

Peer Review File

Article information: <https://dx.doi.org/10.21037/jtd-24-310>

**Reviewer A**

In this narrative review, you discuss ways to detect OSA in the general population, based on analysis of speech sounds or snoring.

However, you mainly address anesthesiologists. The beginning of the abstract and the first two sentences of the introduction should be modified to target all specialties. Indeed, removing the anesthesia and pre-anesthetic assessment section would make the summary lighter, more attractive and less restrictive.

Here is a suggestion (and not a requirement)

In Abstract

Background and Objective:

Obstructive sleep apnea (OSA) is a common chronic disorder characterized by repeated breathing pauses during sleep caused by upper airway narrowing or collapse. The gold standard for OSA diagnosis is the polysomnography test, which is time consuming, expensive, and invasive. In recent years, more cost-effective approaches for OSA detection based in predictive value of speech and snoring has emerged.

In this paper, we offer a comprehensive summary of current research progress on the applications of speech or snoring sounds for the automatic detection of OSA and discuss the key challenges that need to be overcome in future research into this novel approach.

The same changes must be made at the beginning of the introduction.

Page 2, #24# correct “into the patient’s

Page 3 #26-29#

A recent study (29) explored the relationship between voice quality and OSA severity; according to its voice analysis results, there were significant differences between OSA and non-OSA individuals in  $f_0$ , jitter percentage, shimmer percentage, harmonic-noise ratio (HNR), and maximum phonation times as the degree of OSA increased.

Can you remove the abbreviations like  $f_0$  “The fundamental frequency”

Speech contains a large amount of characteristic information regarding anatomical structures of the upper respiratory tract. With the rapid development of machine learning in audio analysis, abnormal changes in speech information or snoring during OSA can be exploited to develop an automatic OSA detection approach. Can you summarize or explain in a small diagram or figure the principle of voice and snoring

analysis in the detection of OAS?

This would make this paper more attractive.

**Reply to Reviewer A:**

Thank you for your positive comments and valuable suggestions to improve the quality of our manuscript. According to your nice suggestions, we have made careful modifications to the original manuscript. the detailed corrections are listed below.

**Changes in the text:**

1. In abstract and the beginning of the introduction, we have removed the anesthesia and pre-anesthetic assessment section. See Page 1, line 26-31
2. We have corrected the “f0” into “fundamental frequency”. See Page 3, line 36
3. We have figured the principle of voice and snoring analysis in the detection of OSA. See Page7, line 34-44

**Reviewer B**

This manuscript is a narrative review of very interesting and new topic of obstructive sleep apnea (OSA) and perhaps an easier way of diagnosing it. We are very well aware of the need to find easier and cheaper ways to diagnose OSA in the sleep field. The conventional polysomnography or cardiorespiratory polygraphy are both expensive and time consuming as mentioned in the manuscript. However, both of them are available as ambulatory also which was not mentioned in the manuscript and should be added there. The criteria for OSA was based on AHI + symptoms of OSA. Today we know based on the research that AHI is not a good indicator for OSA and there are OSA patients with low AHI but as much OSA symptoms as in patients with high AHI. So, the different phenotypes of OSA should be taken into account in these new ways of diagnosing OSA. This is one of the limitations of the review also and should be addressed in the challenges.

The manuscript addresses earlier and resent references very well and as such will be a good reference for future studies in this field. The language is fluent and easy going.

**Reply to Reviewer B:**

Thank you for your nice comments on our article. The manuscript has been revised accordingly to clarify the above concerns

**Changes in the text:**

1. In the introduction, we have added the advantage of conventional polysomnography. Page 2, line 25

2. In the key challenges in future research, we have added that “different phenotypes of OSA should be taken into account in these new ways of diagnosing OSA”. Page 8, line 21-24

### **Reviewer C**

This manuscript elevated automatic detection of obstructive sleep apnea based on speech or snoring sounds: a narrative review. It was reflected valuable results of human health. This study reports interesting and original data adding to our knowledge in the subject.

#### GENERAL COMMENTS:

A general comment is that manuscript it is good. A general comment is that grammatical misspelling improvement is necessary throughout the manuscript by the journal rules.

#### SPECIFIC COMMENTS:

Abstract: Please added method as a detailed. Authors should be emphasized which search of database.

Introduction: Please removed line 94-108.

Methods: This chapter is well written.

Results: This chapter is well written

Discussion: This chapter is well written

Conclusion: This chapter is well written.

The authors have tackled a very important and different topic. I congratulate each of them for drawing attention to such an issue.

Considering the prevalence and consequences of OSAS, it has become a public health problem that needs to be prevented. The fact that PSG test is long and costly and not comfortable for the patient makes the diagnosis difficult.

It is very important that the authors present studies on snoring, which is a part of the studies conducted to diagnose OSAS with a noninvasive method. However, if available in the literature, I suggest that they add artificial intelligence-supported studies examining the relationship between snoring and OSAS. if not, I suggest that they make an addition to draw attention to the importance of this issue.

#### **Reply to Reviewer C:**

We sincerely thank the reviewer for careful reading. The revised text reads as follows.

**Changes in the text:**

1. In abstract, we have added method as a detailed. Page 1, line 35-36.
2. As suggested by the reviewer, we have added one literature which examined the relationship between snoring and OSAS with artificial intelligence method. Page 6, line 6-10.

**Reviewer D**

This review on the ability of speech and snoring sounds is interesting although lot of improvements are required for it to be of value to the reader.

1. Article requires proof-reading by someone proficient in English. The use of propositions and phrases throughout the whole manuscript needs to be reviewed (see examples in minor comments).

2. Regarding the ability of speech to predict snoring, authors need to address the following:

- (1) Given the high prevalence of OSA in the general population, any analysis of speech can show a significant result when it is correlated with presence or absence of OSA? How do these studies address this aspect given the small number of subjects utilized in each of these studies.

- (2) OSA is quite prevalent in Hispanics and Chinese-speaking populations. Again, how can one specific technique be mentioned as being useful for prediction of OSA in these populations?

- (3) Authors should recognize that most readers are not familiar with terms like formants, f0, shimmer etc. What amongst formants, harmonics of certain vowels, certain phonemes etc are important? Do they specifically point to a certain attribute of the airway that correlates with OSA (such as vocal tract length) or do they point to the collapsibility (the term compliance has been used and it is not clear as to what compliance is being referred to).

- (4) It also appears that the type of analysis used for correlating with OSA such PCA, LDA, genetic algorithms also make a difference as far as the results. Authors should come up with some recommendations for which of these techniques should be used for such analyses for future researchers in this field.

3. Regarding the summary of studies looking at snore analysis to predict OSA, authors have commented on “feature extraction” and how the small studies that have looked at

different features do not provide any clear example of which feature is likely to provide the best predictive ability for OSA. A table showing each of these features and what they measure would be helpful to those who are not familiar with the technicalities of sound assessment. Similar to speech, authors have not mentioned what age group and BMI of subjects were included in the listed studies in Table 3. Given the increase in AHI with increasing BMI and age, it is difficult to understand what these studies imply if similar populations are not compared. How do the authors address the problem of studying snore features in populations that differ in terms of their age, gender, BMI, comorbidities etc.? Are there any specifications on location of sound sensor or other environmental attributes that can confound these results?

#### 4. Minor comments

Abstract - last line of Background – key challenges “for” future research (not “in”

Key content – first line – and they have been studied extensively “in its automatic screening”. What does “in its automatic screening” mean?

Key content – last line – have been extensively studied “for” the detection of OSA (rather than “in”).

Introduction – 2nd line – “from the prevalence of the general population” (17-22%) needs to be reworded.

Change occurrences to events when describing apneas and hypopneas.

Introduction-line 17 – “the diagnostic rate of OSA in the general population is comparatively low” needs to be reworded.

Use “severity” of OSA instead of degree of OSA.

#### **Reply to Reviewer D:**

**Reply 1:** We tried our best to improve the manuscript and made some changes to the manuscript. These changes will not influence the content and framework of the paper. And we hope the revised manuscript could be acceptable for you.

#### **Reply 2 (1):**

We sincerely appreciate the valuable comments. It is important to acknowledge that not all studies reported positive results. The largest study to date yielded contradictory findings. We also note that small sample sizes may lead to overfitting, thereby resulting in overly optimistic outcomes.

#### **Changes in the text:**

We have added this as a limitation of this review. Page 8, line 26-32

#### **Reply 2 (2):**

Thank you for pointing this out. We also agree that one particular technology can not be always useful for prediction of OSA in different populations. Indeed, there is no standard protocol for speech analysis to identify OSA. Firstly, the types of speech recordings can be classified into 2 main categories: short sentence and vowel/syllable. As for audio feature extraction, various methods are proposed in the literature to identify acoustic features from the temporal, frequency, cepstral, wavelet, and time-frequency domains. At last, Following the selection of features, machine or deep learning algorithms, such as support vector machines, hidden Markov models and convolutional neural networks can be trained to automatically predict OSA.

So, this is a potential limitation of the study.

**Changes in the text:**

We have added this as a limitation. Page 8, line 33-35

**Reply 2 (3):**

Thank you for this suggestion. Usually, the correct choice of features heavily depends on the disease and type of voice recording. According to the current literature, features including time domain feature, frequency domain features and cepstral domain features are commonly used in the prediction of OSA.

**Changes in the text:**

We have added the suggested content to the manuscript through a table.(Table 4)

**Reply 2 (4):**

Thank you for your valuable suggestions. Feature transformation refers to mapping high-dimensional feature data to low-dimensional features, such as PCA, LDA and ICA. PCA aims to identify the primary variance in data and project it into a new coordinate system to maximize data variance. It's useful for data preprocessing, noise filtering, and feature extraction, especially when dealing with multiple correlated features. Additionally, PCA is an unsupervised algorithm and doesn't require class labels. LDA seeks to maximize the differences between classes and minimize the variance within classes. It's beneficial for feature extraction in classification problems, enhancing classifier performance, particularly when classes are distinctly separated. ICA assumes that observed data are composed of multiple independent sources mixed together, aiming to identify these independent sources. Its goal is to find a linear transformation making data independent components in a new space. It's applicable in signal processing, such as separating independent components of mixed signals, and also for feature extraction.

In the literature on predicting OSA using snoring and speech, there appears to be no comparative research on the effectiveness of different dimensionality reduction

methods. Therefore, the authors are currently finding it difficult to make recommendations.

**Reply 3:**

We feel great thanks for your professional review work on our article.

**Changes in the text:**

We have added age group and BMI of subjects in the listed studies in Table 2 and Table 3.

**Reply 4:** Thank you for pointing this out. According to your feedback, we have made revisions point-by-point.

**Changes in the text:**

Abstract – Page 1, line 34

Key content – Page 1, line 39

Key content – Page 2, line 9

Page 5, line 26: Change occurrences to events when describing apneas and hypopneas.

Introduction- Page 2, line 27

Page 3, line 38: Use “severity” of OSA instead of degree of OSA.

**Reviewer E**

**1. Table 1 and “#Methods”**

a. Please check if “articles” should be “original articles”.

Inclusion criteria	English-language papers including articles, reviews, and editorials related to studies in humans
--------------------	--

6 “deep learning” OR “machine learning” OR “artificial intelligence”). All English-language papers  
7 including articles, reviews, and editorials related to studies in humans were included. Each article  
8 was screened for relevance through a reading of the titles and abstracts. Relevant information was

b. It is suggested to specifying the name of two reviewers in table 1.

Selection process	Each article was screened for relevance by reading the titles and abstracts. Relevant information was independently extracted by two reviewers.
-------------------	---

8 was screened for relevance through a reading of the titles and abstracts. Relevant information was  
9 extracted independently by two reviewers (S Cao, M Xia).

Reply 4: Thanks for your careful checks. We have corrected the “articles” into “original

articles”. We have added the name of two reviewers in table 1.

## 2. Please revise the typos.

Jitter	muscle tension resulting in a more open, turbulent glottis Deviations in individual consecutive fo period lengths, which indicates irregular closure and asymmetric vocal-fold vibrations
Spectral slope	It is the measure of slope of the amplitude of the signal and it is computed by linear regression. This feature is used in speech analysis and in identifying speaker from a speech signal
PR800	The ratio of spectral energy above 800 Hz to that below 800 Hz
(3) Cepstral domain features	A cepstrum is obtained by taking the inverse Fourier transform of the logarithm of the spectrum of the signal.
Linear predictive coding (LPC)	Coefficients that best predict the values of the next time point of the audio signal using the values from the

Reply 5: We feel sorry for our carelessness. In our resubmitted manuscript, the typo is revised.

## 3. Please indicate the citation of “Ding et al.” in this sentence.

*Note: References should be cited consecutively and consistently according to the order in which they first appear in the main text.*

*Nevertheless, it should be noted noting that the study by Ding et al. mainly focused on male patients, and thus it would be interesting to further clarify whether this method is suitable for testing the female population.*

Reply 6: Thank you for this suggestion. We have added two references in the revised manuscript.

## 4. Please confirm if a citation is needed in this sentence, as you mentioned “a study”.

*In a study, they found that abnormal resonance, articulation, or phonation was present in 74% of participants with OSA.*

Reply 7: Thanks for your careful checks. The citation for this sentence is (25). We have corrected the “a study” into “this study”.