

Scientific Paper

Advancing Auscultation Education: Signals Visualization as a Novel Tool for Enhancing Pathological Respiratory Sounds Detection

Anna Katarzyna PASTUSIAK^{1,2,ABD,*}, Honorata HAFKE-DYS^{1,2,AEF}, Jędrzej KOCIŃSKI^{1,2,CEF},
Krzysztof SZARZYŃSKI^{2,BC}, Kamil JANECEK^{3,EF}

¹Department of Acoustics, Faculty of Physics, Adam Mickiewicz University Poznan, Poland

²StethoMe, Poznan, Poland

³Department of Paediatric Pulmonology and Rheumatology, Medical University of Lublin, Poland

*Corresponding author: Anna Pastusiak; anna.pastusiak@amu.edu.pl

(received 21 July 2023; revised 19 October 2023; accepted 18 December 2023)

Abstract

Introduction: Auscultation remains a pivotal diagnostic modality for various respiratory pathologies. To augment its clinical relevance, the continuous expansion of our understanding of pulmonary acoustics, coupled with the advancement of auscultation recording and analysis methodologies, is imperative.

Material and methods: We investigated how the bimodal presentation of auscultatory signals (sound and visual cue perception) influences the subjective efficacy of pathological respiratory sound detection, which is a critical step in the development of a new auscultation tool.

Recordings of pediatric breath sounds were presented in three different forms - audio only, visual representation only (spectrogram) or audiovisual (both together). The F1-score, sensitivity and specificity parameters were calculated and compared to the gold standard (GS). Subsequent to the detection experiment, participants completed a survey to subjectively assess the usability of spectrograms in the procedure.

Results: Over 60% of all responders ranked the spectrogram as important (40.8%) or very important (21.1%). Moreover, 11.3% of all participants found this new form of presentation of auscultation results to be more useful than the evaluation of sound only. The addition of visual information did not statistically significantly change the evaluation of the sounds themselves, an observable trend implies that enhancing audio recordings with visualizations can enhance detection performance. This is evident in the 4 p.p. and 2 p.p. sensitivity increments for physicians and students, respectively, even without specialized visual training.

Conclusions: Our research findings indicate that the integration of spectrograms with conventional auditory assessment, albeit based on observed trends and survey responses, presents a promising avenue for improving the precision and quality of medical education, as well as enhancing diagnosis and monitoring processes.

Keywords: pulmonary; medical education; breath sounds; wheezes; crackles.

Introduction

Auscultation, a process that is strongly subjective by nature and the results of which are thus hard to share, remains an essential part of basic medical examination used to distinguish between normal and adventitious respiratory sounds¹ in various conditions, including e.g. asthma. There are many factors, which may influence the obtained results and make them ambiguous or even not credible. Indeed, the quality of auscultation depends on the perceptual abilities as well as experience of the examiner, is subject to inherent inaccuracies of the process and suffers additionally from the lack of standardized classification used to describe pathological sounds.²⁻⁴ These factors can and often do

lead to improper interpretation of breath sounds and thus directly affect the treatment of patients.

In everyday life scenarios, enhancement of the accuracy of perception is achieved by operating with a number of senses simultaneously utilizing information redundancy, to extract more details from noisy or blurred signals that we experience. While the information accessed by different senses can be markedly different, there is usually a strong overlap e.g. between what people hear, see and identify as a sound source.⁵ This feature is very effective for example in speech intelligibility where observing the lip movements of an interlocutor significantly increases understanding of what was said even in

noisy conditions. Given this natural and effective way of enhancing signal identification (proven in numerous studies⁶⁻⁷) it is natural to expect that auscultation results may be improved by invoking multimodal signals. In this context, sound and its visual representation. The most common way of sound visualization that brings additional information is a spectrogram which is a three-dimensional plot that depicts the energy changes at different frequencies over time. Therefore, visual information should complement audio information making the auscultation process easier thus perception and identification of different pathologies are more effective. An important prerequisite for this is, however, proper spectrogram interpretation by the auscultating person.

Indeed, most recent works show that such multimodal operation may have a positive impact on the performance of medical staff in the distinction of breath sound.⁸ Taking this into account one may state that using this additional visual information may be a good training tool making the quality of medical education higher translating into subsequent better decision-making in primary health care. The lingering inquiry pertains to the necessity of providing support to medical students and personnel, and the methodological approach employed for the evaluation of such a solution. A prevailing consensus exists within the scientific community that auscultation, owing to its inherent subjectivity, spatial localization (primarily centered on the thoracic region), and temporal variability (pathological sounds often manifest intermittently within a respiratory cycle), presents formidable challenges in the realms of pedagogy and acquisition. This issue engenders a ubiquitous issue concerning the identification and comprehension of pathological auscultatory findings, which extends beyond the realm of novice medical trainees and afflicts seasoned healthcare practitioners as well.^{2,9,10}

Within the scope of our research, we investigated not only medical students but also a select cadre of seasoned physicians who routinely utilize both conventional and electronic stethoscopes in their clinical practice. This comprehensive approach was undertaken to ascertain whether these experienced practitioners could effectively leverage supplementary information beyond their extensive clinical acumen. Moreover, by incorporating this group, we aimed to establish a benchmark that could serve as a reference point for evaluating the performance of medical students.

To gain a comprehensive insight into the findings, we conducted a dual assessment encompassing both the participants' proficiency in identifying pathological respiratory system auscultatory sounds and their subjective evaluation of the spectrogram as a supplementary tool. This evaluation encompassed its significance in the context of auscultation and the detection of pathological sounds, as well as an appraisal of the importance of both visual and auditory cues in the identification of specific pathologies.

In recent times, a matter of paramount significance pertains to the substantial advancements in the field of telemedicine, along with the proliferation of tools for remote auscultatory patient assessment, which have now emerged as the prevailing standard. This paradigm shift introduces a new set of challenges, given that the physician is no longer afforded the immediate presence of the patient. Consequently, it becomes imperative for the physician to perform an accurate and efficient auscultatory examination. In this context, the incorporation of spectrograms has the potential to further streamline and expedite the assessment process.

The central thrust of this study lies in providing an initial foundation for the consideration of this critical topic. It primarily revolves around the initial step in comprehending the role of visual cues in the accurate analysis of adventitious sounds within auscultation recordings. Moreover, this research contributes to the broader objective of enhancing the education of this skill among medical students.

The persistence of this problem underscores its status as an ongoing, globally pervasive challenge, with physicians themselves acknowledging the inadequacy of both educational methodologies and their own experiential knowledge in effectively addressing this issue.¹⁰

Materials and methods

The study was approved by the Bioethics Committee at Poznań University of Medical Sciences (no. 193/18).

Subjects

All participants took part in the experiment voluntarily and anonymously and they were not paid for participation in the research. Before the test, detailed instructions and essential definitions were provided to disambiguate the nomenclature. Partakers received additional materials covering the topic of auscultation sounds and example spectrograms. The tested group included 65 medical students, representing different levels of education, but possessing at least basic knowledge of respiratory system auscultation phenomena. Furthermore, we involved an additional group of six physicians (three pediatricians and three internists) as a reference group. These individuals possessed prior exposure to the utilization of spectrograms for the assessment of breath sound recordings, albeit not as a routine component of their clinical practice.

Auscultation recordings

Since no objective measure providing an accurate classification of pathological breath sounds exists, it was essential to establish a point of reference or ground truth (gold standard, GS). The detection test consisted of 20 respiratory system recordings taken at different auscultation chest points. They were chosen from a representative database of over 500 recordings of

pediatric examinations performed using a CE certified electronic stethoscope (StethoMe Smart Stethoscope and Littmann 3200) by trained pediatricians (Department of Pediatric Pulmonology, Karol Jonscher University Hospital in Poznan, Poland). The chosen signals were obtained from the patients between 8 months and 16 years old (median: 3.8 y.o.) who were healthy or suffering from cystic fibrosis, asthma, pneumonia and bronchitis chosen randomly.

Each recording underwent a meticulous assessment conducted by one of five medical doctors, who tagged pathological sounds within the respective recording for use in the GS. Subsequently, two additional medical doctors independently verified each recording to ensure the accuracy of the identified pathological sounds. This evaluation involved an analysis of both the presence and severity (0- no pathological sound, 1- minor pathological sound present in the recording, 2- significant pathological sound present in the recording) of auscultatory phenomena, categorized into four classes: wheezes, rhonchi, coarse crackles, and fine crackles.

The selection of signals presented to listeners was a deliberate process, aiming to closely mimic the occurrence of auscultatory changes in real-life scenarios. It was based on a comprehensive and representative database of authentic recordings. Additionally, precautions were taken to mitigate the risk that any of the auscultatory phenomenon classes did not manifest in the entirety of the recordings chosen for this study. **Table 1** presents information about the presence of each respiratory phenomenon in each recording.

The test material was divided into three settings (audio only, visual only, audiovisual), yielding $3 \times 20 = 60$ samples, each about 15 seconds long. These signals were presented in the online test shared with the participants.

Survey

The test was followed by a short, Google-forms-based online survey with only three questions (Qs). Participants were asked which cues (audio, visual, or audiovisual) were the most useful in phenomena detection. Q1 employed a Likert scale¹¹, while other questions required participants to select a single category. Specific questions were:

Q1: *To what extent do you consider a spectrogram crucial for the assessment of auscultation recordings?* (Provide your answer on a 5-point scale: 1 - Not important at all, 5 - Very important).

Q2: *Which cues in the evaluation of auscultation recordings held the greatest significance for you?* (Select one response: "mostly spectrogram", "mostly sound", "spectrogram only", "sound only", "both (spectrogram and sound)").

Q3: *For which phenomena is the spectrogram considered most critical?* (Select one response: "rhonchi", "wheezes", "coarse crackles", "fine crackles", "all", "not important").

Table 1. Description of pathological respiratory system sounds content and their intensity (0- no pathological sound, 1-minor, 2-significant) present in each recording

rec. no.	intensity level			
	wheezes	rhonchi	coarse crackles	fine crackles
#1	0	0	0	0
#2	0	0	0	0
#3	2	0	0	0
#4	2	2	0	0
#5	2	0	0	2
#6	0	2	0	2
#7	0	2	2	0
#8	2	0	0	2
#9	1	2	0	0
#10	1	2	0	1
#11	2	1	0	0
#12	2	2	0	1
#13	0	0	0	2
#14	2	0	0	2
#15	0	0	2	0
#16	2	0	2	0
#17	2	0	0	2
#18	1	0	0	2
#19	0	2	2	0
#20	0	0	0	2

Study design

The initial phase of the test entailed the presentation of auscultation recordings across three distinct settings: (1) audio-only stimuli - the most common way of auscultation (wav files presented via headphones), (2) visual stimuli only as a spectrogram presented at a computer screen, (3) audiovisual stimuli - sound and spectrogram jointly.

Participants were allowed to use their own listening devices, in particular headsets as well as their own computer screens. In pursuit of a more faithful emulation of real-life auscultation, the utilization of external loudspeakers and the exclusive reliance on visual representations on tablet or mobile devices for listening were strongly discouraged. No predefined sound level was set during the presentation - students and physicians were free to adjust the loudness to their preferred level. Participants were allowed to listen to signals and verify the spectrogram multiple times, if needed.

Spectrograms were generated through the application of the Non-Uniform Fast Fourier Transform (NFFT) method, employing an NFFT analysis with 512 samples. The dynamic range of the spectrogram was established at 130 dB, while the upper frequency limit for display was set at 2000 Hz, encompassing the entire relevant spectrum within which significant auscultatory sounds, both normal and pathological, may manifest.^{12,13} Within the spectrogram, the temporal dimension is mapped along the vertical axis, denoting time in seconds. Simultaneously, the horizontal axis delineates sound frequency (in Hz), spanning from lower frequencies (typically associated with rhonchi) to higher frequencies (e.g., wheezes).

The color or grayscale intensity within the spectrogram serves as an indicator of intensity, with greater intensity at specific time and frequency points manifesting as a more pronounced reddish or black hue in the spectrogram, as exemplified in **Figure 1**.

From a physical point of view, pathological auscultation signals may be divided into two main groups: continuous (wheezes, rhonchi) and transient (coarse and fine crackles). Their spectrograms exhibit variations that are distinguishable from one another, displaying discernible patterns. Notably, in this representation, tonal sounds are observable as horizontal lines, while brief, high-energy impulses manifest as vertical lines, distinctively visualized through specific colors or darkening, as depicted in **Figure 1**. Such distinctions hold the potential to enhance the educational experience of medical students as well as assist in the development of a clinical impression and the classification of medical conditions.

Taking this into account we assumed that the spectrogram can serve as a supporting tool to identify and differentiate abnormal lung sounds by providing direct information on their pitch, level of sound energy, as well as, continuous or discontinuous character. Audiovisual stimuli were synchronized so that the observer could see the exact visual form of the signal at the time he/she heard it.

Each participant was asked to listen to the recordings/look at the spectrogram/both and describe presented samples based on: (1) the appearance of the breathing pathological phenomena of four classes (rhonchi, coarse crackles, fine crackles, wheezes), (2) pathology intensity (0 - no pathological sound, 1 - minor pathological sound present in the recording, 2 - significant pathological sound present in the recording). The signals were randomly presented to the participants, and each individual assessed all 20 available recordings within each of the three previously mentioned settings: audio, visual, and audiovisual. This arrangement yielded a total of 60 presentations for evaluation. To mitigate the potential impact of learning induced by the concurrent exposure to overlapping cues in a short time frame (such as the same signal presented in multiple settings),

the entire procedure was divided into two halves, each separated by a several-day break.

Following the completion of the detection test, participants were presented with a set of survey questions, to which they provided responses through an electronic platform.

Data Analysis

For the pathological respiratory sound detection test confusion matrices were analyzed - the values of sensitivity (TPR - true positive rate; the proportion of actual positives that are correctly identified), specificity (TNR - true negative rate; the proportion of actual negatives that are correctly identified), and F1-score (the harmonic mean of TNR and TPR) were measured with respect to the GS for phenomena detection by medical students and reference group of physicians. Given the unequal group (physicians and students) sizes in the research, separate ANOVA tests were conducted for each group separately as for all participants as one group. Although it was not feasible to conduct a direct statistical comparison between the two groups, it is plausible to consider physicians as an additional reference point in relation to medical students. In both groups, statistical analysis was carried out based on several criteria, specifically: test mode (audio only, visual only, audiovisual) and phenomena type (wheezes, rhonchi, coarse crackles, fine crackles). To account for students' limited auscultation experience and the subtle nature of minor pathological changes, we amalgamated two intensity groups (0 for no changes and 1 for minor pathological changes) and scored them together. This means that, for a specific phenomenon tagged with 0 or 1 in the GS, both responses of 0 and 1 were considered correct.

The survey responses underwent analysis through a standard statistical method that involved tallying the frequency of responses for each scale number for question one, and examining the distribution of responses for questions two and three.

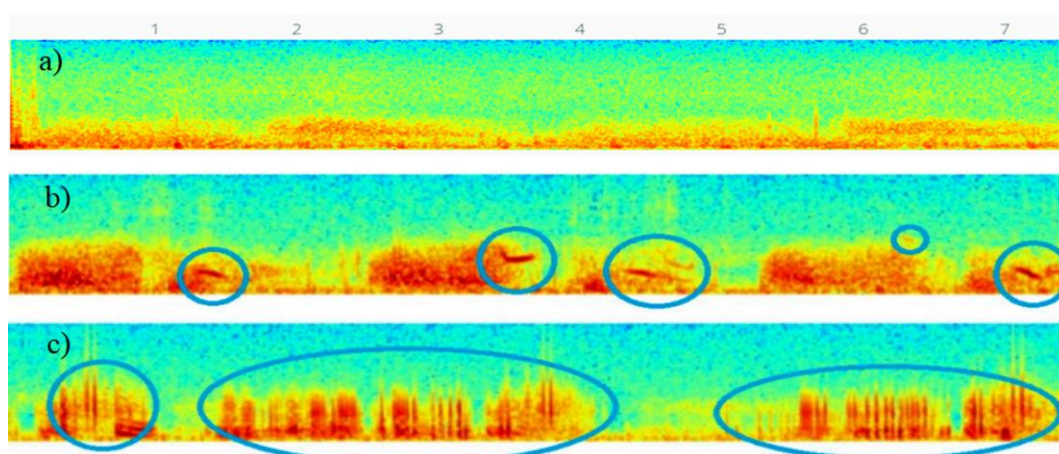


Figure 1. Example of the auscultation recording's spectrogram: a) normal sound, b) wheezes, c) coarse crackles. Characteristic courses (shapes) of both pathologies are marked with blue ellipses.

Results

Pathological respiratory system sounds detection

GS was used as a point of reference (100%) for descriptions made by test participants separately in each group. In this way the values of the parameters describing their performances - F1-score, specificity and sensitivity were established. These statistical measures are commonly used in research on auscultatory sounds and the systems (or devices) that support their detection and/or classification.¹ Detailed values of various statistical parameters corresponding to the three experimental settings in both recipient groups are presented in **Table 2** (F1-score, sensitivity, specificity). F1-score, sensitivity and specificity values depending on the phenomena's type are presented in **Table 3**.

All the results underwent statistical analysis to investigate the differences between three modes of presentation: (1) audio stimuli only, (2) visual stimuli only, (3) audiovisual stimuli. Although the results of the two-factor ANOVA did not reveal statistically significant differences (failing to meet the $p < 0.05$ criterion) between presentation modes, distinct trends were perceptible. Moreover, in each group the phenomena factor was statistically significant ($p < 0.01$).

Analyzing **Table 2** one may state that the lowest sensitivity values were observed in the visual-only setting (0.72 and 0.68 for physicians and students, respectively), but when visual representation becomes a supportive element of conventional audio presentation, the accuracy of auscultation recording description rises in both groups (0.83 and 0.75, respectively). The same tendency is visible for specificity and F1-score. Given

the observed statistical insignificance of these factors, it is not justified to proceed with a more extensive analysis.

In the case of phenomena, the lowest sensitivity values were recorded for fine crackles in medical students and physicians (0.55 and 0.63, respectively) in visual-only mode while traditional listening of auscultatory sounds raises those values to 0.62 and 0.75, respectively. The combination of both senses gives an additional increase only in physicians (3 p.p.), while in students no benefit is observed. On the other hand, the highest scores were obtained for audiovisual presentation of wheezes (0.88 for students and 0.9 for doctors). However, the benefit of bimodal presentation is neglectable. Finally, the highest benefit in that matter can be found in rhonchi giving 5 p.p. for both groups. Those results are confirmed by two other measures, namely F1-score and specificity.

Table 2. F1-score, sensitivity and specificity values for all three types of presentation (visual only, audio only, audiovisual) in physicians (P) and medical students (S).

	mode	P	S
F1-score	visual only	0.70	0.65
	audio only	0.77	0.71
	audiovisual	0.81	0.73
sensitivity	visual only	0.72	0.68
	audio only	0.79	0.73
	audiovisual	0.83	0.75
specificity	visual only	0.74	0.77
	audio only	0.86	0.84
	audiovisual	0.86	0.81

Table 3. F1-score, sensitivity and specificity for all phenomena types in physicians (P) and medical students (M).

pathological sound	mode	F1-score		sensitivity		specificity	
		P	S	P	S	P	S
coarse crackles	visual only	0.62	0.64	0.64	0.68	0.82	0.80
	audio only	0.76	0.71	0.79	0.74	0.86	0.86
	audiovisual	0.80	0.73	0.83	0.77	0.88	0.81
fine crackles	visual only	0.63	0.52	0.63	0.55	0.58	0.57
	audio only	0.73	0.60	0.75	0.62	0.87	0.78
	audiovisual	0.78	0.61	0.78	0.62	0.82	0.70
wheezes	visual only	0.83	0.76	0.85	0.80	0.92	0.95
	audio only	0.86	0.86	0.88	0.89	0.98	0.97
	audiovisual	0.88	0.86	0.90	0.88	0.98	0.97
rhonchi	visual only	0.72	0.66	0.75	0.67	0.64	0.76
	audio only	0.74	0.69	0.75	0.67	0.74	0.74
	audiovisual	0.79	0.72	0.80	0.72	0.76	0.77

Survey

Results are presented in **Figures 2-4**. Each figure depicts the percentage of answers. Only one response was available in each question.

The majority of respondents indicated that the supplementary information provided by the spectrogram when added to the audio is either important (approximately 41%) or very important (about 21%). Only a minor proportion of respondents, specifically 2%, considered the spectrogram to be entirely unimportant, while an additional 8% expressed that it was relatively unimportant. Results indicated that 35% of the participants affirmed the significance of both auditory and visual cues in their perceptual processes. Concurrently, 45% of the participants expressed a predominant reliance on auditory cues, while a minority, approximately 7%, exclusively relied on auditory cues.

Notably, a smaller subgroup, approximately 2% of the respondents, exhibited a distinct preference for spectrogram-based information over auditory cues and bimodal presentation. This finding underscores the presence of a niche preference for visual representations in the domain of aural perception.

Furthermore, when considering the specific utility of spectrograms and bimodal presentation across various phenomena, a majority of the participants (over 31%) recognized the fundamental importance of spectrograms in the detection of all phenomena. In distinct phenomena, such as rhonchi, the bimodal presentation was found to offer significant benefits, with approximately 20% of participants acknowledging its advantageous role. This was followed closely by coarse crackles and wheezes, with slightly more than 20% of participants recognizing the utility of bimodal presentation for these phenomena.

Curiously, a consistent 7% of the participants across all phenomena expressed that bimodal presentation did not significantly contribute to their perceptual processes.

Discussion

In the case of a detection task, the ANOVA in auscultatory test outcomes between audio and audiovisual modes appears to exhibit statistically insignificant disparities, with a discernible inclination towards marginally elevated results in the audiovisual presentation compared to the conventional audio-only approach. However, different phenomena were detected in a significantly different way. Considering these insights, along with the discernible trends we have discussed, we delve into the implications of the results below.

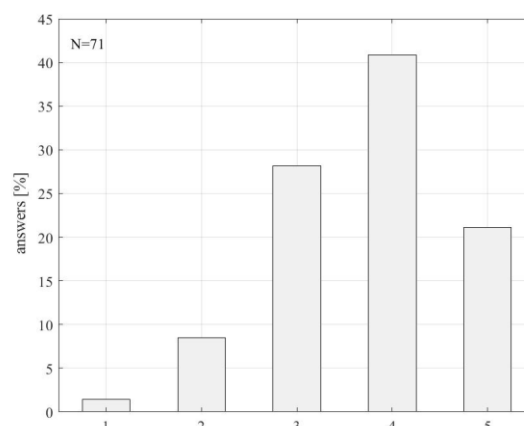


Figure 2. Percentage of participant responses (jointly for 65 students and 6 physicians) to Q1: To what extent do you consider a spectrogram crucial for the assessment of auscultation recordings? [1 - not important at all, 5 - very important].

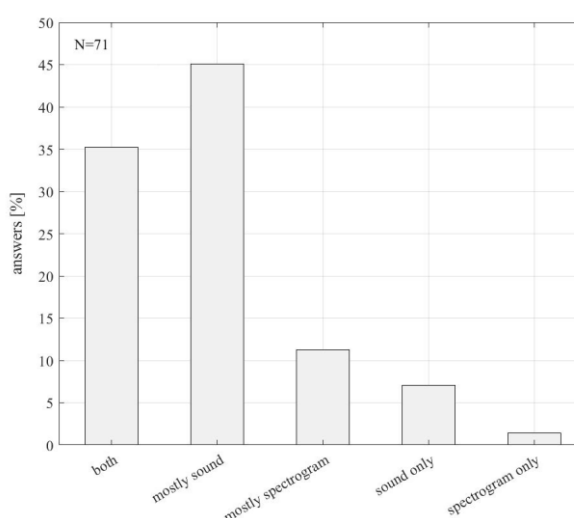


Figure 3. Percentage of participant responses (jointly for 65 students and 6 physicians) to Q2: Which cues in the evaluation of auscultation recordings held the greatest significance for you?

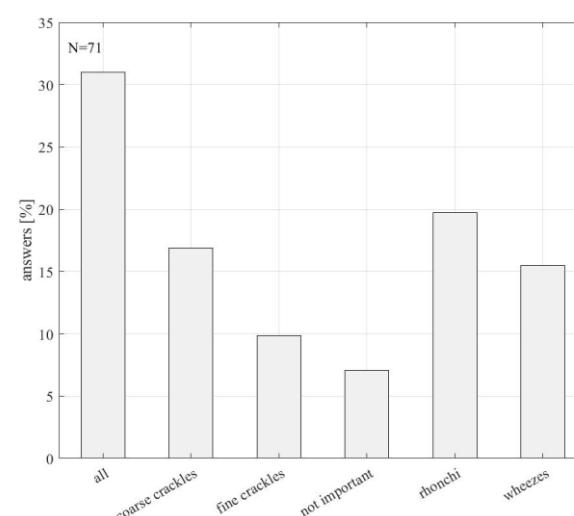


Figure 4. Percentage of participant responses (jointly for 65 students and 6 physicians) to Q3: For which phenomena is the spectrogram considered most critical?

The least sensitive results, encompassing all phenomena presented via audiovisual means, were obtained for fine crackles (0.62 for students and 0.78 for physicians). The interpretation of this outcome can be comprehensively elucidated within the specified context, which encompasses the presence of spectrograms portraying not only the 'relevant' signals such as bronchial sounds, heartbeats, and, in certain instances, breath phenomena, but also internal and/or external noise. Even upon filtering, the presence of such artifacts remains inevitable and even more pronounced in visual presentation, than audio recording alone. This is important mainly in the case of noise sources which are not necessarily linked to ambient conditions (background speech, medical apparatus sounds, etc.), but rather to 'internal' noise caused by unstable chest piece position or tapping of the stethoscope membrane. These kinds of distortions may be similar to transient phenomena like crackles. This results in a noteworthy increase in the false positive ratio, subsequently diminishing the specificity value.

The sensitivity results align with those of specificity, underscoring that the advantage of employing additional sensory modalities beyond the conventional auditory perception is contingent on the type of phenomenon under consideration. This can be related to the ambiguous nature of the sound phenomena - both, when it comes to their acoustic parameters and nomenclature, which remains inconsistent. It is well known in the literature, that diagnosticians use different descriptions of the same sounds or semantically similar terms to describe different phenomena.^{2,9,10} Moreover, it seems that in crackles medical proficiency is crucial in their proper detection - the performance of physicians is better than medical students (even up to 16 p.p. in the case of fine crackles). Their transient nature makes them susceptible to being misinterpreted as noise.¹⁴ Therefore, the utility of visual aids is emphasized with a 4 p.p. advantage observed for both groups in identifying coarse crackles through spectrogram access.

The presence of the spectrogram may be a supporting tool also when it comes to rhonchi. This also may be related to the acoustic characteristics of this type of phenomenon - tonality leads to their better visibility in the spectrograms. Up to 5 p.p. higher accuracy in both groups may be especially important in the light of previous research claiming that rhonchi are the most ambiguous class of breath sounds.¹⁵

In both groups we observe a markedly smaller benefit stemming from the use of spectrograms in the case of wheezes, which as continuous, musical sounds, are quite easy to identify based on sound only. It is noteworthy, however, that the identification of this pathology based solely on spectrogram analysis appears to be relatively straightforward, with a sensitivity of 0.85 for physicians and 0.80 for students. This observation implies that, for the assessment of such breath sounds, spectrogram analysis may offer a more accessible and expedient method for the identification of adventitious sounds.. Such results are particularly significant in the case of asthma

patients, for whom wheezes are a common symptom. In the context of remote monitoring of those patients, the ability to quickly assess the presence of these pathologies, especially when wheezes occur most frequently, will greatly facilitate and expedite clinical practice.

When scrutinizing the survey responses, it becomes evident that for a significant part of the recipients, the most important information during an analysis of auscultation recordings is sound which is consistent with the traditional approach to auscultation and their experience. However, over 60% of all responders ranked the spectrogram as important (almost 41%) or very important (21 %). Interestingly, 11.3% of all participants, although having just limited experience with spectrogram reading, found this new form of presentation of auscultation results more useful than the evaluation of sound signals only. This indicates that using visual representation of the acoustic content is regarded as a potential source of useful, clinical information that can be used by physicians during auscultation. It must be stressed, however, that well-designed training is needed to ensure optimal use of these additional cues.

Significantly, it is noteworthy that a mere 10% of the participants, when queried on the importance of additional visual cues, rated them with only 1 or 2 points in response to Q1. This observation underscores the prevailing sentiment among the majority of respondents regarding the importance of supplementary visual cues, indicating a broad consensus on their significance in the perceptual processes under examination. These findings find support in the broader spectrum of responses obtained in this study. Approximately 47% of participants explicitly recognize the added value of spectrograms as supplementary visual cues in their aural perception, indicating a substantial portion of the sample acknowledging their utility. Conversely, the majority, constituting 53% of the participants, continue to favor the conventional approach of traditional auscultation, a preference rooted in familiarity and established practice. This distribution of preferences underscores the ongoing coexistence of both traditional and technologically augmented methods in the field of aural perception, reflecting the diverse attitudes and practices within the study's participant cohort.

Finally, it is notable that an overwhelming majority of participants, constituting approximately 97%, perceive the utility of additional visual cues in the detection of at least one phenomenon. Within this group, a substantial 31% of respondents recognize the value of these cues in the recognition of all phenomena, highlighting their widespread applicability.

A more granular analysis of these results reveals that 20% of participants specifically appreciate spectrograms for their role in rhonchi recognition. This finding aligns with the outcomes of a recognition test, where the most substantial increase in performance was observed in the context of rhonchi identification. This convergence of results suggests a strong association between the recognition of rhonchi and the value

attributed to spectrograms, lending further empirical support to the significance of visual cues in the perceptual processes and the effectiveness of spectrograms in particular. Furthermore, a similar trend is observed in the case of coarse crackles, where a significant number of participants place importance on the additional visual cues. This sentiment is further mirrored in the recognition test results, which indicate a noteworthy improvement in the identification of coarse crackles when visual cues are incorporated into the process. This is in line with Hafke-Dys et al. findings regarding phenomena recognition in medical staff and students.¹⁰

In the context of wheezes, participants similarly emphasize the importance of visual cues, yet a subtle difference arises in the interpretation of these results. It appears that in the case of wheezes, the visual cues play a distinct role, likely serving as a confirmation or reinforcement of the highest detection scores. Participants seem to rely on visual information to augment their confidence in wheeze identification, further highlighting the value of visual cues in this particular phenomenon. In this context the incorporation of a visual component has the potential to expedite the diagnostic process, as the visual information alone could offer valuable insights, thereby enhancing the expeditious monitoring of conditions like asthma.

Conversely, fine crackles, known for their intricate and challenging identification, were rated the lowest in terms of the importance attributed to visual cues in their recognition. This finding suggests that participants may perceive fine crackles as less reliant on visual assistance, possibly due to the complexities associated with their auditory characteristics. This divergence in importance rankings underscores the nuanced interplay between auditory and visual cues, which varies according to the specific features and recognition challenges of each phenomenon.

Of particular interest is the fact that 11% of participants predominantly rely on spectrograms, while an additional, albeit small, fraction of approximately 2% exclusively utilizes spectrograms as their primary source of information. One plausible explanation for this observation is that these individuals may possess relatively lower proficiency in traditional audio-only auscultation skills, prompting them to perceive spectrograms as a more informative and effective alternative for their specific needs. This observation further supports the argument that a well-structured educational approach has the potential to yield significant enhancements in auscultation skills among both students and practicing physicians.

Finally, these findings shed light on the nuanced and multifaceted interplay between auditory and visual cues in the context of aural perception, with implications for future research in the field. The results suggest that even during the learning phase, various preferences and strategies are associated with the interpretation of auditory assessments and the allocation of attention between the visual and auditory senses. It also implicates the value of incorporating a visual aspect, particularly

to support individuals with hearing impairments and those interpreting test results in environments lacking acoustic conditions for sound perception.

It is important to underscore certain limitations inherent to this study, particularly the dissimilar sizes of the study groups. Nonetheless, it is essential to recognize that the primary focus of this research centers on potential applications in medical education. This emphasis is particularly pertinent to the numerous medical students who comprised a substantial portion of the experimental cohort. While the smaller group of doctors serves as a preliminary reference, their role within this context should not be overlooked.

It seems that many aspects may influence the efficiency of using additional cues in the evaluation of breath sounds. Providing spectrogram information is not sufficient and the potential benefit can not be understood per se. Moreover, one needs to take into account that auscultation is not limited to one location only. Accurate following of the repeatable (or not) visible patterns at different contact points could help in the identification of breathing phenomena from artifacts and help decide which sounds have clinical value. The understanding of such effects, however, would require a more sophisticated approach with a larger group of participants, particularly medical doctors. Further studies are warranted to also explore how developing familiarity with additional cues (spectrograms) influences auscultation accuracy. Comparing our data to those obtained by Aviles-Solis et al.⁸ it seems that training is crucial to take advantage of visual clues. Our results show a lower increase compared to the ones mentioned above. This may be the result of a lack of training. Nevertheless, this approach, as suggested by Andr  s et al.¹⁷ may be a promising adjunct to chest auscultation as a diagnostic aid in both clinical and research settings.¹⁸⁻²¹

Conclusions

After more than 200 years since its introduction, auscultation is still one of the most important examinations, however, it is still totally subjective and relies on the ability of a physician to hear out, detect and discriminate different signals using his/her hearing system only. Considering that hearing begins to decline in one's 20s²² and the projection that a significant portion of the medical workforce will be over 50 in the near future, the potential for misinterpretation in auscultation is likely to increase.

Our findings suggest that integrating spectrograms with traditional auditory assessment, albeit based on trends and survey responses, represents a promising avenue with the potential to enhance the accuracy and quality of medical education as well as further diagnosis and monitoring procedures, offering valuable benefits. It seems that this approach necessitates a comprehensive understanding and, consequently, appropriate training.

Nonetheless, it is essential to emphasize the significance of a systemic approach. This approach must be instilled in medical education to ensure its efficiency and the subsequent benefits it can bring to daily practice. There's a compelling case for an educational paradigm shift from a purely subjective, master-apprentice teaching method, where students rely on hearing alone, to a more objective approach that incorporates additional visible cues. Such a shift warrants consideration to enhance medical education and practice. Leveraging visual cues can significantly elevate the precision of basic physiological data interpretation, thereby positively impacting the accuracy of initial diagnoses. This benefit extends not only to young doctors but also to older practitioners whose hearing naturally deteriorates.

The aspect presented in this study is of exceptional significance in light of the evolution not only of telemedicine per se but also the imperative need for legislative adaptation and the transformation of medical education. Physicians, from the outset, must now contend with situations in which they lack direct patient contact, relying solely on data transmitted via remote communication means. In this context, a well-structured systemic approach, commencing from the early stages of medical training, through specialization, and into daily practice, serves as the fundamental underpinning for the successful, efficient, and secure integration of innovative telemedical technologies.

List of abbreviations

GS - Gold Standard
MD - Medical Doctor
p.p. - percentage point
TPR - True Positive Rate
TNR - True Negative Rate

Ethics approval and consent to participate

All studies were approved by the Bioethics Commission at Poznań University of Medical Sciences (approval number 193/18).

Founding Statement

This work was supported by the grant no. FENG.01.01-IP.02-0093/23 from the Polish Agency for Enterprise Development (PARP) as part of the European Funds for a Modern Economy Programme.

Acknowledgements

The Authors would like to thank the Management of Karol Jonscher Clinical Hospital in Poznań, Poznań University of Medical Sciences and the physicians from the Department of Pediatric Pulmonology, Allergology and Clinical Immunology of this hospital for their help in recording and description of the acoustic signals used for GS determination. We would also like to thank all the participants.

References

1. Grzywalski T, Piecuch M, Szajek M, et al. Practical implementation of artificial intelligence algorithms in pulmonary auscultation examination. *European Journal of Pediatrics*. 2019;178(6):883-890. <https://doi.org/10.1007/s00431-019-03363-2>
2. Pasterkamp H, Brand PLP, Everard M, Garcia-Marcos L, Melbye H, Priftis KN. Towards the standardisation of lung sound nomenclature. *European Respiratory Journal*. 2016;47(3):724-732. <https://doi.org/10.1183/13993003.01132-2015>
3. Grzywalski T, Szajek M, Hafke-Dys H, et al. Respiratory system auscultation using machine learning - a big step towards objectivisation? In: *M-Health/e-Health*. European Respiratory Society; 2019: PA2231. <https://doi.org/10.1183/13993003.congress-2019.PA2231>
4. Melbye H, Garcia-Marcos L, Brand P, Everard M, Priftis K, Pasterkamp H. Wheezes, crackles and rhonchi: simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings. *BMJ Open Respiratory Research*. 2016;3(1):e000136. <https://doi.org/10.1136/bmjresp-2016-000136>
5. Welch R, Warren D. Intersensory interactions. In: Boff K, Kaufman L, Thomas J, eds. *Handbook of Perception and Performance*. Vol 1. Wiley, 1981:251-253.
6. McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*. 1976;264(5588):746-748. <https://doi.org/10.1038/264746a0>
7. Sumbly WH, Pollack I. Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*. 1954;26(2):212-215. <https://doi.org/10.1121/1.1907309>
8. Aviles-Solis JC, Storvoll I, Vanbelle S, Melbye H. The use of spectrograms improves the classification of wheezes and crackles in an educational setting. *Scientific Reports*. 2020;10(1):8461. <https://doi.org/10.1038/s41598-020-65354-w>
9. Mangione S, Nieman LZ. Pulmonary Auscultatory Skills During Training in Internal Medicine and Family Practice. *American Journal of Respiratory and Critical Care Medicine*. 1999;159(4):1119-1124. <https://doi.org/10.1164/ajrccm.159.4.9806083>

10. Hafke-Dys H, Bręborowicz A, Kleka P, Kociński J, Biniakowski A. The accuracy of lung auscultation in the practice of physicians and medical students. *PLOS ONE*. 2019;14(8):e0220606. <https://doi.org/10.1371/journal.pone.0220606>
11. Likert R. A technique for the measurement of attitudes. *Archives of Psychology*, 1932;22(140):1-55.
12. Reichert S, Gass R, Brandt C. Analysis of respiratory sounds: state of the art. *Clinical medicine. Circulatory, respiratory and pulmonary medicine*. 2008;2:45-58. <https://doi.org/10.4137/ccrpm.s530>
13. Pramono R, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*. 2017;12(5):e0177926. <https://doi.org/10.1371/journal.pone.0177926>
14. Kiyokawa H, Greenberg M, Shirota K, Pasterkamp H. Auditory Detection of Simulated Crackles in Breath Sounds. *Chest*. 2001;119(6):1886-1892. <https://doi.org/10.1378/chest.119.6.1886>
15. Wilkins RL, Dexter JR, Murphy RLH, DelBono EA. Lung Sound Nomenclature Survey. *Chest*. 1990;98(4):886-889. <https://doi.org/10.1378/chest.98.4.886>
16. Pasterkamp H, Montgomery M, Wiebicke W. Nomenclature used by health care professionals to describe breath sounds in asthma. *Chest*. 1987;92(2):346-352. <https://doi.org/10.1378/chest.92.2.346>
17. Andr  s E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J Med Life*. 2018;11(2):89-106.
18. Kim Y, Hyon Y, Jung S.S. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Sci Rep*. 2021;11:17186. <https://doi.org/10.1038/s41598-021-96724-7>
19. Kevat A, Kalirajah A, Roseby R. Artificial intelligence accuracy in detecting pathological breath sounds in children using digital stethoscopes. *Respir Res*. 2020;21:253. <https://doi.org/10.1186/s12931-020-01523-9>
20. Ahmed S, Mitra DK, Nair H. Digital auscultation as a novel childhood pneumonia diagnostic tool for community clinics in Sylhet, Bangladesh: protocol for a cross-sectional study. *BMJ Open*. 2022;12(2):e059630. <https://doi.org/10.1136/bmjopen-2021-059630>
21. Falter M, Gruwez H, Young J. The future is more than a digital stethoscope. *European Heart Journal - Digital Health*. 2021;2(4):557-558. <https://doi.org/10.1093/ehjdh/ztab077>
22. Hoffman HJ, Dobie RA, Losonczy KG et al. Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years. *JAMA Otolaryngol Head Neck Surg*. 2017;143(3):274-285. <https://doi.org/10.1001/jamaoto.2016.3527>