Peremans: Towards modelling active sound localisation based on Bayesian inference in a static environment. Acta Acustica 5 (2021) 45. https://doi.org/10.1051/aacus/2021039.
15. B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, P. Samarasinghe: Spatial audio signal processing for binaural reproduction of recorded acoustic scenes-review and challenges. Acta Acustica 6 (2022) 47. https://doi.org/10.1051/aacus/2022040.
16. M. Döllinger, Z. Zhang, S. Schoder, P. Šidlof, B. Tur, S. Kniesburges: Overview on state-of-the-art numerical modeling of the phonation process. Acta Acustica 7 (2023) 25. https://doi.org/10.1051/aacus/2023014.
17. E. Pain: How to (seriously) read a scientific paper. Science 10 (2016). https://doi.org/10.1126/science.caredit.a1600047.
18. S.D. Frank: Remember everything you read: The Evelyn Wood 7 day speed reading and learning program. Crown, 2012.
19. L. Tang, Z. Sun, B. Idnay, J.G. Nestor, A. Soroush, P.A. Elias, Y. Peng: Evaluating large language models on medical evidence summarization, npj Digital Medicine 6, 1 (2023) 158.
20. C.A. Gao, F.M. Howard, N.S. Markov, E.C. Dyer, S. Ramesh, Y. Luo, A.T. Pearson: Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digital Medicine 6, 1 (2023) 75.
21. A. Lahat, E. Shachar, B. Avidan, Z. Shatz, B.S. Glicksberg, E. Klang: Evaluating the use of large language model in identifying top research questions in gastroenterology. Scientific Reports 13, 1 (2023) 4164.
22. D. Duong, B.D. Solomon: Analysis of large-language model versus human performance for genetics questions. European Journal of Human Genetics (2023) 1–3. https://doi.org/10.1038/s41431-023-01396-8.
23. S. Syed, R. El Baff, J. Kiesel, K. Al Khatib, B. Stein, M. Potthast: News editorials: Towards summarizing long argumentative texts, in: Proceedings of the 28th International Conference on Computational Linguistics. 2020, pp. 5384–5396. https://doi.org/10.18653/v1/2020.coling-main.470.

24. T. Goyal, J.J. Li, G. Durrett: News summarization and evaluation in the era of gpt-3, 2022. arXiv preprint arXiv:2209.12356.
25. A. dos Santos, P. de Oliveira, B. Masiero: A retrospective on multichannel speech and audio enhancement using machine and deep learning techniques, in: Proceedings of the 24th International Congress on Acoustics. 2022, pp. 173–184.
26. A. Radford, K. Narasimhan, T. Salimans, I. Sutskever: Improving language understanding by generative pre-training. 2018.
27. H. Alkaissi, S.I. McFarlane: Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 15, 2 (2023) e35179.
28. N.F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, P. Liang: Lost in the middle: How language models use long contexts. 2023. arXiv preprint arXiv:2307.03172.

Cite this article as: dos Santos A. Pereira J. Nogueira R. Masiero B. Sander Tavallaey S, et al. 2024. An experiment on an automated literature survey of data-driven speech enhancement methods. Acta Acustica, **8**, 2.