Diagnostic accuracy of audio-based seizure detection in patients with severe epilepsy and an intellectual disability

Citation for published version (APA):
Arends, J. B., van Dorp, J., van Hoek, D., Kramer, N., van Mierlo, P., van der Vorst, D., & Tan, F. I. Y. (2016). Diagnostic accuracy of audio-based seizure detection in patients with severe epilepsy and an intellectual disability. *Epilepsy and Behavior*, *62*, 180-185. https://doi.org/10.1016/j.yebeh.2016.06.008

DOI:
10.1016/j.yebeh.2016.06.008

Document status and date:
Published: 01/09/2016

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Epilepsy & Behavior 62 (2016) 180–185

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Diagnostic accuracy of audio-based seizure detection in patients with severe epilepsy and an intellectual disability

Johan B. Arends [a,b,*], Jasper van Dorp [c], Dennis van Hoek [a], Niels Kramer [d], Petra van Mierlo [a], Derek van der Vorst [c], Francis I.Y. Tan [a]

[a] Kempenhaeghe/Heeze, The Netherlands
[b] Technological University Eindhoven, Eindhoven, The Netherlands
[c] Sound Intelligence BV/Amersfoort (part of CLB), The Netherlands
[d] CLB BV/Raamsdonksveer, The Netherlands

ARTICLE INFO

Article history:
Received 20 March 2016
Revised 12 June 2016
Accepted 13 June 2016
Available online 1 August 2016

Keywords:
Epilepsy
Seizure detection
Audio monitoring
Diagnostic study

ABSTRACT

We evaluated the performance of audio-based detection of major seizures (tonic–clonic and long generalized tonic) in adult patients with intellectual disability living in an institute for residential care.

Methods: First, we checked in a random sample (n = 17, 102 major seizures) how many patients have recognizable sounds during these seizures. In the second part of this trial, we followed 10 patients (who had major seizures with recognizable sounds) during four weeks with an acoustic monitoring system developed by CLB ('CLB-monitor') and video camera. In week 1, we adapted the sound detection threshold until, per night, a maximum of 20 sounds was found. During weeks 2–4, we selected the epilepsy-related sounds and performed independent video verification and labeling ('snoring', 'laryngeal contraction') of the seizures. The video images were also fully screened for false negatives. In the third part, algorithms in the CLB-monitor detected one specific sound (sleep-related snoring) to illustrate the value of automatic sound recognition.

Results: Part 1: recognizable sounds (louder than whispering) occurred in 23 (51%) of the 45 major seizures, 20 seizures (45%) were below this threshold, and 2 (4%) were without any sound. Part 2: in the follow-up group (n = 10, 112 major seizures; mean: 11.2, range: 1–30), we found a mean sensitivity of 0.81 (range: 0.33–1.00) and a mean positive predictive value of 0.40 (range: 0.06–1.00). All false positive alarms (mean value: 1.29 per night) were due to minor seizures. We missed 4 seizures (3%) because of lack of sound and 10 (9%) because of sounds below the system threshold. Part 3: the machine-learning algorithms in the CLB-monitor resulted in an overall accuracy for 'snoring' of 98.3%.

Conclusions: Audio detection of major seizures is possible in half of the patients. Lower sound detection thresholds may increase the proportion of suitable candidates. Human selection of seizure-related sounds has a high sensitivity and moderate positive predictive value because of minor seizures which do not need intervention. Algorithms in the CLB-monitor detect seizure-related sounds and may be used alone or in multimodal systems.

1. Introduction

Nocturnal seizures often go unnoticed and are associated with SUDEP [1]. For detection of these seizures, heart rate [2–4] and movement [5] are the physical signs most often used. Audio detection has become popular in many fields of health care, because it only uses the traditional acoustic monitoring systems for night-care and is a non-intrusive method. Until now, audio-based detection of epileptic seizures has been disappointing because of the plethora of noise that is received during the night in many hospitals. Audio-based seizure detection, however, remains attractive, because some of the patients have specific seizure-related sounds, which easily can be identified by the human ear if heard in situations where only a few sounds are passed though the widely used audio-based surveillance systems.

In a previous study [6] in our institute, automatic detection of a number of specific sounds (by matching their frequency spectrum) resulted in high performances. However, we still do not know what the audio detection will miss or detect falsely, because of lack of sounds, minor epileptic seizures such as myoclonic or short tonic seizures (which do not need intervention), or nonepileptic events.

Therefore, we studied the usefulness of audio-based nocturnal seizure detection in patients with severe epilepsy in a residential setting with video as the gold standard. All of the patients were adults with an intellectual disability and had been previously studied by EEG/video.

* Corresponding author at: Kempenhaeghe/Technological University, Eindhoven, Sterkselseweg 65, 5591 VE, Heeze, The Netherlands. Tel.: +31 40 227 92 15.
E-mail address: ArendsJ@kempenhaeghe.nl (J.B. Arends).

http://dx.doi.org/10.1016/j.yebeh.2016.06.008

The STARD criteria [7] and ILAE classification of 2010 [8] were used. To assess the representativeness of our study population, we present an earlier trial in which we assessed the prevalence of sounds in a broader population of adults having an intellectual disability. Finally, to check the potential value of completely automatic sound recognition, we analyzed the performance of automatic analysis of 'snoring' sounds.

2. Methods

The research received prior approval by the institutional review body of Kempenhaeghe, and informed consent was obtained from each patient's representative. In Kempenhaeghe, we run a continuous program for seizure detection investigating 40–50 patients each year. The 17 initial subjects in Section 2.1 entered our program in 2008, while the 10 subjects investigated in Section 2.2 were consecutively selected during 2014.

We followed a stepped approach in our in-hospital population of patients with intellectual disability and severe epilepsy.

2.1. Representativeness check

To assess the representativeness of our study population, we determined the proportion of patients that produced sounds during their major seizures: generalized tonic–clonic (>30 s) or generalized tonic seizures of long duration (>30 s). In 17 patients, 102 major seizures occurred during 4 weeks. Forty-five seizures were classified as major motor seizures (tonic–clonic or long generalized tonic seizures). All other (nonmajor) seizures (50) were labeled as minor seizures, and 7 could not be classified (not included in the analysis). The perceived loudness of the seizure sounds was subjectively judged by a panel (PvM and JA) on a relative scale of loudness (0–100% in steps of 10%). We used the following loudness reference sounds: whispering (20–30% of the scale), talking (40–60%), and screaming (70–90%). The panel concentrated during the trial on the sounds and was not looking at the video, but they were not completely blinded; in a later stage, the video was scored by the same panel.

2.2. Human sound recognition and analysis

From the population of 284 patients with intellectual disability and severe epilepsy, we randomly selected 10 patients (12–65 years old) who were known to make audible sounds during their seizures and suffer at least one major seizure a week. The patients were diagnosed during our clinical seizure detection program where seizures were detected during a clinical study of 1 week (2 days EEG/video followed by 5 days video and multimodal non-EEG sensors for accelerometry and heart rate which were not used in this study). During the trial period, we used the threshold-based CLB-monitor to collect sounds with simultaneous continuously recorded nocturnal video monitoring during 4 weeks. In the first week, we collected noise fragments with a sound pressure level above a predefined threshold. This threshold was set manually per patient, varying from a level corresponding to whispering, up to a level corresponding to shouting. The integration times used for each patient varied in the range 0.0 to 2.5 s. Furthermore, we identified sounds that were specifically related to the videotaped epileptic seizures (for example, due to laryngeal spasm, a myoclonic hiccup, coughing). During the first week, the audio threshold was adapted until, per night, a maximum of 20 sounds were above the threshold. Generally, between 2 and 20 sounds were detected by the system, of which 0–2 could be linked to epileptic seizures (0–10% of unselected sounds). During the following 3 weeks, all sounds were collected and classified as belonging to seizures or not. Afterwards, the relation of the presumed seizure-related sounds to the real seizures was independently verified (by video), the seizures were classified, and the nature of the sounds labeled. To avoid missed seizures, all video recordings were screened for each night (at 16× normal speed). When doubtful episodes were suspected at high speed, we went back to a normal speed for a period of 5 min around the event. Suspected or possible seizures were classified by a panel (one epileptologist and at least two nurses specializing in epileptology). The sensitivity and positive predictive value of the seizure-related sounds for the detection of major seizures were determined. Furthermore, the number of false sound alarms per night was assessed.

2.3. Automatic sound selection

An automated sound event detection system by Sound Intelligence was tested on the collected audio data as well. The system is based on machine learning, making it necessary to have sufficient amounts of data for a particular sound category in order to be able to train and test the system. Depending on the sound class, a specific combination of decision tree algorithms and/or neural network algorithms is chosen to achieve optimal results. In the data collected in this trial, not enough epilepsy-related sounds were collected, making it impossible to train the system on these categories. However, sufficient data were available for 'snoring' (including sleep-related snoring), which was the most prominent sound in three patients (see Results, Table 2). For these patients, a snoring detector would be relevant in detecting seizures.

The machine learning algorithm used consisted of a neural network and was trained and tested only for snoring, as a proof of concept for detecting other types of epilepsy-related sounds in the future, once sufficient amounts of epilepsy-related sounds have been collected. To train and test the SI-monitor, the available audio data were annotated manually and split randomly (not by patient) in training and validation sets using a 70/30 ratio. This resulted in a training set consisting of 3760 events (of which, 936 were annotated as 'snoring') and a validation set consisting of 1608 events (of which, 338 were annotated as 'snoring'). The system was trained to classify snoring and classify all other sounds just as 'other'. After training, the system was validated using the validation set.

Because this was a diagnostic study, only descriptive statistics are presented.

3. Results

3.1. Representativeness check

Results of the perceived loudness of the seizure sounds are depicted in Fig. 1.

At least one sound event was found in 60 of the 95 seizures (63%), in 43 of the 45 major seizures (96%), and in 17 of the 50 other minor seizures (34%). Recognizable sounds with a perceived loudness above the level of whispering occurred in 23 (51%) of the major and 6 (12%) of the minor seizures.

In other words, 96% of major seizures were accompanied by sound, of which about half had a sound perception level (SPL) above the detection threshold (whispering). Only 34% of the minor motor seizures were accompanied by sound, of which only 12% had an SPL above the detection level.

The types of the sounds are depicted in Fig. 2.

From this figure, one can see that screaming and bed sounds are predominantly related to major seizures.

3.2. Human sound recognition and analysis in our selected patient population

The mean age of the patients was 34 years (range: 18–42 years); 6 patients were female, and 4 were male. All patients were known to suffer from symptomatic generalized or multifocal epilepsies and had a moderate-to-severe intellectual disability. In Table 1, the results of the manual audio analysis are shown. From this table, it is obvious that manual selection of seizure-related sounds is a sensitive procedure. False alarms are related to less severe, minor seizures that do not require an intervention. In Table 2, the most frequent types of seizure-related sounds

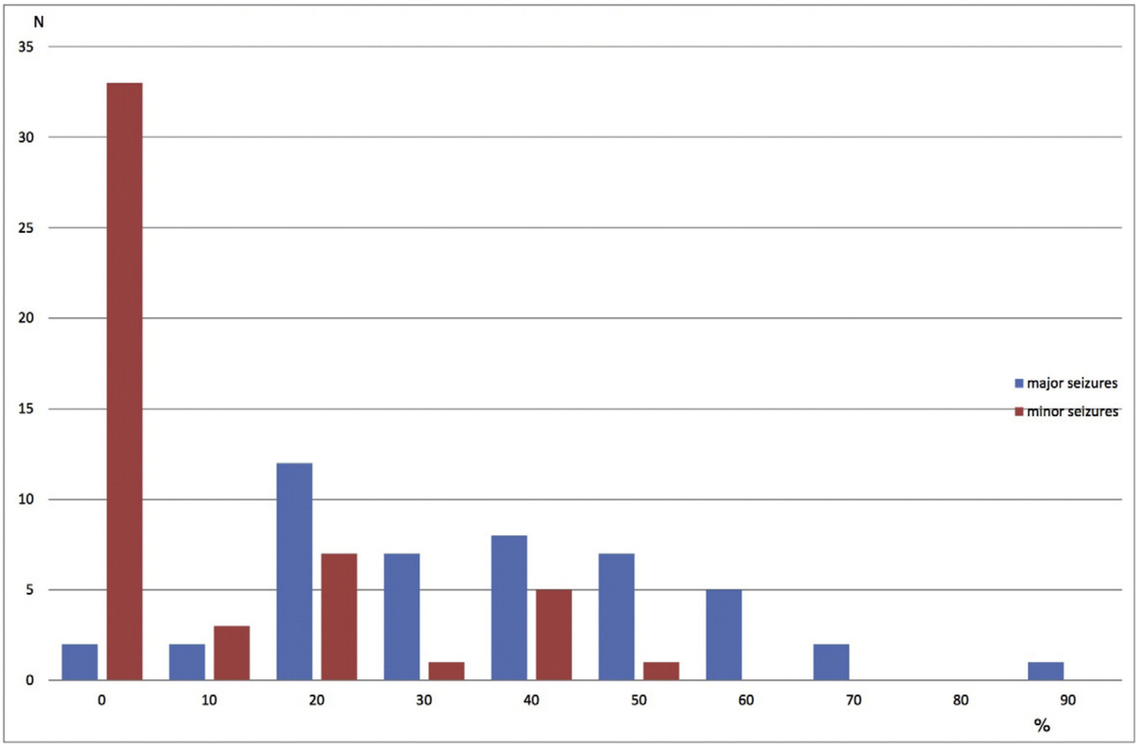

Fig. 1. Number of patients with audible sounds. x-axis: relative loudness.

are depicted per patient. Some patients (3 of 10) had more than one sound type. From Table 2, it can be concluded that almost exclusively laryngeal contractions or respiratory sounds are detected by the human observer. The laryngeal contractions were due to (myo)clonic or tonic mechanisms. Bed-related sounds were not found as the most dominant one.

3.3. Results after testing with a sound event detection system

The results of the automatic snoring detection using machine learning are shown in Table 3. The overall validation accuracy of the system was 98.3%; all true detections (= 1580) divided by all detections (= 1608). For snoring, the sensitivity was 0.95 (BI: 0.92–0.97); true

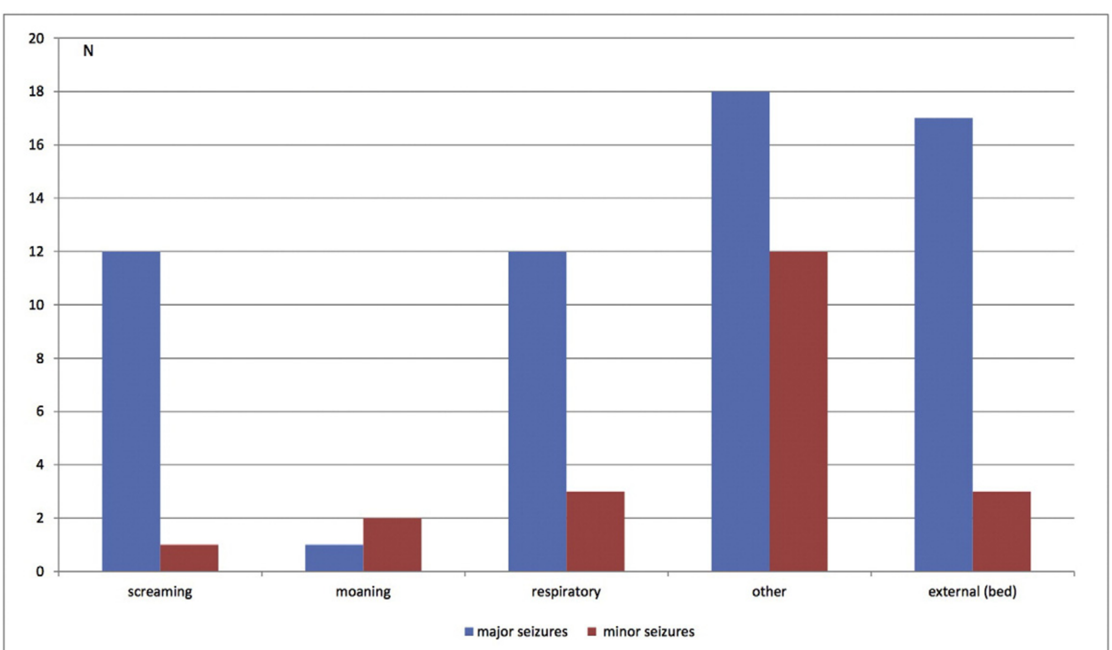

Fig. 2. The types of sounds found in the population. On the vertical axis, the number of seizures is shown. Respiratory-related sounds include the typical laryngeal tonic of myoclonic sounds. The external bed sounds have a rhythmical character corresponding to the frequency of the (myo)clonic jerks.

Table 1
Results of the manual audio analysis.

Patient number	Number of seizures				Sensitivity	Positive predictive value	Number of false alarms\night[b]
	Major[a]	Major without sounds	Major with subthreshold sounds	Minor			
1	3	0	0	45	1.00	0.06	2.5
2	18	1	4	7	0.44	0.72	0.5
3	3	0	2	2	0.33	0.60	0.1
4	30	0	0	67	1.00	0.31	3.5
5	10	0	0	33	0.90	0.23	1.38
6	9	0	4	40	0.56	0.18	1.48
7	10	0	0	9	1.00	0.53	0.33
8	25	3	0	62	0.88	0.29	2.48
9	3	0	0	0	1.00	1.00	0
10	1	0	0	16	1.00	0.06	0.59
Total	112	4	10	281			
Mean	11.20	0.40	1.00	28.10	0.81	0.40	1.29
Maximum	30	3	4	67	1.00	1.00	3.5
Minimum	1	0	0	0	0.33	0.06	0

Sensitivity and positive predictive value of manual audio classification by listening to detected sounds.
[a] Major = tonic–clonic seizures, clonic seizures, or tonic generalized seizures >30 s.
[b] False alarms = number of minor seizures\night.

positive detections (=322) divided by all detections of snoring (=338). The positive predictive value was equal to 0.96 (BI: 0.93–0.98); true positive detections (=322) divided by all snoring sounds (=334).

4. Discussion

Audio detection of seizures is feasible if patients initially are selected on the basis of audible seizure-related sounds. In our population of intellectually disabled persons, this is feasible for about half of the patients with major seizures. A detection procedure by a human observer results in a high sensitivity of 0.81 and a moderate positive predictive value of 0.40. False positives are infrequent (1.9 false alarms per night on average) and due to minor seizures, which emphasizes the strong seizure-related character of seizure-specific sounds. The majority of false negatives are caused by major seizures without (3%) sounds or with sounds below the set detection threshold (9%). One should bear in mind that the performance figures in this study are obtained by using video alone as reference. Furthermore, the panel was not completely blind: they assessed the sounds and later classified the seizures, which may have caused observation bias. Although clonic and hypermotor phases are easily recognized by video, tonic elements may be missed or misclassified.

Automatic detection of one particular sound event class (sleep-related, not specific postictal ‘snoring’) using a machine learning-based system yielded a high sensitivity of 0.95 and positive predictive value of 0.96. This indicates that seizure-related sounds carry a high potential for automatic detection, if these sounds are identified by an experienced observer.

Table 2
Most frequent seizure-related sounds.

Patient	Number of sounds	Screaming	Snoring	Laryngeal contraction	‘Forced’ inspiration
1	1			1	
2	1		1		
3	1				1
4	2		1	1	
5	2		1	1	
6	1			1	
7	2	1		1	
8	1			1	
9	1			1	
10	1				1
Total	13	1	3	7	2

Type of sounds detected. The predominant types of sounds are shown. For each patient, a maximum of two types was chosen.

The automatic analysis of the snoring events in this study and other specific sounds [6] shows that automatic sound detection in patients with tonic–clonic seizures may become a useful tool in the surveillance of patients with epilepsy with high SUDEP risk. To succeed, we will need to develop a prior sound identification phase for each patient.

4.1. Comparison with other detection modalities

Compared with the detection of major seizures with other non-EEG modalities, such as movement, heart rate, galvanic skin resistance, or electromyography [9,10], our method of audio detection shows a good sensitivity, a moderate positive predictive value with an almost perfect association with seizures in general, and a high potential for automatic detection. The findings in this study confirm that audio detection is a highly personalized form of seizure detection, contrary to most other detection modalities that use the ‘generic’ properties of the seizures such as the motor phenomena of classical tonic (–clonic) signs. Dynamic algorithms that adapt to the personal and local situation therefore are essential in audio detection, while in other modalities, rather ‘fixed’ algorithms may suffice.

4.2. How many patients are suitable for audio analysis

The fact that audio analysis is suitable for only half of the patients is a drawback. This may be partly overcome by using low detection thresholds in the machine learning algorithms and improved sound detection by more sophisticated array microphones and improving the acoustics and/or lowering noise levels in the bedroom. Improved sound detection will also diminish the current false negative detection rate (9% of seizures were not detected because of subthreshold sounds).

4.3. Other audio detection methods

Compared with those of other published audio detection systems [11] (sensitivity: 0.33 and positive predictive value: 0.33) [12], our results seem to be more promising. This is probably due to the fact that

Table 3
The results of the automatic detection of snoring sounds.

		Reported sound class	
		Other	Snoring
Actual sound class	Other	1258	12
	Snoring	16	322

two former systems only respond to the frequency of the sound caused by the clonic phase of the seizures (between 3 and 12 Hz), while the CLB system uses specific sounds related to the seizure semiology itself. However, of course, we cannot directly compare our study with the other ones [11,12], because the number of sounds in our population was not large enough for a complete automatic evaluation.

4.4. Types of sounds detected

Although the audio signals from seizures may come from different internal and external sources, in practice, respiratory and laryngeal sounds were most often selected by the sound filter and our manual assessment. Apart from snoring and heavy breathing, laryngeal myoclonic or tonic contractions are important phenomena in this selected group of patients. Probably, these sounds are most recognizable for the human observer. Sounds labeled as 'screaming' in the first population may have been assigned to 'laryngeal' sounds in the second one. We have no explanation for the differences in proportion of bed sounds between both populations other than a coincidental finding.

4.5. False alarms

We considered alarms due to minor seizures as false positives. We did not see them as true positives, because we only wanted to detect acute 'need for care'. In other situations such as diagnostic or therapeutic evaluations, detection of minor seizures is desirable resulting in 'true' positives.

Because the false alarms were caused by minor seizures, it might be useful to verify them in clinical practice by a video camera with inbuilt memory to allow for replay of prior events. This is especially useful when many false alarms occur (in this study, 3 of 10 patients showed more than 2 false alarms per night). Another method of lowering the number of false alarms is to integrate audio detection in multimodal detection systems that are currently in development.

4.6. Population

The population in this study consisted of adults with intellectual disability. This population was initially chosen because of the high frequency of unattended tonic–clonic seizures associated with a high risk of SUDEP. In children or other adults (often with fewer seizures), the results of audio detection may differ from our study. Because the mechanism and semiology of tonic and tonic–clonic seizures is similar in all groups, there will be no major differences; the sounds produced will be altered largely by age (children) and sometimes by abnormal neurological development associated with intellectual disability.

4.7. Automatic analysis

Even though an insufficient number of epilepsy-related sound events was available as part of this study to train a machine learning-based automatic detection system, it was still possible to train and test the Sound Intelligence system using sleep-related 'snoring', which was the most prominent epilepsy-related sound in 30% of our patients. The resulting detection accuracy (98.3%) is promising and shows that, when sufficient data are available, a machine learning-based algorithm could be used to detect at least some forms of epileptic seizures as well. The resulting sensitivity (0.95) of the system is comparable with the ones obtained in an earlier study [6], which used a Bayesian classifier to detect screams (sensitivity = 0.98), smacking of the lips (sensitivity 0.98), noises due to bronchial secretion (sensitivity = 0.95), and movements of the bed (sensitivity = 0.97). The positive predictive value (PPV) for sleep-related snoring (0.96) is higher than the ones found previously [6]: PPV = 0.30 for screaming, 0.02 for smacking of the lips, 0.02 for noises due to bronchial secretion, and 0.40 for movements of the bed. This means that the Bayesian system used by [6] will result in much more false alarms compared with the CLB-monitor. It has to be noted, though, that their algorithm was able to recognize four different sounds instead of one. Furthermore, the types of sound events were different and not the same as the predominant sounds in this study. The CLB-monitor could be trained for other types of epilepsy-related sound events once sufficient data of these events are available and compared more thoroughly with the Bayesian system [6].

The sleep-related snoring detection results of the proposed machine learning approach are comparable with those found in an earlier study by Dafna et al. [13], which proposed an AdaBoost-based classifier to detect snoring with a resulting overall accuracy of 0.98 and a sensitivity of 0.98. However, a major difference between that and the present study is that the AdaBoost-based classifier is feature-based, whereas the neural network-based classifier proposed in this study 'finds' its own features. So when expanding the classifier with other seizure-related sounds, it may be necessary for the AdaBoost-based classifier to be expanded manually with new features as well. Also, in [13], more audio recordings were available (>76,600 recordings versus 5368 in this study); usually with machine learning, an even better detection accuracy/sensitivity will be reached when more training data are used. Another difference between the present study and the AdaBoost-based classifier is that the latter included an adaptive noise reduction algorithm, which was not used in the neural network-based classifier.

4.8. Integration

We foresee a future in which these dynamic audio detection algorithms will coexist within the more general multimodal systems. For each client, the presence and type of particular sounds can be identified and the optimal detection method chosen. Audio detection alone may be useful for some patients, especially in hospitals or institutions where no one is woken up unnecessarily. This article shows that dedicated audio analysis may be an important part of future seizure detection systems.

Acknowledgment

We acknowledge the critical review of the manuscript by H. van Dijk.

Conflicts of interest

Author JB Arends has served as a paid consultant for UCB (development of an educational program) and Eisai (lectures). Authors J van Dorp and D van der Vorst are employed by CLB and Sound Intelligence (who own the CLB detection system). The remaining authors have no conflicts of interest.

References

[1] Monté C, Arends J, Tan I, Limburg M, De Krom M. Sudden unexpected death in epilepsy patients: risk factors, a systematic review. Seizure 2007;16:1–7.
[2] Zijlmans M, Flanagan D, Gotman J. Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign. Epilepsia 2002;43:847–54.
[3] Leutmezer F, Schernthaner C, Lurger S, Pötzelberger K, Baumgartner C. Electrographic changes at the onset of epileptic seizures. Epilepsia 2003;44:348–54.
[4] Van Elmpt W, Nijsen T, Griep P, Arends J. A model of heart rate changes to detect seizures in severe epilepsy. Seizure 2006;15:366–75.
[5] Nijsen T, Arends J, Griep P. The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy. Epilepsy Behav 2005;7:74–84.
[6] De Bruijne G, Sommen P, Aarts R. Detection of epileptic seizures through audio classification. 4th European conference of the International Federation for Medical and Biological Engineering. Berlin Heidelberg: Springer; 2009. p. 1450–4.
[7] Bossuyt PM, Reitsma JB, Bruns DE, et al, For the STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015; 351:h5527.
[8] Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 2010;51:676–85.
[9] Van de Vel A, Cuppens K, Bonroy B, et al. Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art. Seizure 2013;22:345–55.

[10] Van Andel J, Thijs R, De Weerd A, Arends J, Leijten F. Ambulatory seizure detection: what is available and how will it influence epilepsy care? Epilepsy Behav 2016;57:142–5.
[11] Carlson C, Arnedo V, Cahill M, Devinsky O. Detecting nocturnal convulsions: efficacy of the MP5 monitor. Seizure 2009;18:225–7.
[12] Fulton S, Van Poppel K, McGregor A, Ellis M, Patters A, Wheless J. Prospective study of 2 bed alarms for detection of nocturnal seizures. J Child Neurol 2012;28:1430–3.
[13] Dafna E, Tarasiuk A, Zigel Y. Automatic detection of whole night snoring events using non-contact microphone. PLoS One 2013;8(12):1–14.