



**HAL**  
open science

# Comparing Clustering Methods Applied to Tinnitus within a Bootstrapped and Diagnostic-Driven Semi-Supervised Framework

Robin Guillard, Adam Hesses, Louis Korczowski, Alain Londero, Marco  
Congedo, Vincent Loche

► **To cite this version:**

Robin Guillard, Adam Hesses, Louis Korczowski, Alain Londero, Marco Congedo, et al.. Comparing Clustering Methods Applied to Tinnitus within a Bootstrapped and Diagnostic-Driven Semi-Supervised Framework. *Brain Sciences*, 2023, 13 (4), pp.572. 10.3390/brainsci13040572. hal-04261381

**HAL Id: hal-04261381**

**<https://cnrs.hal.science/hal-04261381v1>**

Submitted on 27 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparing clustering methods applied on tinnitus within a bootstrapped and diagnostic-driven semi-supervised framework.

Robin Guillard<sup>1,2,\*</sup>, Adam Hexas<sup>2</sup>, Louis Korczowski<sup>2</sup>, Alain Londero<sup>3</sup>, Marco Congedo<sup>1</sup> and Vincent Loche<sup>4</sup>

<sup>1</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

<sup>2</sup> Siopi, 15 rue des Halles, 75001, Paris, France

<sup>3</sup> Service ORL et CCF, hôpital Européen G.-Pompidou, AP-HP, 20, rue Leblanc, 75015 Paris, France.

<sup>4</sup> Service d'ORL, Hôpital Claude Huriez, CHU Lille, France

\* Correspondence: robin.guillard@grenoble-inp.fr

**Abstract:** The understanding of tinnitus has always been elusive and is largely prevented by its intrinsic heterogeneity. To address this issue, scientific research has aimed at defining stable and easily identifiable subphenotype of tinnitus. This would allow to better disentangle the multiple underlying pathophysiological mechanisms of tinnitus. In this study, 3 dimensionality reduction techniques and 2 clustering methods were benchmarked on a database of 2772 tinnitus patients in order to obtain a reliable segmentation of subphenotypes. In this database, tinnitus patients endotypes (i.e. parts of a population with a condition with distinct underlying mechanisms) were reported when diagnosed by an ENT expert in tinnitus management. This partial labelling of the dataset enabled the design of an original semi-supervised framework. The objective was to perform a benchmark of different clustering methods to get as close as possible to the initial ENT expert endotypes. To do so, two metrics were used : a primary one, the quality of the separation of the endotypes already identified in the database as well as a secondary one, the stability of the obtained clusterings. The relevance of the results was finally reviewed by two ENT experts in tinnitus management. A 20-clusters clustering was selected as the best performing, the most clinically relevant, and the most stable through bootstrapping. This clustering used a T-SNE method as dimensionality reduction technique and a k-means algorithm as clustering method. The characteristics of this clustering are presented in this article.

**Keywords:** Tinnitus; Semi-supervised clustering; Subphenotype; Bootstrap; Benchmark; Expert validation

**Citation:** Guillard, R.; Hexas, A.; Korczowski L.; Londero, A.; Congedo, M.; Loche, V. Tinnitus bootstrapped and diagnostic-driven semi-supervised clustering. *Journal Not Specified* 2023, 1, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

**Copyright:** © 2023 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Tinnitus can be defined as “the conscious awareness of a tonal or composite noise for which there is no identifiable corresponding external acoustic source” [1]. It is a debilitating symptom that affects 14% of the adult population, 2% experiencing a severe form of it [2]. Tinnitus can have disastrous effects on quality of life of people suffering from it [3]. One of the main characteristics of this symptom is its intrinsic heterogeneity [4], a challenge that is currently being tackled by multiple coordinated efforts especially in the European research community, through the Tinnitus Database initiative [5], the European School on Interdisciplinary Tinnitus Research (ESIT) [6] and recently the Unification of Treatments and Interventions for Tinnitus Patients (UNITI) project [7].

It has been suggested that tinnitus heterogeneity is partially responsible for the lack of significant treatment outcomes in various clinical trials for tinnitus [5,8–10]. Also, several sources advocate to stop seeing tinnitus as a symptom that would admit a one size fits all solution [11–13]. Moreover, it has been suggested that tinnitus interventions might prove to be more specifically effective for particular subtypes of tinnitus [14]. A tinnitus subtype refers here to a part of the tinnitus population with a condition that can be used to guide treatment decisions [15,16].



Likewise, there are convergent opinions on the fact that the observed tinnitus heterogeneity derives from a heterogeneity of underlying mechanisms, potentially cumulative [17–19]. In fact, research to identify the underlying pathophysiological mechanisms has led to the identification of several distinct etiologies accounting only for a part of the tinnitus population [20]. Such endotypes (parts of a population with a condition with distinct underlying mechanisms [15,21,22]) are close to have reached consensus. Those endotypes can be diagnosed through specific medical examinations : otosclerosis [23], vestibular schwannoma [24], pulsatile tinnitus [25]. As consequence, the quest to obtain a tinnitus endotypes segmentation appears as a partially solved puzzle [20]. It is interesting to note that for some cases, the link between phenotypes and potential associated pathophysiological mechanism is a matter of intense debate. For example, although it seems that Menière’s disease and hydrops (which can now be objectively diagnosed [26]) are linked, the relation between the two entities is not totally bijective and is still being discussed [27].

Various methods have been attempted to identify a global tinnitus segmentation : defining tinnitus subgroups on the response to a tinnitus treatment [28,29], exploring in a data-driven approach the potential inherent structure of tinnitus symptom profiles (i.e subphenotypes) [30–32], highlighting the existence of a particular subtype in a hypothesis-driven fashion [33]. Such attempts often used algorithmic methods designated as clustering to try to infer such segmentations [15,34]. Tinnitus clustering was performed either on the base of questionnaire information [30,34], on medical imaging [35,36] or audiological explorations [31,37] and more rarely, tinnitus clustering has been performed on the base of free text contents on social networks through natural language processing method [38,39].

An important methodological limitation to such studies is the limited supervision or the absence of supervision used by those clustering methods to compare their outputs to one another. This makes it difficult to concretely define what makes a clustering closer to an ideal segmentation. At best, some studies have used clusters silhouette as clustering quality criteria [30,40]. In addition, few of those studies have large sample sizes, this being another limitation [41,42].

More fundamentally, this raises the question of the definition of the ideal segmentation. Such clusterings create subphenotype segmentations that would convey limited practical clinical value in themselves unless they aimed to get as similar as possible either the optimal subtype segmentation or the optimal endotype segmentation. Here, subphenotype refers to a segmentation based on the distinction between sets of observable characteristics [15,22]. In fact, Genitsaridi legitimately presents in her thesis that nothing proves that the three conceptual segmentations (in subphenotypes, subtypes and endotypes) are identifiable to one another. To illustrate that, it can be plausible that one treatment can work for two different endotypes, or that two different endotypes express the same phenotypes or even that according to individuals, a unique endotype presents itself with varying phenotypes (as for example the subphenotype of tinnitus with otosclerosis is probably not the same before and after surgical intervention).

Table 1, quoted as an excerpt of [43] enables to anchor this important semantic distinction.

Term	Definition
Subgroup	A part of a population (generic term)
Subphenotype	A part of a population with a distinct set of observable characteristics (based on Lötvall et al., 2011)
Subtype	A part of a population with a condition that can be used to guide treatment decisions (based on Saria and Goldenberg, 2015)
Endotype	A part of a population with a condition with distinct underlying mechanisms (based on Anderson, 2008; Lötvall et al., 2011)

**Table 1.** Subgrouping rules semantic distinctions, approved excerpt from (Genitsaridi, 2021).

In this study, the main objective was to establish a tinnitus subphenotype clustering that would get as close as possible to the ideal tinnitus endotypes segmentation. To achieve this goal, the fact that the tinnitus endotype segmentation is already a partially resolved problem, as some etiologies are already known (Menière disease, otosclerosis etc...), was exploited. This was done within an original semi-supervised framework to drive the evaluation of clusterings quality. Thus, setting the resolution of this problem within this partially labelled framework enabled quantitative comparison and thus a proper benchmark of different algorithmic methods.

## 2. Materials and Methods

### 2.1. Population and data acquisition

The investigated database initially comprised 3703 entries of a case history questionnaire recorded by tinnitus patients over the last 20 years of practice by VL, ENT specialist at the Lille University Hospital, France. Entries were paper-filled then computer-recorded and carefully anonymized before being compiled for analysis in a comma separated value (csv) format file. Patients gave their informed consent that their collective entries could be aggregated, anonymized and then used for clinical research purposes.

Features of the database included demographic variables (age, sex, etc), characteristics of the tinnitus (side, frequency, loudness, etc), subjective visual analog scales measuring the impact of tinnitus on patients' quality of life (on anxiety, sleep disturbance etc...), information on patients' lifestyle (sleep duration, practice of sports, etc), as well as, whenever it was possible, the diagnostic of VL. The tables 2 and 3 summarize the main features characteristics over the pre-processed database.

This database offers the perk of gathering a variety of tinnitus-related descriptors as well as, when possible, the medical diagnosis given by an expert medical doctor. The latter offers the opportunity to test diagnostic-driven semi-supervised approaches of clustering.

Table 2, Sample characteristics for quantitative features (N =2772)

	Mean	Std	Min	Max	Missing values (%)
Age (in years)	51,3	14,74	18	95	1,41
Tinnitus Handicap Inventory (THI)	50,98	23,83	0	100	4,4
Tinnitus duration (in months)	67,79	93,12	0	852	9,92
What percentage of time is tinnitus present (in %) ?	66,34	28,19	0	100	16,31
VAS scale Tinnitus annoyance (0 to 10)	6,89	2,17	0	10	4,98
VAS scale Anxiety (0 to 10)	4,89	3,25	0	10	5,81
VAS scale Sleep quality (0 to 10)	5,25	3,35	0	10	6,28
VAS scale Hyperacusis (0 to 10)	5,05	3,41	0	10	5,77
VAS scale Vertigo (0 to 10)	2,36	2,94	0	10	5,77
VAS scale Headache (0 to 10)	2,16	3,23	0	10	0,29
VAS scale Jaw pain (0 to 10)	1,9	2,79	0	10	6,78
VAS scale Cervical pain (0 to 10)	2,63	3,39	0	10	0,4
Khalfa questionnaire score (hyperacusis)	19,72	9,73	0	42	11
HAD questionnaire on anxiety	9,18	4,24	0	21	12,59
HAD questionnaire on depression	6,47	4,6	0	21	12,73
Insomnia Sleep Index (ISI) score	12,39	6,62	0	28	8,44
Epworth scale score	7,62	4,7	0	24	7,65
Sleep onset latency (in min)	35,72	27,23	0	121	11,62

Abbreviations : HAD : Hospital Anxiety and Depression scale, VAS : Visual Analog Scale

Table 2. Sample characteristics : quantitative features of the pre-processed database

Table 3, Sample characteristics for categorical features (N =2772)

Occurrences count	Yes / Choice 1	Sometimes / Choice 2	No / Choice 3	Missing values (%)
Gender (Choice 1 : Female ; Choice 2 : Male )	1491	1226		1,95
Tinnitus apparition ( Choice 1 : brutal, Choice 2 : progressive)	1665	1099		0,3
Tinnitus side (Choice 1 : left, Choice 2 : bilateral, Choice 3 : right)	1015	910	760	3,14
Tinnitus lateralisation (1 : bilateral, 2 : partially lateral, 3 : lateral)	910	743	1032	3,14
Tinnitus intensity (1 : Low, 2 : Medium; 3 : Elevated)	1282	990	500	0,3
Was tinnitus caused by an acoustic trauma ? (Yes, No)	460		2304	0,3
Is the tinnitus pulsatile ? (Yes, No)	231		2533	0,3
Is the tinnitus a narrow band noise ? (Yes, No)	283		2481	0,3
Do you have somato-sensory modulations ? (Yes, No)	446		2318	0,3
Do you have jaw somato-sensory modulations ? (Yes, No)	317		2447	0,3
Do you have neck somato-sensory modulations ? (Yes, No)	256		2508	0,3
Do you have often headaches (Yes, No)	826		1938	0,3
Do you have often cervical pain/rigidity? (Yes, No)	1198		1566	0,3
Do you have trouble falling asleep ? (Yes, Sometimes, No)	972	505	1134	5,81
Do you have nocturnal awakenings ? (Yes, Sometimes, No)	425	530	1644	6,24
Are you feeling tired when awake ? (Yes, Sometimes, No)	653	574	1360	6,67
Do you often snore ? (Yes, No)	1152		1438	6,57
Do you have sleep apnea syndrome ? (Yes, Probably, No)	217	238	2102	7,76
Normal hearing ? (Yes, No)	337		2427	0,3
Sensorineural hearing loss (Yes, No)	1563		656	19,95
Transmissionnal hearing loss (Yes, No)	197		2022	19,95
Notch Hearing Loss (Yes, No)	251		1968	19,95
High Frequencies hearing loss (Yes, No)	146		2073	19,95
Cophosis (Yes, No)	38		2181	19,95

Table 3. Sample characteristics : categorical features of the pre-processed database

Only a minority subset of the dataset was labelled with a diagnostic by the expert doctor. Some endotypes were less represented than others reflecting their unequal prevalence in clinical practice. The different endotypes considered in this study and their prevalence in the dataset are presented in table 4.

Table 4, Endotypes prevalence in the dataset (N=2772)

Endotype	Occurrences	Percentage of the samples
Pulsatile tinnitus	231	8,33%
Otosclerosis	38	1,37%
Menière's disease	35	1,26%
Neurinoma	24	0,87%
Petrous bone fracture	14	0,51%
Eustachian tube dysfunction	7	0,25%
<b>Total</b>	<b>349</b>	<b>12,59%</b>

Table 4. Endotypes prevalence in the database

## 2.2. Preprocessing

Several pre-processing steps were needed to make raw data exploitable. Such steps are described hereafter. Pre-processing steps were conducted with the objective to limit the introduction of additional bias in the dataset.

### 2.2.1. Missing Value Imputation

Case history questionnaire collection was paper-based before being computer-recorded. Such procedure induced, at both steps, missing values in the raw dataset. Missing values imputation is a classic problem in data analysis and machine learning [44,45]. Mainly, two methods have been proposed to remediate to this issue : removing or replacing the missing values. The main objective of missing value management is to limit as much as possible the introduction of additional bias in the dataset while keeping a maximum of patients and features available for analysis. To meet these two objectives, the following method has been implemented :

- First, each patient that had more than 40 percent of missing values in the initial 107 features was deleted to ensure overall data consistency. Through this procedure, out of the 3703 initial patients, 2772 were retained (74,9% of the initial 3703 patients sample).
  - Then, for each feature, if the number of patients for which the feature was missing was higher than 25 percent, the feature was excluded from the analysis. Through this procedure, out of the 107 features considered, 44 were retained (41% features retained).
  - filling the remaining missing values using the MissForest algorithm specially designed to fill missing values with an iterative imputation method based on a random forest [46].
- At the end of this process, the final dataset no longer contained any missing value and consisted of 2772 patients and 44 features.

### 2.2.2. Bootstrap procedure

In the aim to perform an objective comparison of performances between combinations of dimensionality reduction methods and clustering algorithms, stability assessment of the clusterings was performed. This stability assessment was realized through a bootstrap procedure. The dataset was split into a training set and a test set three times, with a resampling with replacement procedure. Such partitionings all contain the entirety of the dataset, split randomly in different ways according the procedure described hereafter.

In classical cross-validation practice, the design of such partitionings results in a higher percentage of data attributed to the training set and a lower proportion to the test set. In the case of the present study, the t-distributed stochastic neighbor embedding (t-SNE) method was included in the benchmark. This method relies on the high dimensional topological structure and local density of the data. Due to this fact, having significantly different number of samples in the training sets and test sets would result in an increased probability of divergence in the high-dimensional spatial distribution of samples and as a consequence an increased divergence in the projection spaces. It was hence necessary for this method to have a 50-50 split of the dataset. Hence, we could not perform a classic cross-validation procedure to be able to compare all combinations of methods tested in the benchmark in the same conditions. This is why it was decided instead to bootstrap three 50-50 partitionings between the training and test sets for all combination of methods.

As the evaluation of clustering performances is based on the diagnoses of an ENT tinnitus expert, it was important that each endotype was equally represented in each partitioning of the data between training and test subdatasets. Hence, a random assignment procedure was applied to each endotypes and the unlabelled data group to equally divide them between the three bootstrapped training and test subdatasets. Following presented results are the average of the performances between the three partitionings of the dataset.

### 2.3. Dimensionality reduction methods

Even after excluding features with missing values, the remaining number of variables is important and justifies the use of dimensionality reduction methods to achieve best performances of the clustering algorithms.

There are methods that try to reduce the number of features ahead of the clustering method while trying to preserve as much information as possible from the initial dataset. In this study, three different dimensionality reduction methods were used and compared : PCA [47], Factor analysis of mixed data (FAMD) [48] and finally t-distributed stochastic neighbor embedding (t-SNE) [49].

The number of dimension is tested repeatedly for a defined range of values and treated as an hyperparameter, thus validating *a posteriori* the quality of the dimensionality reduction method.

Each three methods are hereafter described in further details.

#### 2.3.1. Principal Components Analysis or PCA

PCA is a classical dimension reduction technique that is typically used to summarize large datasets with a small number of descriptors while retaining the maximum amount of



information. PCA is not aiming at selecting the best characteristics while dropping others : instead, it constructs some new characteristics named principal components, based on the original features of the dataset. More formally, PCA is a linear dimension-reduction technique that creates a set of pairwise orthogonal axes that maximize the variance in the data. Thus, this method helps remove redundancy in the new reduced features space.

### 2.3.2. Factor analysis of mixed data or FAMD

PCA was designed for quantitative and non-categorical data analysis. However, our dataset is composed of a mix of categorical and non-categorical variables. FAMD was chosen for our benchmark of dimensionality techniques as it was similar to the PCA but taking into account the mixed aspect of the data. The implementation used is reachable in the Python library Prince (<https://github.com/MaxHalford/prince>).

### 2.3.3. t-distributed stochastic neighbor embedding t-SNE

The last two methods allow an analysis of the global structure of the data, but do not provide an insight into the local topological structure of the data points. Given the objective to create an accurate clustering of endotypes for tinnitus, the local grouping of data points may be of particular importance. Thus, a third dimensionality reduction techniques was considered.

The dimensionality reduction algorithm called t-distributed stochastic neighbor embedding (t-SNE) is an unsupervised learning algorithm. Developed by Laurens van der Maaten and Geoffrey Hinton [50], it enables the analysis of data described in high-dimensional spaces (via a large number of descriptors) to be mapped into a reduced dimensional space. This algorithm is widely used because it facilitates the visualization of data with many descriptors. Through this method, data points which are close in the original highly dimensional space have higher probability of remaining close to one another in the two or three dimensional space of projection. Conversely, data points that are distant in the original space have a low probability of having close representations in the space of projection.

There is a hyperparameter that has a strong impact on the output of the t-SNE algorithm: the perplexity. It characterizes the balance of importance between the local neighborhood structure versus the global neighborhood structure of the data. A large perplexity will lead the algorithm to put the emphasis on the global neighborhood structure of the data. On the opposite, a small perplexity will bring out the local structures of the original data. In this study, the t-SNE method was tested for a wide-range of perplexity values (from 5 to 200).

## 2.4. Clustering the dataset

In this study, a clustering step is applied after the dimension reduction. Thus, the best clustering will be selected and the adequate parameters to reduce dimensionality will be inferred from it.

Two different clustering algorithms were compared: the K-means clustering algorithm and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm [51].

### 2.4.1. K-Means clustering

K-means is a clustering algorithm that aims at partitioning a dataset of observations into k clusters : each data point is attributed to the cluster with the nearest mean. It is a classic method of clustering that has already been tried in several former studies for tinnitus subphenotyping [34,52]. The three main assets of this method are its easiness of interpretation, simplicity of implementation and speed of convergence. Due to the nature of this study, the number of clusters was not predefined. Yet, it is necessary to specify to the algorithm the number k of clusters to find for it to be able to run, a range from 2 to 20

was hence implemented for the grid search (as shown in table 5). The distance measure on numeric attributes was the square Euclidean distance.

#### 2.4.2. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a non-parametric clustering algorithm that takes a given set of points in an initial space, then groups together points that are densely packed together (points with many nearby neighbors), and marks as outliers points that are isolated in low-density regions (whose nearest neighbors are too far away).

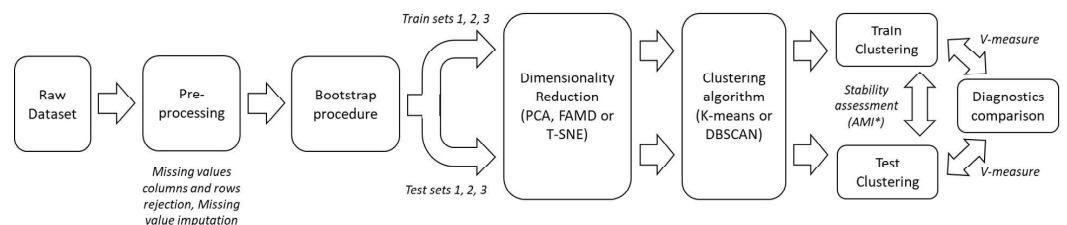
It was selected for its different characteristics from K-means : it is a density method and is able to find arbitrary shaped clusters while K-means produces Voronoi cells shaped clusters. DBSCAN has been used once to try to perform tinnitus subphenotyping [53].

DBSCAN exploration is driven by two hyperparameters: first minsamples, the minimal number of samples (or total weight) in a neighborhood for a point to be considered as a core point. This includes the point itself; then eps (standing for epsilon) : the maximum distance between two samples for one to be considered as in the neighborhood of the other. The ranges of exploration of this hyperparameters for the grid search are shown in table 5.

#### 2.5. Quantified evaluation of the quality of the clusterings

The following procedure has been applied for the benchmark. For each type of clustering and dimension reduction technique tested, best hyperparameters for clustering were identified through grid search exploration presented in table 5. Six combinations of dimensionality reduction techniques and clustering methods were hence tested and compared according to the different mathematical methods presented hereafter.

Then, a shortlist of best clusterings was obtained according to these mathematical criteria and was then presented and analyzed by two tinnitus ENT experts to check their clinical relevance. Out of this comparison, a single clustering was selected and is presented hereafter.



**Figure 1.** Overview of the processing pipeline

To check the quality of the clusterings, three separate quantified evaluation criteria were used.

##### 2.5.1. Silhouette score

The “silhouette measure of cohesion and separation” is a measure for the overall goodness-of-fit of the cluster structure that is described in [54,55]. More precisely, the silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from -1 to +1, with a high value indicating that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is considered appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. Due to its construction, this score will advantage clustering algorithm that will tend to form spherical-like clusters, such as k-means, over less regular shape clustering methods, such as DBSCAN. Moreover, its interpretation is less obvious in the framework of some dimensionality reduction t-SNE algorithm used in this study. Hence, this score will only be shown because it is a classical



way to evaluate clustering, and used in a former study on tinnitus clustering [30] but will not be eventually determinant for the choice of the final clustering.

### 2.5.2. Stability assesment

For each combination of dimensionality reduction technique and clustering algorithm, it is necessary to ensure the stability of the clusterings obtained. To achieve that, a bootstrap procedure was designed.

It is common for artificial structures to emerge which do not correspond to the real separation of the data. In the case of the present study, an output would be deemed stable through bootstrapping if the structure of the clusterings obtained through the application of the same method with the same parameters on the training and test sets of data would be similar. The similarity between two clusterings, which can be identified to two data partitions, can be reliably measured by the adjusted mutual information (AMI) [56]. This measure was hence selected for this study.

To measure the agreement of two data partitions (i.e., clusterings)  $U$  and  $V$ , the AMI takes a value of 1 when the two partitions are identical and 0 when the AMI between two partitions equals the value expected due to chance alone. Its calculation is implemented by this formula :

$$AMI(U, V) = \frac{MI(U, V) - E\{MI(U, V)\}}{\max\{H(U), H(V)\} - E\{MI(U, V)\}} \quad (1)$$

The bootstrap procedure was the following : given a partitioning  $i$  of the data (between 1 and 3), the clustering method is fitted on the (already projected in reduced dimension) training and the test sets. The prediction method of the obtained two models is then applied to the training and the test set. The output of this procedure are four clusterings : two clusterings on the training set (the one fitted on training set and the one fitted on the test set) and two clusterings on the test set (the one fitted on training set and the one fitted on the test set).

The AMI is applied between the two pairs of clusterings stemming from the same datas (i.e., respectively between the two clusterings of the training set and between the two clusterings of the test set). The average of two scores AMI is then taken as the final score.

This method was applied for the K-means clustering yet it was not possible to apply it for the DBSCAN method as it is a transductive method hence in this procedure the fitting method cannot be disentangled from the prediction method. In this case, a proxy of cross-validation was performed by calculating the difference of number of clusters between the training and test clusterings outputs of the DBSCAN applied with the same hyperparameters.

Here, an eligibility threshold of at least 0.7 for the averaged AMI score and a difference of 0 between the number of clusters emerging from the training and test sets was applied to filter for the best solutions of the benchmark.

### 2.5.3. Clustering similarity to the endotype labelling clustering enabled by the partial medical diagnostic of the patients

As a key goal of this study was to take advantage of the fact the endotype segmentation is a partially resolved problem, the quality of the obtained clusterings was evaluated through this partial knowledge. The dataset considered contains, whenever possible, a diagnosis of the endotype of the patient provided by ENT specialist (VL).

An optimal clustering of the whole dataset would be a clustering that would be at least able to separate each of the known endotypes and assign it to a separate cluster. In order to be able to quantify how close a clustering is to an optimal clustering of reference, a metric is needed. Such an evaluation criterion is given by the V-measure, an entropy-based cluster evaluation measure, presented by Rosenberg and Hirschberg [57].

The V-measure between a clustering considered as the reference and a clustering obtained as the output of an experimental algorithmic procedure is defined as the weighted

harmonic mean of two other metrics called homogeneity (h) and completeness (c) and is given by this formula :

$$V - measure = (1 + \beta) * \frac{h * c}{\beta * h + c} \quad (2)$$

Where  $\beta$  is an hyperparameter  $\in [0, \infty[$ , quantifying the relative importance of homogeneity and completeness. The V-measure score is bounded between 0 and 1, a score of 1 corresponds to a perfectly complete and homogeneous matching between clusterings. If  $\beta \leq 1$ , the emphasis is on homogeneity. Otherwise, completeness is highlighted. One can note the analogy between this metric and the F-score used in classification and being composed of precision and recall. A clustering result is homogeneous when all of the clusters it formed contain only data points which are members of a single class. A clustering result satisfies completeness if all the data points that are members of a given class are elements of the same cluster.

In this study, we want to isolate to the best of our ability the known endotypes clusters. After careful examination of the results produced by various  $\beta$  values,  $\beta = 0.1$  was chosen as the better proxy.

The whole process is summarized in Figure 1. The range of exploration for hyperparameters used for the grid search is presented in table 5.

Table 5, Hyperparameters and associated Grid search ranges

Hyperparameter	Associated algorithm(s)	Range of exploration	Steps of exploration
Number of components	PCA, FAMD	[ 2, 20 ]	1
Number of components	t-SNE	[2, 6]	1
Perplexity	t-SNE	[5, 200]	5 by 5 to 40 then to 25 by 25 from 50 to 200
Number of clusters	K-means, DBSCAN*	[ 2, 20 ]	1
Epsilon	DBSCAN	[ 0, 10 ]	0,1
Minsample	DBSCAN	[ 2, 15 ]	1

\*For DBSCAN, the number of cluster cannot be directly assigned, hence only the solutions in the range from 2 to 20 clusters were selected.

Table 5. Range of hyperparameters

## 2.6. Qualitative evaluation of the obtained clusterings

The grid search exploration of all hyperparameters for all combinations of method of the benchmark and their ranking using the evaluation criteria led to a subset of best performing clusterings. To be able to present the overall best achieving clustering obtained by this study, a last step of qualitative evaluation was achieved by ENT tinnitus experts AL and VL.

The subset of the three overall best performing clusterings obtained at the end of the quantitative evaluation procedure was selected and presented to the quantitative evaluation of the appointed experts.

Each clustering was presented with the following information : the combination of hyperparameters and methods leading to this clustering, an illustration of the clustering in the reduced dimensional projection of the dataset when possible (as in the figure 2 of the article), a table as table 7 presenting the specific characteristics of each clusters.

The procedure of the qualitative evaluation of the clusterings was performed by each expert independently. The evaluation consisted in assigning a score between 0 and 5 for each cluster of each clustering. A score 0 means that the cluster appears incoherent and does not relate to anything encountered in their clinical practice, a score 5 means that the cluster fits perfectly a well-known and potentially documented specific subphenotype of patients often encountered in clinical practice.. Then a general appreciation of each clustering was freely given to each clustering.

The final score for each clustering was obtained by averaging the scores for all clusters and for the two experts.

The final best clustering obtained by this procedure is presented in the results section, by figures 2 3 and table 7.

### 3. Results

#### 3.1. Results Overview

Table 6 presents the results of best achieved performances for each combination of methods of the benchmark explored in this study, for the evaluation metrics presented in the methods.

Table 6, Algorithms best performances comparison

Dimensionality reduction method	Clustering algorithm	V-Measure (Beta = 0,1)	Stability assessment (AMI*)	Silhouette score
Principal Component Analysis	K-means	0,142	0,716	0,321
Principal Component Analysis	DBSCAN	0,035	0 clusters difference	-0,17
Factor Analysis of Mixed Data	K-means	0,146	0,707	0,324
Factor Analysis of Mixed Data	DBSCAN	0,049	0 clusters difference	-0,15
t-distributed stochastic neighbor embedding	K-means	<b>0,381</b>	<b>0,728</b>	<b>0,351</b>
t-distributed stochastic neighbor embedding	DBSCAN	0,346	0 clusters difference	0,008

*AMI : adjusted mutual information, calculated between k-means clustering, not possible for DBSCAN clusterings comparison*

Table 6. Performance benchmark

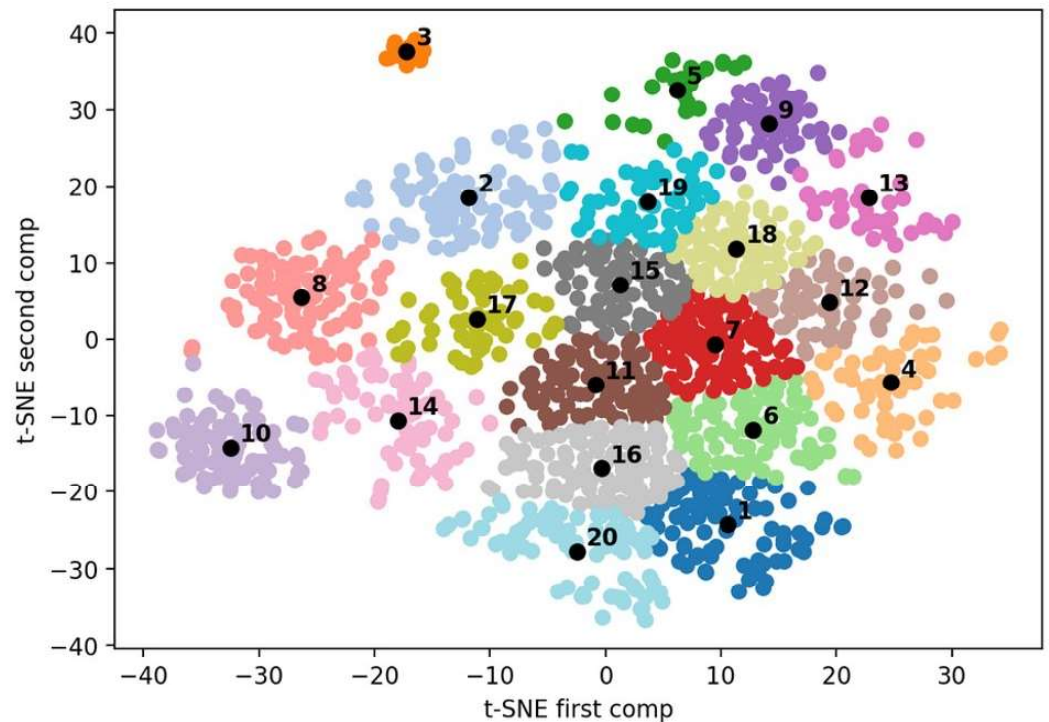


Figure 2. Selected clustering representation of the two-dimensional t-SNE projection. The black points are representing each of the clusters centers (i.e their centroids). The numbers are labels of each cluster. Table 6 is referring to these cluster numbers to characterise each of the clusters.

Best performances on all three evaluation criteria was achieved simultaneously by the combination of t-SNE dimensionality reduction technique and K-means clustering algorithm. These performances were evaluated on the averaged performances on the three partitionings of the dataset obtained by the stratified random subsampling procedure. A set of three best performing clusterings was constituted at the end of this procedure and was then presented to the quantitative evaluation of the two ENT tinnitus experts.

The three best clusterings were all three obtained from the combination of t-SNE dimensionality reduction technique and K-means clustering algorithm. The sets of hyperparameters of these three clusterings were (Clustering 1 : perplexity = 40, number of

components for t-SNE : 2, number of clusters of the k-means : 20); (Clustering 2 : perplexity = 75, number of components for t-SNE : 2, number of clusters of the k-means : 20); (Clustering 3 : perplexity = 5, number of components for t-SNE : 2, number of clusters of the k-means : 18). The three clusterings presented to the ENT tinnitus experts were the ones of the partitionings performing the best V-measure score for the given set of hyperparameters.

The overall averaged qualitative scores between the ENT Tinnitus experts evaluations gave the following scores : Clustering 1 : 4.18/5, Clustering 2 : 4.53/5, Clustering 3 : 4.22/5.

The best performing clustering at the end of this procedure is the Clustering 2 with hyperparameters (perplexity = 75, number of components for t-SNE : 2, number of clusters of the k-means : 20). Both ENT tinnitus experts ranked first this clustering as the most clinically relevant.

As this clustering is projected in a two-dimensional space by the t-SNE method, it was possible to present it in figure 2. The details of the characteristics of each cluster of this clustering are presented in table 7. The endotypes repartition associated with this clustering is presented in figure 3

Table 7, Individual clusters characteristics of the finally selcted clustering

Cluster number	Cluster general description	Number of patients	Positive differentiation		Negative differentiation	
			Feature	Effect size	Feature	Effect size
1	Acoustic trauma without annoyance	83	Initial acoustic trauma	0.96	ISI	-1.3
			Progressive apparition	0.8	VAS Sleep quality	-1.2
			Tinnitus duration (in months)	0.71	Trouble falling asleep	-1.1
			Men	0.69	THI	-1.1
					Tired when awake	-1.0
					HAD anxiety	-0.98
					VAS Anxiety	-0.96
					Nocturnal awakenings	-0.86
					HAD depression	-0.84
					Brutal onset	-0.71
		VAS Tinnitus annoyance	-0.7			
2	Notch hearing losses	91	Notch Hearing Loss	4.6	Sensorineural hearing loss	-1.5
			VAS Anxiety	0.56	Age (in years)	-0.73
			Trouble falling asleep	0.53	Tinnitus pulsatile	-0.32
			THI	0.49	Neck'somato-sensory modulations	-0.31
			Tired when awake	0.43		
			HAD anxiety	0.41		
			ISI	0.35		
			VAS Sleep quality	0.33		
Sleep onset latency	0.33					
3	Cophosis group	18	Cophosis	>10	Sensorineural hearing loss	-1.6
			Hyperacusis Khalifa	0.65	Initial acoustic trauma	-0.45
			VAS Tinnitus annoyance	0.64	Normal hearing	-0.38
			VAS Vertigo	0.63	Jaw somato-sensory modulations	-0.36
			Tinnitus lateralised	0.62	VAS Headache	-0.33
			THI	0.58	Notch Hearing Loss	-0.32
			Age (in years)	0.39	Tinnitus pulsatile	-0.31
			VAS Hyperacusis	0.39		
			Tinnitus duration (in months)	0.34		
			Nocturnal awakenings	0.33		
			% time tinnitus present	0.32		
4	Loud narrow band tinnitus in OSA elderly	57	Narrow band noise	2.7	Headaches	-0.34
			OSA	0.62	Notch Hearing Loss	-0.33
			High tinnitus intensity	0.54	VAS Headache	-0.32
			Age (in years)	0.52		
			VAS Tinnitus annoyance	0.47		
			Sensorineural hearing loss	0.45		
			Snoring	0.44		
			VAS Cervical pain	0.42		
			Cervical pain/rigidity	0.36		
			% time tinnitus present	0.33		
5	Right pulsatile tinnitus	28	Tinnitus pulsatile	3.8	Sensorineural hearing loss	-1.2
			Normal hearing	1.4	Snoring	-0.62
			Tinnitus side right	0.8	Age (in years)	-0.62
			Tinnitus lateralised	0.56	VAS Cervical pain	-0.62
			Notch Hearing Loss	0.31	Men	-0.54
					Cervical pain/rigidity	-0.53
					HAD depression	-0.49
					Initial acoustic trauma	-0.46
					Tinnitus duration (in months)	-0.39
					VAS Anxiety	-0.37
		THI	-0.34			
		VAS Tinnitus annoyance	-0.33			
6	Acoustic trauma with hearing loss	81	Initial acoustic trauma	1.4	Brutal onset	-0.59
			Sensorineural hearing loss	0.65	VAS Headache	-0.58
			Men	0.49	VAS Anxiety	-0.55
			Age (in years)	0.43	Headaches	-0.52
			Tinnitus duration (in months)	0.41	HAD anxiety	-0.5
			Progressive apparition	0.35	Tinnitus lateralised	-0.42
					VAS Sleep quality	-0.42
					Sleep onset latency	-0.42
					Somato-sensory modulations	-0.42
					Normal hearing	-0.4
		VAS Jaw pain	-0.4			
7	Tinnitus and insomnia	90	Sensorineural hearing loss	0.63	VAS Headache	-0.58
			ISI	0.51	Headaches	-0.54
			VAS Sleep quality	0.51	Somato-sensory modulations	-0.46
			Tired when awake	0.42	Normal hearing	-0.4
			Nocturnal awakenings	0.38	Jaw somato-sensory modulations	-0.38
			Trouble falling asleep	0.37	Snoring	-0.38
					Neck'somato-sensory modulations	-0.35
					Notch Hearing Loss	-0.34
					Narrow band noise	-0.32
					Tinnitus pulsatile	-0.32
		High Frequencies hearing loss	-0.31			

385



8	Neck Somatosensory tinnitus	95	Neck somato-sensory modulations Somato-sensory modulations Jaw somato-sensory modulations VAS Cervical pain Cervical pain/rigidity HAD anxiety VAS Jaw pain THI VAS Vertigo ISI VAS Anxiety	5.3 3.2 1.8 0.93 0.74 0.7 0.64 0.6 0.57 0.56 0.55		
9	Pulsatile tinnitus	59	Tinnitus pulsatile Trouble falling asleep Tinnitus lateralised Sensorineural hearing loss Sleep onset latency Age (in years)	4.8 0.68 0.6 0.54 0.39 0.31	High tinnitus intensity Men Initial acoustic trauma Somato-sensory modulations Cervical pain/rigidity Normal hearing Tinnitus duration (in months) Jaw somato-sensory modulations	-0.53 -0.52 -0.47 -0.35 -0.34 -0.34 -0.32 -0.31
10	High frequency hearing loss with tinnitus	85	High Frequencies hearing loss Normal hearing Tired when awake	8.5 2.3 0.33	Sensorineural hearing loss % time tinnitus present Age (in years) Snoring OSA	-1.8 -0.43 -0.39 -0.32 -0.31
11	Sudden hearing loss	82	Brutal onset Sensorineural hearing loss Tinnitus lateralised Headaches	0.94 0.68 0.55 0.39	Progressive onset ISI Somato-sensory modulations VAS Jaw pain Normal hearing Sleep onset latency Jaw somato-sensory modulations HAD anxiety Tired when awake Neck somato-sensory modulations Notch Hearing Loss	-0.8 -0.53 -0.45 -0.45 -0.4 -0.37 -0.37 -0.36 -0.36 -0.35 -0.34
12	Elderly people with bothersome tinnitus and OSA	61	OSA HAD depression Snoring THI % time tinnitus present Hyperacusis Khalifa VAS Tinnitus annoyance Sensorineural hearing loss Epworth Age (in years) VAS Anxiety	1.8 0.73 0.73 0.73 0.69 0.63 0.62 0.62 0.6 0.59 0.59	Somato-sensory modulations Normal hearing Jaw somato-sensory modulations Brutal onset Notch Hearing Loss Tinnitus pulsatile	-0.41 -0.39 -0.37 -0.34 -0.33 -0.32
13	Otosclerosis oriented group	43	Transmissiional hearing loss Progressive onset Tinnitus pulsatile Tinnitus side right Tired when awake VAS Sleep quality ISI VAS Vertigo	4.7 0.59 0.56 0.47 0.39 0.35 0.34 0.32	OSA Brutal onset Initial acoustic trauma Somato-sensory modulations Jaw somato-sensory modulations Notch Hearing Loss Normal hearing	-0.43 -0.42 -0.4 -0.38 -0.36 -0.33 -0.32
14	Somatosensory tinnitus without a nnoyance	73	Somato-sensory modulations Jaw somato-sensory modulations Neck somato-sensory modulations Men	2.8 2.7 1.1 0.58	THI ISI VAS Sleep quality VAS Anxiety VAS Tinnitus annoyance HAD anxiety Trouble falling asleep Hyperacusis Khalifa HAD depression % time tinnitus present VAS Hyperacusis	-0.79 -0.75 -0.67 -0.64 -0.62 -0.61 -0.6 -0.55 -0.54 -0.51 -0.46
15	Tinnitus and headaches	70	Headaches VAS Headache ISI THI Sensorineural hearing loss VAS Sleep quality VAS Tinnitus annoyance % time tinnitus present VAS Hyperacusis Trouble falling asleep Hyperacusis Khalifa	1.2 0.9 0.63 0.62 0.6 0.6 0.55 0.54 0.43 0.42 0.42	Cervical pain/rigidity VAS Cervical pain Somato-sensory modulations Normal hearing Jaw somato-sensory modulations Neck somato-sensory modulations Notch Hearing Loss Tinnitus pulsatile High Frequencies hearing loss	-0.54 -0.53 -0.46 -0.4 -0.37 -0.35 -0.33 -0.32 -0.3

386



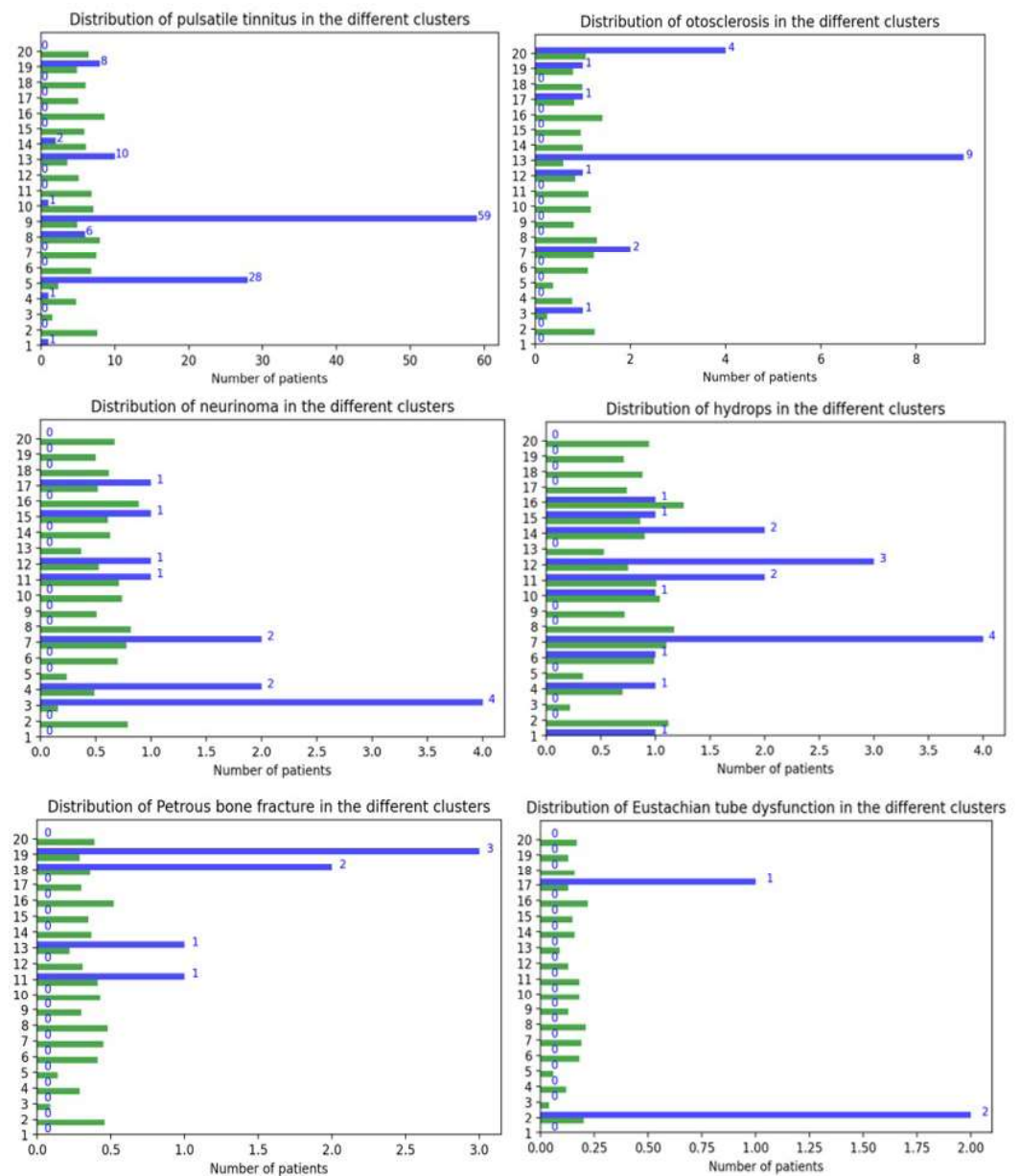
16	Habituated patients	103	Sensorineural hearing loss	0.44	VAS Sleep quality ISI Tired when awake THI VAS Anxiety Nocturnal awakenings VAS Tinnitus annoyance Hyperacusis Khalifa HAD anxiety Trouble falling asleep VAS Hyperacusis	-1.2 -1.1 -1.0 -0.92 -0.88 -0.87 -0.86 -0.83 -0.78 -0.74 -0.72
17	Young people with normal hearing with bothersome tinnitus and hyperacusis	60	Normal hearing VAS Headache HAD anxiety THI VAS Jaw pain VAS Hyperacusis Hyperacusis Khalifa VAS Anxiety	2.0 0.46 0.44 0.43 0.42 0.36 0.33 0.33	Age (in years) Sensorineural hearing loss Snoring Tinnitus lateralised Initial acoustic trauma Notch Hearing Loss Tinnitus pulsatile	-1.3 -1.2 -0.48 -0.42 -0.37 -0.33 -0.32
18	Tinnitus patients with depression	71	HAD depression THI Sleep onset latency HAD anxiety VAS Anxiety ISI VAS Sleep quality VAS Tinnitus annoyance VAS Cervical pain % time tinnitus present Trouble falling asleep	1.6 1.3 1.2 1.2 1.1 1.1 1.1 0.95 0.94 0.83 0.81	Somato-sensory modulations Normal hearing Jaw somato-sensory modulations Neck somato-sensory modulations Epworth Notch Hearing Loss Tinnitus pulsatile Narrow band noise High Frequencies hearing loss	-0.46 -0.4 -0.37 -0.35 -0.33 -0.33 -0.32 -0.3 -0.3
19	Head trauma tinnitus (Majority of petrous bone fractures)	58	VAS Headache VAS Jaw pain Headaches VAS Vertigo HAD anxiety HAD depression THI ISI VAS Sleep quality VAS Anxiety VAS Cervical pain	2.0 1.4 1.3 1.1 1.1 1.1 1.0 1.0 0.98 0.96 0.92	Progressive onset Men Somato-sensory modulations OSA Neck somato-sensory modulations	-0.52 -0.47 -0.35 -0.34 -0.34
20	Tinnitus with OSA seemingly affecting middle ear conduction, not bothersome	77	Transmissional hearing loss Narrow band noise OSA Snoring	1.0 0.86 0.56 0.39	THI VAS Sleep quality ISI VAS Anxiety Hyperacusis Khalifa Tired when awake VAS Hyperacusis VAS Tinnitus annoyance Trouble falling asleep HAD depression HAD anxiety	-1.1 -1.1 -1.1 -1.0 -1.0 -1.0 -0.99 -0.84 -0.81 -0.76 -0.71

THI : Tinnitus Handicap Inventory, ISI : Insomnia Severity Index, HAD : Hospital Anxiety and Depression scale score, OSA : Obstructive sleep apnea, VAS : Visual Analog Scale, Hyperacusis Khalifa : Khalifa questionnaire score on hyperacusis

**Table 7. Characteristics of the clusters of the selected clustering.** Pooled Cohen D was used as effect size measurement. If a feature is present the positive differentiation column, it means that the mean score of this feature for the patients in this cluster is more elevated than the mean of this feature for the rest of the patients, with an Cohen D effect size given in the associated Effect size column. Conversely, if a feature is present the negative differentiation column, it means that the mean score of this feature for the patients in this cluster is less elevated than the mean of this feature for the rest of the patients. As an illustration : in cluster 2, THI is present in the positive differentiation feature. It means that the patients in cluster 2 have a more elevated THI score than average, while as the feature Age is present in the negative differentiation column, it means that the patients of cluster 2 are younger than the rest of the tinnitus patients sample.

#### 4. Discussion

In this study, the main objective was to establish a tinnitus subphenotype clustering (i.e., a clustering of parts of a population with a distinct set of observable characteristics [22,43] that would get as close as possible to the ideal tinnitus endotypes segmentation (i.e., a segmentation between parts of a population with a condition with distinct underlying mechanisms [21,43]. To achieve this goal, a bootstrapped semi-supervised and diagnostic-driven benchmark of combinations of algorithmic methods was performed to obtain the best possible clusterings of a given dataset of tinnitus patients. The final choice presented in the results was selected qualitatively by ENT tinnitus experts among almost equally performing clusterings on the basis of its most accurate clinical relevance. The best clusterings were performed by a combination of t-SNE dimensionality reduction and k-means clustering and successfully separated the known endotypes of tinnitus within different



**Figure 3. Endotypes repartition.** The blue bars are representing the number of samples of a specific endotype present in the cluster of the number it faces. The green bars are representing the number of samples of the same specific endotype that this cluster would contain if the patients of this endotype were randomly assigned to the clusters. For example, on the top left, the cluster 9 contains 59 patients presenting a pulsatile tinnitus, although if the assignment of pulsatile tinnitus patients per cluster was random, we should have expected to have between five and six pulsatile patients in this cluster.

clusters. The important number of clusters (18 to 20) of the final subset of best performing 400  
 clusterings tends to highlight the clinically-observed and reported highly heterogeneous 401  
 nature of tinnitus [4]. The obtained subphenotypes yielded interesting bases for further 402  
 explorations of the underlying pathophysiological mechanisms of tinnitus on specific tinni- 403  
 tus homogeneous subpopulations of patients. This could facilitate the discovery of new 404  
 endotypes of tinnitus. 405

#### 4.1. Final clustering description

As the finally selected clustering was projected in a two-dimensional space by the t-SNE dimensionality reduction step, it can easily be displayed and analyzed. It is important to specify that the presented clustering is applied on the training set of the third partitioning of the data. This is the reason why only half of the whole sample of patients is included in it.

The general spatial organization of this clustering, as well the two others that were evaluated, is composed of a big core (here the grouping of clusters 1, 4, 6, 7, 11, 12, 15, 16, 18, 19 and 20) and gravitating satellites (satellite clusters 2, 3, 5, 8, 9, 10, 13, 14, 17).

Only the finally selected clustering is presented in this study, yet it is important to note that some regularities were observed in all three clusterings presented for the evaluation by the ENT Tinnitus experts.

- The cophosis cluster (cluster 3) is always dense and isolated in all clusterings. These patients have most probably an unilateral cophosis, and it makes sense that it contains neurinoma patients (probably post neuro-radiological procedure).
- The High frequencies hearing loss and tinnitus cluster (cluster 10), the pulsatile group composed of clusters 5 and 9, the somatic group composed of clusters 8 and 14 were always satellites of the clusterings. The pulsatile and somatosensory groups were sometimes partitioned in two as in this clustering and sometimes unified within only one cluster.
- The main core group of patient had always a gradient structure where at one extremity the impact of tinnitus and associated symptoms is very important at one end (as in clusters 12, 15, 18 and 19) and the impact is either mild or absent at the other end (patients not impaired as in clusters 1, 16 and 20). These "poles" have an influence on the general spatial organization : the otosclerosis satellite cluster 13 where the patients are symptomatic and annoyed is close to clusters 12 and 18, and similarly, none disturbed somatic tinnitus patients cluster 14 is close to clusters 16 and 20. It is worth mentioning that the gradient of impact on quality of life of tinnitus is almost aligned and coincident with the gradient of associated quality of sleep measured by ISI, VAS Sleep quality, sleep latency, Nocturnal awakenings... In fact, sleep is disturbed in clusters 7, 12, 15, 18, 19 in opposition to clusters 1, 14, 16 and 20 (except that in cluster 20, patients have OSA but do not seem to be disturbed by it).
- This clustering seemed to highlight and isolate clusters centered on sudden hearing loss (cluster 11), head trauma (cluster 19), and to show an association between hyperacusis and headaches (clusters 15 and 17). It was also interesting to observe that one of the pulsatile tinnitus cluster had the feature "side right" presenting an important effect size. In fact, it is reported in the literature that very often the tinnitus is localised on the right side, due to the important prevalence of venous origin pulsatile tinnitus in the population [58]. Likewise, it is no surprise that for both pulsatile tinnitus groups, effect sizes highlight a over-prevalence of women in these groups as well as lateralized tinnitus.

#### 4.2. Merits

##### 4.2.1. Changing framework from non-supervised to semi-supervised enabling benchmark on performance

This study was the first in the field of tinnitus to leverage the partial existing knowledge of tinnitus heterogeneity to drive the clustering procedure. Although this procedure can probably be largely improved in the future (see the limits section and suggestions for future research), it opens the path to a new framework of analysis on the issue of tackling tinnitus heterogeneity.

The main perspective that this new framework brings for future research is a (partial) ground base for evaluating the performance of tinnitus patient segmentation (and in the case of the present study, tinnitus clustering). This basis enables to quantitatively evaluate

and thus compare the performances between different competing algorithmic methods or combination of methods to solve the task at hand. For the first time a benchmark of performances between several combinations of methods was made possible to evaluate which combination best reproduces the partially-known diagnostic segmentation. It should be noted that the best performing clustering obtained in the present study achieved a 0.386 score on the main criterion V-measure score. Such score is comparable to best performances achieved for spondyloarthritis clustering by a recent study with a similar database size of 3438 patients. The best clustering of this study obtained a V-measure score of 0.588 [59].

#### 4.2.2. Stability assessment

This study is also the first to have performed an equivalent to a cross-validation procedure (here, a bootstrap procedure) aiming at ensuring the quantified stability of its clusterings, following the suggestion of [15]. Here, three partitionings of the dataset in training and test subdatasets were conducted following a resampling with replacement procedure. The evaluation of stability involved the use of the adjusted mutual information score between clusterings, as described in the stability assessment section.

Applying such a procedure is a well-known safeguard against over-fitting of the model and ensures its replicability to some extent.

It is also important to mention that in the present study, adjusted mutual information scores between best performing clusterings were surprisingly quite elevated ( $>0.7$ ), suggesting a quite stable structure of the dataset.

#### 4.2.3. Performing clustering on a large and rich patient questionnaire sample

Such semi-supervised paradigm adoption would hardly have been possible without the importance of the unified dataset analysed in this study. It is also important to stress out that the 44 features reported per sample in this database were covering a wide spectrum of tinnitus patient characteristics from hearing, hyperacusis to somato-sensory modulations and to associated sleep disturbance and depression. Such specter covers the majority of items suggested to be addressed by the review on tinnitus clustering published by [15].

#### 4.2.4. Reproducing in the benchmark already tested dimensionality reduction methods and clustering methods and proposing original combinations

In the present study, the methods (PCA, k-means and DBSCAN) used by [52,53] were included in the benchmark and applied on the same dataset. Likewise, the silhouette score evaluation method presented in [30] was reproduced and presented, although it was not the main criteria for discriminating the performances of the clusterings.

New original methods have also been proposed and included in the benchmark of this study : the Factor Analysis for Mixed Data framework that appeared as a promising improvement to the classic PCA approach. It actually performed slightly better than PCA when combined with either k-means and DBSCAN. Likewise, t-SNE appeared as a relevant dimension reduction technique to be applied on such mixed datas and performed best in combination with k-means clustering method.

### 4.3. Limits

#### 4.3.1. Missing values

A first limit of this study was the non-negligible presence of missing data in the initial dataset that required to exclude some features as well as a non-negligible amount of patients (929 rows), and the use of missing value imputation. Although an adaptative and stable method was used for missing value imputation (MissForest [46]), even more accurate results could have been obtained with a complete and larger dataset.

#### 4.3.2. Limited number of diagnostics

The main aim of the study was to create a clustering that would find and separate the known endotypes of tinnitus. Yet, due to the retrospective nature of this work achieved



on an already acquired dataset, several limitations are to be considered on this matter in the present study. First, although this labelling was performed by a skilled ENT specialist and on the base of objective diagnostic measurement, it cannot be excluded that some diagnostics were wrongly attributed to some patients. Even more, some patients were probably not diagnosed although they were part of a known endotype. The diagnosis partitioning that served as reference in this study is hence limited in its validity.

Secondly, prevalence of some endotypes are naturally rare in the tinnitus population. This dataset was obtained over several years of clinical practice taking care of any patient that sought the help of this practitioner. Consequently, there has been no specific selection procedure applied that would have led to a more advantageous representation of known endotypes in the dataset. As a consequence, the total percentage of labelled data in the dataset was quite low (12,59%) and the prevalence of some endotypes were extremely low (seven patients for eustachian tube dysfunction).

Last, a question that remains open is the boundary between what can be considered as a "known endotype" and a "comorbidity". It could be argued that "ototoxicity", a good potential candidate, should not be included in such work as several different drugs could lead to such attribution of label without affecting similarly the auditory system. Yet, the question poses when it comes to more prevalent subgroups such as presbycusis, acoustic trauma, somato-sensory tinnitus or sudden hearing loss to a lesser extent. Should such conditions be considered as endotypes of tinnitus? More refined criteria to delimit between what should be considered eligible labels or comorbidity should be proposed in the future.

On the other hand, some etiologies chosen in the present study could be considered as heterogeneous. For example, pulsatile tinnitus patients all received the same diagnostic label although it could be argued that one should have separated patients affected by semicircular canal dehiscence, carotid aneurysm, neurovascular conflict... This example highlights the question of the level of granularity one should give to the definition of a tinnitus etiology, a question that goes beyond the scope of the present work. Here, the decision to take pulsatile tinnitus as a unique diagnostic label was a pragmatic choice acknowledging the limited number of labels available to map such pulsatile tinnitus sub-etiologies. The limited size of the database in itself also conditioned the maximum number of clusters that could be allowed for the grid search. Moreover, only a weak constraint from the V-measure was placed on the algorithms on the number of clusters that could include pulsatile tinnitus. As a consequence, the best clustering presented in the results exhibits two different clusters with pulsatile tinnitus patients.

#### 4.3.3. Mono-label clustering

In this study, a methodological choice was to attribute each patient to one endotype and one endotype only. It was hence a mono-label clustering. Yet it could be easily argued that this methodological choice over-simplifies the sometimes complex and intricate nature of clinical presentations of some patients. For example, as the ENT tinnitus experts commented on the results of the final clustering, it was reassuring to observe that some "pulsatile tinnitus" patients were attributed to the cluster where otosclerosis was predominant (cluster 13). Indeed, in some cases otosclerosis can produce a tinnitus presenting with a pulsatile sound.

#### 4.3.4. Biases of the questionnaire

There is no ideal questionnaire to make an optimal anamnesis of tinnitus. However, some good directions were given regarding the important components such questionnaire should at least contain so that the chances to perform a good clustering are maximized [15]. Here, due to the retrospective nature of this study concerning the database acquisition, the questionnaire used, although reasonably complete, did not include all the items suggested by this review.

Another more subtle bias in the questionnaire is the heterogeneity in the level of scrutiny given to some dimensions of the symptomatology compared to others. In the case

of the questionnaire used for this study, only one feature in the dataset was associated with vertigo (VAS scale on vertigo) whereas eight features were devoted to the impact of tinnitus on sleep (ISI, epworth, VAS scale on sleep disturbance, Nocturnal awakenings, sleep latency, quality of the sleep onset, snoring, sleep apnea). This bias was most probably introduced by the fact VL has studied sleep medicine in addition to his ENT specialization. Such over-representation of sleep features in the dataset, comparatively to others, induced biases in the dataset that may not have been solved by the redundancy limitations induced by dimensionality reduction techniques.

It could also of course be argued that the focus given on tinnitus interactions with sleep in the present study brings additional value to this study. Indeed, until now, the description of the nature of those interactions had been limited [60].

#### 4.3.5. Limited range of the hyperparameters for the grid search

In the present study, the grid search parameters were set to explore a range of values that may have constrained the exploration to a space yielding sub optimal performances. A wider range for grid search was explored in testing the dataset before the decision was set on these ranges. From these explorations, it seems that widening the ranges for the dimensionality reduction hyperparameters PCA, FAMD or t-SNE components, as well as for perplexity for t-SNE did not seem to produce better performances. It was also remarkable to note that the best performances on V-measure were always obtained on the lowest possible numbers of components (mostly 2, sometimes 3).

Notably, it is unsure whether the ranges explored for minimum sample and epsilon hyperparameters for DBSCAN were optimal. Above all, the range of exploration for k-means number of clusters was selected as a trade-off and not as an optimal of performance. In fact, simulations of V-measure best performances was an increasing function of maximum number of clusters. This function, for the different combinations of algorithms, had an asymptotic convergence. The asymptotic value as well as the speed of convergence depended on the combination of algorithm, as well as the value of the parameter  $\beta$  of the V-Measure. For the different simulations performed, it was observed that the combination t-SNE + k-means that elicited the best results for this study was also the combination of algorithms that converged the most rapidly to its asymptotic value and had the most elevated asymptotic value.

The maximal value of 20 for the range of exploration of the number of clusters is the result of a trade-off : this number was enough to reach 75% of the asymptotic value for this method, while breaking the samples into groups of around 70 samples on average. It made it possible to characterize the symptomatic specificities of each cluster. Likewise, it brought a fine granularity for the evaluation of clusterings for the ENT tinnitus experts and separated the endotypes without dividing each endotype too many times into different groups. What was surprising was that while simulating the clusterings for a greater number of clusters (for example taking a maximal number of clusters yielding at least 95% of the asymptotic value of the V-measure), the obtained clustering appeared to divide too much the endotypes. This suggests that the V-measure might not be totally optimal in achieving the desired goal. It is important to precise that increasing the value of  $\beta$  instead (to favour less the optimization on homogeneity) resulted in another problem : the results regrouped different (and clinically incompatible) endotypes in the same clusters.

#### 4.3.6. Limited stability assessment for the DBSCAN output

In the present study, a rigorous bootstrap procedure was applied to assess the stability of the k-means clustering method output. Yet, such method could not be applied for DBSCAN algorithm. In fact, due to the transductive nature of DBSCAN (i.e. it cannot predict the labels of new datas), the fit and predict methods of DBSCAN cannot be disentangled so it was not possible to make a prediction of labels of a subset of samples with a model fitted on the other sample subset. Counting the difference of clusters in the outputs proposed by the DBSCAN was used as a replacement for the stability assessment : indeed



the number of clusters of the output clustering is a free parameter of DBSCAN. Yet it can be questioned whether such replacement constitutes the best possible stability assessment for a transductive method.

#### 4.4. Suggestions for future research

##### 4.4.1. Widen hyperparameters search on larger databases

As stated in the limits, our grid search ranges were limited, especially for the maximal number of clusters which was chosen as trade-off considering the sample size of the database.

On the other hand, it seems intuitive to induce that if six different endotypes of tinnitus are already present in 12,59% of the dataset as in the present study, it could be expected that the total number of clusters of an ideal clustering should be above 20. Future research should hence work to apply such techniques on wider datasets to enable widening the range of hyperparameters.

Similarly, the set of features of our database was limited and was not covering all the important dimensions advised in [15]. Future research should constitute their databases prospectively and hence actively shape the questionnaires so as to cover globally the anamnesis of a tinnitus patient. Focus should also be put on not over-weighting the exploration of some dimensions compared to others (as illustrated in the limits for vertigo and sleep in our case).

##### 4.4.2. Change framework to longitudinal data

A supplementary important suggestion for future research would be to switch from an initial dataset composed of a unique point in time (and thus a unique completions of a given questionnaire) per patient to a longitudinal dataset in which each patient should answer to a given set of questionnaires at some strategic points in time. In fact, the clustering chosen to be presented in this study tries to capture the time dimension by the feature "tinnitus duration". Yet such metric poorly captures the level of intrusiveness of a patient which can evolve at different speeds according to the psychological and behavioural adaptation to the tinnitus and to the efficiency of therapeutic interventions. Likewise, some clusters tend to reflect such a difference of temporality, with clusters that evoke a state of initial crisis and others where one could hypothesize that patients are habituated to their condition (i.e. where the tinnitus annoyance is significantly lower).

It would naturally require more resources to lead such a study over a longer time frame to enable to map trajectories of patient symptoms. But such initiatives could be led by digital mobile platforms such as TrackYourTinnitus [61] or Siopi [62]. Also, it would require to reorganize the framework of analysis of such characterization of each patient so as to achieve a clustering of different trajectories of patients rather than a clustering of questionnaire entries.

##### 4.4.3. Going further in the semi-supervised framework

In the present study, the semi-supervision of the algorithm only intervenes in the evaluation of the clusterings obtained at the end of a systematic unsupervised process set in motion on a constrained grid search exploration space. This is the reason why the description of the procedure is only referred as a semi-supervised *framework*. Yet, future research should investigate on how to include semi-supervision directly into the pipeline of clustering. It could be at the level of the dimensionality reduction or at the level of the clustering algorithms [63,64].

##### 4.4.4. Going from mono-label to multi-label clustering and beyond

As suggested among the limits of the present study, the methodological choice of applying a mono-label clustering might not be well fitted to describe the clinical reality of patient symptomatology. When defining endotypes more restrictively, it becomes possible for a patient to be part of several endotypes : for example a patient presenting presbycusis with

somato-sensory tinnitus due to a jaw instability would hence be assigned to two endotypes as these two reasons in themselves can lead to an emergence of tinnitus.

It would hence appear useful for future research to adapt their clustering framework to perform a multi-label clustering.

To go even further, it can be questioned that segmentation, due to its discrete and non-continuous nature, is the best methodological tool to address tinnitus heterogeneity. Another possibility to address the heterogeneity problem would be to tackle it with a more continuous and local approach by only looking locally for best neighbors to a given patient. Such approach has been attempted by Siopi for its mutual-help community could involve metric learning methods (supervised or weakly supervised) such as Metric Learning with Application for Clustering with Side Information (MMC) [65], Large Margin Nearest Neighbor Metric Learning (LMNN) [66] or other deep metric learning methods.

#### 4.4.5. Apply the semi-supervised clustering to other labels

It seems important to point out that a semi-supervised framework for tinnitus clustering can also be applied taking in reference something different from endotypes labels. For example, taking individual treatment responses as labels could also be very interesting to try to define subtypes of tinnitus.

#### 4.4.6. Explore other dimensionality reduction techniques and clustering techniques

Here, methods that had been tested in past studies were replicated and two new original dimensionality reduction techniques were introduced. Yet, a great amount of other dimensionality reduction techniques and clustering methods could be tested in the future so as to outperform the results obtained in the present study. Interesting candidates would be HDBSCAN [67] for which in some references its transductive nature could be overcome for cross-validation ([https://hdbscan.readthedocs.io/en/latest/prediction\\_tutorial.html](https://hdbscan.readthedocs.io/en/latest/prediction_tutorial.html)). This method has the advantage not to tend to form Voronoi cell shaped clusters that tend to have similar numbers of samples per clusters (like DBSCAN). Such methods appear to be more adapted in the case of subgrouping samples in groups that can have diverging prevalence in the population.

Additionally, a very good semi-supervised candidate could be the Heterogeneity through Discriminative Analysis (HYDRA) method [68] that proved to be quite efficient on subtyping Schizophrenia [69,70]. However, this method would be best fitted to be applied on neuroimaging datasets rather than questionnaire datasets.

## 5. Conclusion

The present study is aimed at presenting a new semi-supervised framework to bring guidance when facing the issue of heterogeneity in the population of tinnitus patients. It was the first to achieve such a clustering while enabling quantified comparisons between performance of different algorithmic combinations. Through this process and through a bootstrap procedure for stability assessment, a 20-cluster solution was selected and presented. Among this solution, most clusters were confirmed by ENT tinnitus experts to convey strong clinical relevance. Such clusters define homogeneous subphenotypes of patients. Those are relevant to be studied independently to look for objective signatures of different potential underlying pathophysiological mechanisms.

**Author Contributions:** Conceptualization, R.G., A.H. and L.K.; methodology, R.G., A.H., A.L., L.K., M.C. and V.L.; software, A.H. and R.G.; validation, R.G., A.H., A.L. and V.L.; formal analysis, A.H. and R.G.; data curation, A.H., V.L. and R.G.; writing—original draft preparation, A.H. and R.G.; writing—review and editing, R.G.; visualization, A.H. and R.G.; supervision, R.G., V.L. and A.L.; project administration, R.G.; funding acquisition, R.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Felicia and Jean-Jacques Lopez-Loreta Fondation, as well as BPI France bourse French Tech Emergence, that both invested in Siopi who coordinated this research.

**Informed Consent Statement:** Informed consent was obtained from all subjects that answered the questionnaire of this study. Anonimization of the dataset was performed prior to any scientific work involved in this study.

**Data Availability Statement:** Anonimized dataset is available upon request.

**Acknowledgments:** The authors wish to thank Mr Nadji Souilamas for his contribution in the early steps of conception of this study. Likewise, the authors wish to thank French AFREPA association for their constructive feedback that helped shape this research work. Last, the authors would like to thank warmly Mrs Marie-Anne Mirabeau for improving the language quality of the manuscript.

**Conflicts of Interest:** R.G., A.H. and L.K. are employees of Siopi, who's mission is to create a mutual-help community to help tinnitus patients. One of the attributed mission of Siopi is to bring light on the intrinsic heterogeneity of tinnitus. R.G. and L.K. are shareholders of Siopi.

## Abbreviations

The following abbreviations are used in this manuscript:

THI	Tinnitus Handicap Inventory
VAS	Visual Analog Scale
HAD	Hospital Anxiety and Depression scale
ISI	Insomnia Sleep Index
ENT	Ear Nose and Throat
PCA	Principal Component Analysis
FAMD	Factor Analysis of Mixed Data
t-SNE	t-distributed stochastic neighbor embedding
DBSCAN	. Density-Based Spatial Clustering of Applications with Noise
AMI	Adjusted Mutual Information
OSA	Obstructive Sleep Apnea

## References

- De Ridder, D.; Schlee, W.; Vanneste, S.; Londero, A.; Weisz, N.; Kleinjung, T.; Shekhawat, G.S.; Elgoyhen, A.B.; Song, J.J.; Andersson, G.; et al. Chapter 1 - Tinnitus and tinnitus disorder: Theoretical and operational definitions (an international multidisciplinary proposal). In *Progress in Brain Research*; Schlee, W.; Langguth, B.; Kleinjung, T.; Vanneste, S.; De Ridder, D., Eds.; Elsevier, 2021; Vol. 260, *Tinnitus - An Interdisciplinary Approach Towards Individualized Treatment: From Heterogeneity to Personalized Medicine*, pp. 1–25. <https://doi.org/10.1016/bs.pbr.2020.12.002>.
- Jarach, C.M.; Lugo, A.; Scala, M.; van den Brandt, P.A.; Cederroth, C.R.; Odone, A.; Garavello, W.; Schlee, W.; Langguth, B.; Gallus, S. Global Prevalence and Incidence of Tinnitus: A Systematic Review and Meta-analysis. *JAMA Neurology* **2022**, *79*, 888–900. <https://doi.org/10.1001/jamaneurol.2022.2189>.
- Baguley, D.; McFerran, D.; Hall, D. Tinnitus. *The Lancet* **2013**, *382*, 1600–1607. [https://doi.org/10.1016/S0140-6736\(13\)60142-7](https://doi.org/10.1016/S0140-6736(13)60142-7).
- Cederroth, C.R.; Gallus, S.; Hall, D.A.; Kleinjung, T.; Langguth, B.; Maruotti, A.; Meyer, M.; Norena, A.; Probst, T.; Pryss, R.; et al. Editorial: Towards an Understanding of Tinnitus Heterogeneity. *Frontiers in Aging Neuroscience* **2019**, *11*, 53. <https://doi.org/10.3389/fnagi.2019.00053>.
- Landgrebe, M.; Zeman, F.; Koller, M.; Eberl, Y.; Mohr, M.; Reiter, J.; Staudinger, S.; Hajak, G.; Langguth, B. The Tinnitus Research Initiative (TRI) database: A new approach for delineation of tinnitus subtypes and generation of predictors for treatment outcome. *BMC Medical Informatics and Decision Making* **2010**, *10*, 42. <https://doi.org/10.1186/1472-6947-10-42>.
- Schlee, W.; Hall, D.A.; Canlon, B.; Cima, R.F.F.; de Kleine, E.; Hauck, F.; Huber, A.; Gallus, S.; Kleinjung, T.; Kypraios, T.; et al. Innovations in Doctoral Training and Research on Tinnitus: The European School on Interdisciplinary Tinnitus Research (ESIT) Perspective. *Frontiers in Aging Neuroscience* **2018**, *9*, 447. <https://doi.org/10.3389/fnagi.2017.00447>.
- Schlee, W.; Schoisswohl, S.; Staudinger, S.; Schiller, A.; Lehner, A.; Langguth, B.; Schecklmann, M.; Simoes, J.; Neff, P.; Marcrum, S.C.; et al. Chapter 21 - Towards a unification of treatments and interventions for tinnitus patients: The EU research and innovation action UNITI. In *Progress in Brain Research*; Schlee, W.; Langguth, B.; Kleinjung, T.; Vanneste, S.; De Ridder, D., Eds.; Elsevier, 2021; Vol. 260, *Tinnitus - An Interdisciplinary Approach Towards Individualized Treatment: From Heterogeneity to Personalized Medicine*, pp. 441–451. <https://doi.org/10.1016/bs.pbr.2020.12.005>.
- Langguth, B. Treatment of tinnitus. *Current Opinion in Otolaryngology & Head and Neck Surgery* **2015**, *23*, 361–368. <https://doi.org/10.1097/MOO.0000000000000185>.
- McFerran, D.J.; Stockdale, D.; Holme, R.; Large, C.H.; Baguley, D.M. Why Is There No Cure for Tinnitus? *Frontiers in Neuroscience* **2019**, *13*.
- Londero, A.; Hall, D.A. Call for an evidence-based consensus on outcome reporting in tinnitus intervention studies. *Frontiers in Medicine* **2017**, *p.* 42.

11. Hall, D.A.; Hibbert, A.; Smith, H.; Haider, H.F.; Londero, A.; Mazurek, B.; Fackrell, K. One Size Does Not Fit All: Developing Common Standards for Outcomes in Early-Phase Clinical Trials of Sound-, Psychology-, and Pharmacology-Based Interventions for Chronic Subjective Tinnitus in Adults. *Trends in Hearing* **2019**, *23*, 2331216518824827. Publisher: SAGE Publications Inc, <https://doi.org/10.1177/2331216518824827>. 756-759
12. Beukes, E.W.; Manchaiah, V.; Allen, P.M.; Andersson, G.; Baguley, D.M. Chapter 4 - Exploring tinnitus heterogeneity. In *Progress in Brain Research*; Schlee, W.; Langguth, B.; Kleinjung, T.; Vanneste, S.; De Ridder, D., Eds.; Elsevier, 2021; Vol. 260, *Tinnitus - An Interdisciplinary Approach Towards Individualized Treatment: From Heterogeneity to Personalized Medicine*, pp. 79–99. <https://doi.org/10.1016/bs.pbr.2020.05.022>. 760-763
13. Stropahl. Why the One-Size-Fits-All Approach to Tinnitus is Not Successful, 2022. Stropahl M. Why the one-size-fits-all approach to tinnitus is not successful. *Hearing Review*. 2022;29(9):28-31. 764-765
14. D'Arcy, S.; Hamilton, C.; Hughes, S.; Hall, D.A.; Vanneste, S.; Langguth, B.; Conlon, B. Bi-modal stimulation in the treatment of tinnitus: a study protocol for an exploratory trial to optimise stimulation parameters and patient subtyping. *BMJ Open* **2017**, *7*, e018465. Publisher: British Medical Journal Publishing Group Section: Neurology, <https://doi.org/10.1136/bmjopen-2017-018465>. 766-769
15. Genitsaridi, E.; Hoare, D.J.; Kypraios, T.; Hall, D.A. A Review and a Framework of Variables for Defining and Characterizing Tinnitus Subphenotypes. *Brain Sciences* **2020**, *10*, 938. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, <https://doi.org/10.3390/brainsci10120938>. 770-772
16. Saria, S.; Goldenberg, A. Subtyping: What It is and Its Role in Precision Medicine. *IEEE Intelligent Systems* **2015**, *30*, 70–75. Conference Name: IEEE Intelligent Systems, <https://doi.org/10.1109/MIS.2015.60>. 773-774
17. Landgrebe, M.; Azevedo, A.; Baguley, D.; Bauer, C.; Cacace, A.; Coelho, C.; Dornhoffer, J.; Figueiredo, R.; Flor, H.; Hajak, G.; et al. Methodological aspects of clinical trials in tinnitus: A proposal for an international standard. *Journal of Psychosomatic Research* **2012**, *73*, 112–121. <https://doi.org/10.1016/j.jpsychores.2012.05.002>. 775-777
18. Genitsaridi, E.; Partyka, M.; Gallus, S.; Lopez-Escamez, J.A.; Schecklmann, M.; Mielczarek, M.; Trpchevska, N.; Santacruz, J.L.; Schoisswohl, S.; Riha, C.; et al. Standardised profiling for tinnitus research: The European School for Interdisciplinary Tinnitus Research Screening Questionnaire (ESIT-SQ). *Hearing Research* **2019**, *377*, 353–359. <https://doi.org/10.1016/j.heares.2019.02.017>. 778-780
19. Beukes, E.W.; Baguley, D.M.; Manchaiah, V.; Andersson, G.; Allen, P.M.; Kaldo, V.; Jacquemin, L.; Lourenco, M.P.C.G.; Onozuka, J.; Stockdale, D.; et al. Investigating tinnitus subgroups based on hearing-related difficulties. *International Journal of Clinical Practice* **2021**, *75*, e14684. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijcp.14684>, <https://doi.org/10.1111/ijcp.14684>. 781-783
20. Makar, S.K. Etiology and Pathophysiology of Tinnitus - A Systematic Review. *The International Tinnitus Journal* **2021**, *25*. <https://doi.org/10.5935/0946-5448.20210015>. 784-785
21. Anderson, G.P. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *The Lancet* **2008**, *372*, 1107–1119. [https://doi.org/10.1016/S0140-6736\(08\)61452-X](https://doi.org/10.1016/S0140-6736(08)61452-X). 786-787
22. Lötvall, J.; Akdis, C.A.; Bacharier, L.B.; Bjermer, L.; Casale, T.B.; Custovic, A.; Lemanske, R.F.; Wardlaw, A.J.; Wenzel, S.E.; Greenberger, P.A. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. *Journal of Allergy and Clinical Immunology* **2011**, *127*, 355–360. <https://doi.org/10.1016/j.jaci.2010.11.037>. 788-790
23. Virk, J.S.; Singh, A.; Lingam, R.K. The Role of Imaging in the Diagnosis and Management of Otosclerosis. *Otology & Neurotology* **2013**, *34*, e55. <https://doi.org/10.1097/MAO.0b013e318298ac96>. 791-792
24. Goldbrunner, R.; Weller, M.; Regis, J.; Lund-Johansen, M.; Stavrinou, P.; Reuss, D.; Evans, D.G.; Lefranc, F.; Sallabanda, K.; Falini, A.; et al. EANO guideline on the diagnosis and treatment of vestibular schwannoma. *Neuro-Oncology* **2020**, *22*, 31–45. <https://doi.org/10.1093/neuonc/noz153>. 793-795
25. Houdart, E. Acouphènes pulsatiles chroniques : diagnostic étiologique et traitement endovasculaire. *JMV-Journal de Médecine Vasculaire* **2017**, *42*, 94. <https://doi.org/10.1016/j.jdmv.2017.01.096>. 796-797
26. Lopez-Escamez, J.A.; Attyé, A. Systematic review of magnetic resonance imaging for diagnosis of Meniere disease. *Journal of vestibular research* **2019**, *29*, 121–129. <https://doi.org/10.3233/ves-180646>. 798-799
27. Gluth, M.B. On the relationship between Meniere's disease and endolymphatic hydrops. *Otology & Neurotology* **2020**, *41*, 242–249. 800
28. Côté, C.; Baril, I.; Morency, C.; Montminy, S.; Couture, M.; Leblond, J.; Roos, M.; Roy, J.S. Long-Term Effects of a Multimodal Physiotherapy Program on the Severity of Somatosensory Tinnitus and Identification of Clinical Indicators Predicting Favorable Outcomes of the Program. *Journal of the American Academy of Audiology* **2019**, *30*, 720–730. Publisher: Thieme Medical Publishers, <https://doi.org/10.3766/jaaa.17147>. 801-804
29. Klooster, F.; Arnold, R.; Hofman, R.; Burgerhof, J.; van Dijk, P. Models to predict positive and negative effects of cochlear implantation on tinnitus. *Laryngoscope Investigative Otolaryngology* **2019**, *4*, 138–142. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/lio2.224>, <https://doi.org/10.1002/lio2.224>. 805-807
30. van den Berge, M.J.C.; Free, R.H.; Arnold, R.; de Kleine, E.; Hofman, R.; van Dijk, J.M.C.; van Dijk, P. Cluster Analysis to Identify Possible Subgroups in Tinnitus Patients. *Frontiers in Neurology* **2017**, *8*, 115. <https://doi.org/10.3389/fneur.2017.00115>. 808-809
31. Langguth, B.; Landgrebe, M.; Schlee, W.; Schecklmann, M.; Vielsmeier, V.; Steffens, T.; Staudinger, S.; Frick, H.; Frick, U. Different Patterns of Hearing Loss among Tinnitus Patients: A Latent Class Analysis of a Large Sample. *Frontiers in Neurology* **2017**, *8*. <https://doi.org/10.3389/fneur.2017.00046>. 810-812



32. Tyler, R.; Coelho, C.; Tao, P.; Ji, H.; Noble, W.; Gehringer, A.; Gogel, S. Identifying Tinnitus Subgroups With Cluster Analysis. *American Journal of Audiology* **2008**, *17*, S176–S184. Publisher: American Speech-Language-Hearing Association, [https://doi.org/10.1044/1059-0889\(2008/07-0044\)](https://doi.org/10.1044/1059-0889(2008/07-0044)). 813
33. Vielsmeier, V.; Strutz, J.; Kleinjung, T.; Schecklmann, M.; Kreuzer, P.M.; Landgrebe, M.; Langguth, B. Temporomandibular Joint Disorder Complaints in Tinnitus: Further Hints for a Putative Tinnitus Subtype. *PLOS ONE* **2012**, *7*, e38887. Publisher: Public Library of Science, <https://doi.org/10.1371/journal.pone.0038887>. 814
34. Niemann, U.; Brueggemann, P.; Boecking, B.; Mebus, W.; Rose, M.; Spiliopoulou, M.; Mazurek, B. Phenotyping chronic tinnitus patients using self-report questionnaire data: cluster analysis and visual comparison. *Scientific Reports* **2020**, *10*, 16411. <https://doi.org/10.1038/s41598-020-73402-8>. 815
35. Schecklmann, M.; Lehner, A.; Poepl, T.B.; Kreuzer, P.M.; Hajak, G.; Landgrebe, M.; Langguth, B. Cluster analysis for identifying sub-types of tinnitus: A positron emission tomography and voxel-based morphometry study. *Brain Research* **2012**, *1485*, 3–9. <https://doi.org/10.1016/j.brainres.2012.05.013>. 816
36. Zimmerman, B.J.; Abraham, I.; Schmidt, S.A.; Baryshnikov, Y.; Husain, F.T. Dissociating tinnitus patients from healthy controls using resting-state cyclicality analysis and clustering. *Network Neuroscience* **2019**, *3*, 67–89. [https://doi.org/10.1162/netn\\_a\\_00053](https://doi.org/10.1162/netn_a_00053). 817
37. Santacruz, J.L.; de Kleine, E.; van Dijk, P. Investigating the relation between minimum masking levels and hearing thresholds for tinnitus subtyping. In *Progress in Brain Research*; Elsevier, 2021; Vol. 263, pp. 81–94. <https://doi.org/10.1016/bs.pbr.2021.04.011>. 818
38. Palacios, G.; Noreña, A.; Londero, A. Assessing the Heterogeneity of Complaints Related to Tinnitus and Hyperacusis from an Unsupervised Machine Learning Approach: An Exploratory Study. *Audiology and Neurotology* **2020**, *25*, 174–189. <https://doi.org/10.1159/000504741>. 819
39. Manchaiah, V.; Londero, A.; Deshpande, A.K.; Revel, M.; Palacios, G.; Boyd, R.L.; Ratinaud, P. Online Discussions About Tinnitus: What Can We Learn From Natural Language Processing of Reddit Posts. *American journal of audiology* **2022**, *31*, 993–1002. [https://doi.org/10.1044/2021\\_aja-21-00158](https://doi.org/10.1044/2021_aja-21-00158). 820
40. Handl, J.; Knowles, J.; Kell, D.B. Computational cluster validation in post-genomic data analysis. *Bioinformatics* **2005**, *21*, 3201–3212. <https://doi.org/10.1093/bioinformatics/bti517>. 821
41. Newman, C.W.; Wharton, J.A.; Jacobson, G.P. Self-focused and somatic attention in patients with tinnitus. *Journal of the American Academy of Audiology* **1997**, *8*, 143–149. 822
42. Rizzardo, R.; Savastano, M.; Maron, M.B.; Mangialaio, M.; Salvadori, L. Psychological distress in patients with tinnitus. *The Journal of otolaryngology* **1998**, *27*, 21–25. 823
43. Genitsaridi, E. Novel Approaches for Tinnitus Subphenotyping: Evidence Synthesis, Standardised Assessment, and Supervised and Unsupervised Machine Learning Applications **2021**. p. 338. 824
44. Lin, W.C.; Tsai, C.F. Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review* **2020**, *53*, 1487–1509. <https://doi.org/10.1007/s10462-019-09709-4>. 825
45. Armina, R.; Zain, A.M.; Ali, N.A.; Sallehuddin, R. A Review On Missing Value Estimation Using Imputation Algorithm. *Journal of Physics: Conference Series* **2017**, *892*, 012004. Publisher: IOP Publishing, <https://doi.org/10.1088/1742-6596/892/1/012004>. 826
46. Stekhoven, D.J.; Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112–118. <https://doi.org/10.1093/bioinformatics/btr597>. 827
47. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **1987**, *2*, 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9). 828
48. Pagès, J. Analyse factorielle de données mixtes. *Revue de statistique appliquée* **2004**, *tome 52*, p. 93–111. 829
49. Gisbrecht, A.; Schulz, A.; Hammer, B. Parametric nonlinear dimensionality reduction using kernel t-SNE. *Neurocomputing* **2015**, *147*, 71–82. <https://doi.org/10.1016/j.neucom.2013.11.045>. 830
50. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9*. 831
51. Khan, K.; Rehman, S.U.; Aziz, K.; Fong, S.; Sarasvady, S. DBSCAN: Past, present and future. In Proceedings of the The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), 2014, pp. 232–238. <https://doi.org/10.1109/ICADIWT.2014.6814687>. 832
52. Erlandsson, S.I.; Rubinstein, B.; Axelsson, A.; Carlsson, S.G. Psychological dimensions in patients with disabling tinnitus and craniomandibular disorders. *British Journal of Audiology* **1991**, *25*, 15–24. <https://doi.org/10.3109/03005369109077860>. 833
53. Baloi, J. Statistical Analysis and Clustering of Tinnitus related Data with respect to the perceived Symptoms. Master's thesis, Ulm University, 2020. 834
54. Ogbuabor, G.; F. N, U. Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value. *International Journal of Computer Science and Information Technology* **2018**, *10*, 27–37. <https://doi.org/10.5121/ijcsit.2018.10203>. 835
55. Shahapure, K.R.; Nicholas, C. Cluster Quality Analysis Using Silhouette Score. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747–748. <https://doi.org/10.1109/DSAA49011.2020.00096>. 836
56. Vinh, N.X.; Epps, J.; Bailey, J. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance **2010**. p. 18. 837
57. Rosenberg, A.; Hirschberg, J. V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure **2007**. p. 11. 838
58. Sonmez, G.; Basekim, C.C.; Ozturk, E.; Gungor, A.; Kizilkaya, E. Imaging of pulsatile tinnitus: a review of 74 patients. *Clinical imaging* **2007**, *31*, 102–108. 839

59. Azevedo, D.; Rodrigues, A.M.; Canhão, H.; Carvalho, A.M.; Souto, A. Zgli: A Pipeline for Clustering by Compression with Application to Patient Stratification in Spondyloarthritis. *Sensors* **2023**, *23*, 1219. 872
60. Asnis, G.M.; Majeed, K.; Henderson, M.A.; Sylvester, C.; Thomas, M.; La Garza, R.D. An Examination of the Relationship Between Insomnia and Tinnitus: A Review and Recommendations. *Clinical Medicine Insights: Psychiatry* **2018**, *9*, 1179557318781078. 873  
Publisher: SAGE Publications Ltd STM, <https://doi.org/10.1177/1179557318781078>. 874
61. Schlee, W.; Pryss, R.C.; Probst, T.; Schobel, J.; Bachmeier, A.; Reichert, M.; Langguth, B. Measuring the Moment-to-Moment Variability of Tinnitus: The TrackYourTinnitus Smart Phone App. *Frontiers in Aging Neuroscience* **2016**, *8*. 875
62. Guillard, R.; Decobecq, F.; Fraysse, M.J.; Favre, A.; Congedo, M.; Loche, V.; Boyer, M.; Londero, A. Traduction française validée du questionnaire standardisé d'anamnèse d'acouphène ESIT-SQ. *Annales françaises d'Oto-rhino-laryngologie et de Pathologie Cervico-faciale* **2022**. Publisher: Elsevier, United States. 876
63. Bair, E. Semi-supervised clustering methods. *WIREs Computational Statistics* **2013**, *5*, 349–361. \_eprint: <https://onlinelibrary.wiley.com/doi/10.1002/wics.1270>. 877
64. Zhang, D.; Zhou, Z.H.; Chen, S. Semi-Supervised Dimensionality Reduction. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM)*; Proceedings, Society for Industrial and Applied Mathematics, 2007; pp. 629–634. <https://doi.org/10.1137/1.9781611972771.73>. 878
65. Xing, E.; Jordan, M.; Russell, S.J.; Ng, A. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems* **2002**, *15*. 879
66. Weinberger, K.Q.; Saul, L.K. Distance metric learning for large margin nearest neighbor classification. *Journal of machine learning research* **2009**, *10*. 880
67. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* **2017**, *2*, 205. <https://doi.org/10.21105/joss.00205>. 881
68. Varol, E.; Sotiras, A.; Davatzikos, C. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *NeuroImage* **2017**, *145*, 346–364. <https://doi.org/10.1016/j.neuroimage.2016.02.041>. 882
69. Chand, G.B.; Dwyer, D.B.; Erus, G.; Sotiras, A.; Varol, E.; Srinivasan, D.; Doshi, J.; Pomponio, R.; Pigoni, A.; Dazzan, P.; et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain* **2020**, *143*, 1027–1038. <https://doi.org/10.1093/brain/awaa025>. 883
70. Chew, Q.H.; Prakash, K.N.B.; Koh, L.Y.; Chilla, G.; Yeow, L.Y.; Sim, K. Neuroanatomical subtypes of schizophrenia and relationship with illness duration and deficit status. *Schizophrenia Research* **2022**, *248*, 107–113. <https://doi.org/10.1016/j.schres.2022.08.004>. 884