Energy Systems
https://doi.org/10.1007/s12667-022-00529-0

ORIGINAL PAPER

# Structured vocabulary specific to power operation control centers

**Adrisson C. Floriano**[1] · **Sérgio L. Avila**[2] · **Rubipiara C. Fernandes**[2]

Received: 2 December 2021 / Accepted: 30 July 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

The power operation sector is facing an evolution regarding the adoption of new technologies that, although necessary and unavoidable, add operational complexity to the system. Due to the essence of their activities and the systemic orientation required for their execution, the operation centers are natural information hubs. As far as audio calls are concerned, much of the communication between companies' real-time teams is done over the voice channel. The audios are evidence for confirmation of events, procedures, changes in operating states, or any interventions in the power system. Due to the evolution of techniques for processing large amounts of data, it has become possible to extract structured information from audio files. The main contribution of this work is a methodology to obtain a structured vocabulary for power operation centers, according to the governmental requirements of the National Electric System Operator, and that takes both the technical vocabulary, the plurality of accents, and Brazilian–Portuguese language regionalisms. This structured vocabulary has four entities: a general one dedicated to the power operation center and three specifics to hydro, thermoelectric, and photovoltaic plants. Metrics to assess the quality of the vocabulary obtained are also discussed. The definition of a metric adapted to the context is necessary to compare solutions and measure the evolution. The methodology developed seeks evidence of a more dynamic and efficient power operation and can be adapted to any language.

**Keyword** Structured vocabulary · Power operation · Control centers · Speech-to-text · Audio

✉ Sérgio L. Avila
Sergio.Avila@ifsc.edu.br

[1] ENGIE Brasil Energia S.A., Florianópolis 88025-255, Brazil

[2] Federal Institute of Santa Catarina, Florianópolis 88020-300, Brazil

Published online: 16 August 2022

## 1 Introduction

The power operation sector has been restructured to meet the needs of a society in constant development. Power operation companies seek modernization on several fronts, such as:

1. From the operation assets' perspective, teleassistance from power plants and substations has become a recurrent solution among the sector agents. They must regard the grid procedures for this new operation mode [1].
2. Regarding energy planning, intermittent sources are increasing their relevance. Despite the economic and environmental advantages, their non-dispatchable characteristics demand a more efficient operation [2, 3].
3. From an economic point of view, in search of efficiency and competitiveness, the weekly price of energy and dispatch schedule are migrating to half-hour intervals in Brazil. Since 2020, the governmental National Electric System Operator (ONS) has been using a daily generation schedule, with 30-min-windows [4]. Despite benefits such as transparency and reproducibility of processes and a reduction in the occurrence of out-of-merit generation, these changes affect real-time operation teams, who need to adjust power generation more frequently; and the pre-operation teams, who need to agree with the generation program, and with the ONS for the following day.
4. Distributed generation is increasing, and consumers-producers are more numerous every day [5]. Advantages like power generation close to the load, reduction of technical losses, and service to isolated loads demand new solutions for controlling voltage levels, protection of distribution networks, and load forecasting [6, 7].

Desirable and inevitable, these initiatives add operational complexity to the system. In addition to the procedures and tools already consolidated, the challenge of the operating power systems activity is not to become a limiting factor for these evolutions.

Thanks to the systemic orientation necessary for daily tasks execution, the operation centers are natural information hubs. Concerning audio calls, a large part of the communication between the real-time operation companies' teams with field technicians or ONS operators is the voice channel. These recordings are stored [1], creating a large data volume, and selected on-demand. They are evidence for the operation analysis reports [8]. These audio files contain other information relevant to the system's operation and companies' billing. However, they are few exploited due to the difficulty of extracting, tracking, and associating these audios with data from other operating supervisory systems.

In recent years, the evolution and popularization of numerical techniques capable of processing large amounts of data have been applied to the management of equipment and functions of systems, including in the power system control center. In a general way, it is possible evaluating the behavior of a control center to manage them promptly using SCADA systems [9, 10]. Other works have made time-series

forecasts in a dynamic decision support system for different information, such as infrastructure prognosis [11, 12] or electricity demand and prices [13, 14].

It is also possible to extract structured information from audio files of call centers. By the way, the main element that influences the possibility of automation of contact center processes probably is the call transcription methods implemented by automatic speech recognition (ASR) systems, also called speech-to-text (StT).

Overview studies present recent progress in models for one spoken language processing [15, 16]. A multilingual modeling overview of ASR is presented by [17]. A live streaming speech recognition is proposed by [18]. Multilingual and live streaming make the problem more realistic. These works indicate that convolutional neural networks (CNN) have excelled over other computational intelligence techniques, like the end-to-end speech recognition with deep multipath CNN [19].

The current ASR systems do not guarantee high transcription quality for call centers [15–17]. It can be due to the sound signal generated there, whose quality parameters differ from studio recordings. Call centers have poor quality recordings, like our power operation control centers. A lot of people talking together with different vocabulary is the regular operation. A recent work presents a call transcription methodology for a generic call center [20]. They propose a system with a preprocessing of sound recordings and postprocessing of produced transcriptions. As for preprocessing, they separate the client and agent channels into two independent signals, training ASR systems, and doing an audio signal correction. The postprocessing actions are to eliminate the inarticulate sounds, the standard phrases (numbers, dates, times, etc.), and the close sound phrases.

According to our research, most studies on StT are about natural generic language [19–21]. Some works are dedicated to specific languages, such as Mandarin [22], Serbian [23], Colombian [24], and Brazilian Portuguese languages [25]. A few works are about dedicated to engineering applications. Two examples are the audiovisual speech recognition in-vehicle cabin [26] and vocabulary learning in games [27]. Our study gains even more originality and relevance because there is nothing published for the energy systems yet.

In addition, the ONS regulation on technical verbal communication [28], published in 2020, was considered here. It establishes the rules, procedures, and standard phraseology, of mandatory compliance, that apply to the real-time teams of the ONS Operations Centers and Agent Operations Centers, in any communication voice operation, within the scope of the National Interconnected System (SIN). Among many determinations, the ONS [28] requires that all voice communication must always have the same structure: 1st. Identification of interlocutors; 2nd. Transmission of the message by the sender; 3rd. Repetition of the message received by the receiver; 4th. Confirmation by the sender. If the receiver repeats the message incorrectly, the sender will transmit a denial term such as, for example, "negative", "that's not it", followed by the repetition of the initial message, until there is confirmation, at the sender, of the correct understanding of the message by the receiver; 5th. Closing. The ONS [28] objectives are to ensure the standardization and uniformity of operative communications; reduce the transmission time of messages, without compromising the content and awareness; and provide clear and concise operating orders and information. Although the technicians have adequate training, the dynamics and

intimacy throughout the operations end up making the procedure more user-friendly and, therefore, less rigorous. For all these reasons, the requirement to record, transcribe and identify the occurrences is explicit.

In this context, we propose a structured vocabulary to power operation control centers. The methodology must have natural language processing, which takes care of both the technical vocabulary of the power sector and the plurality of accents and regionalisms in the Brazilian territory. We use audio samples and documentation from ENGIE's Systems Operation Center (COS) and ONS requirements [1, 28]. We are not yet focusing on the development of an automatic speech recognition tool as an end-to-end strategy right now. Other objectives are:

1. To transform a regulatory requirement for the storage of communication audio files, reusing them as a data source for the StT application, applying machine learning techniques to convert raw files into an operational skill, through audio transcription and information structuring.
2. To define quantitative metrics for the quality of the transcripts, allowing the comparison between different vocabularies, techniques, and tools; and
3. To obtain relevant information using audio files in power control centers from a regulatory and technical perspective.

This work is organized as follows. Section 2 presents the best practices for audio transcription, as well as audio file processing, ground truth (GT) definition, and comparisons among StT market solutions. Hereafter, the personalized vocabularies can be conceived to improve the transcription quality. Section 3 shows the proposed methodology and possible applications. The results (transcriptions) are stored in text format and indexed with the original audios and can serve as input for several other applications, such as call labeling, operational indicators, identification of changes in generators operation states, among other possibilities. Section 4 is the conclusion.

## 2 Methodology development

The ONS regulates the operations of power control centers, being responsible for the power operation planning and programming, as well as the generation of centralized dispatch control, according to the Brazilian Law 10,848/2004. The ONS communicates to the companies' operation center by voice call [29]. The ONS and companies must be keeping these voice records for two years minimum [1]. ENGIE has had an audio database since 2017, currently with ~ 90 GB distributed in more than 550,000 MP3 files. These numbers are incremented at an average rate of 700 files, increasing 9 h of recordings and 100 MB of storage per day. This data mass is stored and consulted on demand for the operation analysis reports preparation [8]. Here, we chose samples in WAV format with a bit rate of 256 kbps and MP3 with a bit rate of 64 kbps. Bit rate means bit transfer stream and is associated with the quality of the audio files [29]. In addition to the audios themselves, the recorder used by ENGIE provides some metadata such as calls' date and time, duration in seconds, origin,

and destination. A set of 200 audios and their respective manual transcriptions are used as ground truth (GT) for exploratory activities. GT refers to information that is known to be real or true. The next section presents the quality metrics and techniques used in the audios and their transcriptions.

### 2.1 Transcript's quality measuring

There are several criteria to quantify the quality of the transcripts [31]. Defining a context-adapted metric is required to compare solutions and assess model evolutions.

*Word error rate (WER)* To quantify the difference between a hypothesis text (automated transcriptions) and the reference text (GT), the number of substitutions, deletions, and insertions necessary to make the hypothesis text equal to the GT is considered, like that:

$$WER = \frac{(R + E + I)}{N}, \tag{1}$$

where $R$ is the number of words replaced in the hypothesis text, $E$ is the number of words excluded in the hypothesis text; $I$ is the number of words inserted in the hypothesis text; $N$ is the number of words in GT [32].

The GT has the lowest possible WER, zero. The higher this number, the greater the distance between the hypothesis text and the reference. Acceptable values for WER depend on the context of each audio [33]. WER shows the need for pre-processing actions that help the transcriptions manipulations. These are converting all content to lowercase or uppercase, removing multiple spaces, converting phrases to word lists, removing specific words, removing special characters like tabs and line breaks, removing punctuation, and replacing words with synonyms [34].

*Keyword error rate (KER)* Although the WER is a widely used indicator for the validation of the StT model, it uniformly considers each transcribed word, regardless of its relevance in the transmitted information. Thus, for WER calculation purposes, generic words such as definite or indefinite articles, pronouns, and connectives are as significant as specific technical terms, although this second category adds more value to the information. Thus, there may be situations where a high WER value does not match the transcription system's ability to extract relevant information. Considering this as a drawback, there is the KER metric, which computes the errors and missing keywords in the hypothesis text, compared to the GT:

$$KER = \frac{(A + M)}{N}, \tag{2}$$

where $A$ is the number of added keywords in the hypothesis text or false positives, $M$ is the number of missing keywords in the hypothesis text, and $N$ is the number of keywords in the reference text [35]. The [35] presents a comparison between the WER and KER metrics. The conclusion is that KER, although it can give better results, is more volatile than WER. In [35], the GT was compared to three transcription sets. It can be inferred that for low values of WER, KER presents better results. For average values, WER and KER have similar performances. For high WER

values, as the study of this paper, KER shows lower performances. This conclusion does not discourage the use of KER. The approach used in this work stands out from the proposal in [35] by three aspects:

*Definition of keywords* A statistical method for defining keywords is proposed by [35]. It consists of the probability ratio of a word being present in a generic transcript and a domain-specific transcript. Terms whose probabilities in specific texts are much higher than probabilities in generic texts are considered keywords. Although this method allows the listing of keywords by unsupervised learning methods, it had as results some language vices ("um", "yeah", "ah") and words that, although belonging to the study domain, add little information ("click" or "server" in the context of a computer help center). Here, the keywords were listed based on the experience of the power operation center, such as the familiarity with the audio's content, technical terms, and proper names of the assets. Appendix A shows an example of the chosen keywords list. This approach avoids discrepancies such as the fact that the top five keywords are stop words in [35] (definition below).

*Removal of stop words* when analyzing the occurrence of words in a text, the first assumption may be that the most recurrent terms are the most relevant. This intuition, however, is the opposite of reality. The most frequent alternatives in any document will be common words such as "the", "a", "and" among others, that help to build ideas but which do not have meanings by themselves. Sometimes the relevance of some words can vary depending on the analyzed context. But it is usual to filter the most common words of a language (or Stop Words), during natural language preprocessing. Before analyzing KER, a Brazilian–Portuguese language Stop Words filter is implemented in our work. The objective is to remove pronouns, prepositions, conjunctions, and articles that have low relevance in the information transmitted by the text and avoid results as in [35], in which, among the 30 keywords listed, there are terms such as "wanna", "gotta" and "thank", examples of English stop words.

*Steamming* It is the process of agglutinating words, which have the same meaning, into a single representation, even if that representation has no meaning in the original language. This strategy plays a relevant role in quantifying the quality of the transcripts since the WER considers as substitutions the words transcribed with a degree, gender, or number different from the original, although the information itself is preserved. The application of stemming techniques in the pre-processing of transcripts is another item that complements the studies carried out by [35]. However, it is necessary to pay attention to the two types of errors that can happen in the process: over stemming, when there is excessive removal of letters so that words with different meanings are merged into the same stem; under stemming, when many letters are preserved so that words with the same meaning are grouped under different stems. For the steps of removing stop words and stemming, we use the MAC-MORPHO, a Brazilian–Portuguese part-of-speech tag [25]. It also has a set of functions for large-scale natural language data manipulation [36].

*Named entity recognition (NER)* Designed to extract structured information from unstructured texts, NER consists of identifying information units and classifying them into entities such as names and numerical expressions, such as times and dates [25, 36]. These features can be used to identify specific entities in useful contexts and compare data extracted from the hypothesis and GT transcripts. In this work,

three entities were listed: person (PER), organization (ORG), and location (LOC). Figure 1 shows the NER results found in a reference transcript and a candidate transcript from real audios of ENGIE's COS.

NER classification is not highly accurate. For example, in Fig. 1a the greeting "Hi" was classified as ORG. It is because the NER model also needs to be trained and adjusted. For call classification applications, for example, an adjusted model is critical for reliable results. However, for transcript comparisons, a functional model can be employed. By applying the same NER model to the GT and hypothesis transcripts, it is possible to highlight the differences between the transcripts. It is expected that the classification errors, when they exist, are equally applied to both texts.

To quantify the divergences or similarities of the NER applied to two texts, this work proposes indicators of the ratio for the occurrence of generic entities and by type, recognized in the GT text and the hypotheses transcripts, such as:

$$ER_{ent} = \frac{\left(E_{GT} - E_{hyp}\right)}{\left(E_{GT} + E_{hyp}\right)}, \tag{3}$$

$$ER_{ent} = \frac{\left(E_{GTent} - E_{hipent}\right)}{\left(E_{GT} + E_{hyp}\right)}, \tag{4}$$

where $ER$ is the ratio for entities recognized in hypothesis texts and GT, $E_{hyp}$ is the number of entities identified in the hypothesis text, $E_{GT}$ is the number of entities recognized in GT text, and $_{ent}$ is the predefined entities (PER, ORG, or LOC).

The ratios of recognized entities are values that can range from 1 to $-1$. The closer to zero, the more accurate the recognition is. A value of zero means that the same number of entities were recognized in the GT and the hypothesis. Positive values indicate that some GT entities were not recognized in the hypothesis. Negative values indicate the presence of false positives, that is, the hypothesis has more entities recognized than the GT.

In conclusion, the most suitable method for transcript's quality measuring is the WER. The KER is not suitable when used standalone, because it depends on the




(a) reference transcript.




(b) hypothesis transcription.

**Fig. 1** Examples of named entity recognition (NER)

list of keywords, so strongly task-dependent. NER does not evaluate directly, but it evaluates the performance of one specific system for entity recognition connected with the system.

### 2.2 Audios files processing

The control centers use a relatively restricted vocabulary and a fixed structure, with a proposal for standardized phraseology [37]. Many proper names are used, such as names of operators, assets, cities, and sometimes there is an overlap between speakers and background conversations. As an aggravating issue, the audios of the interlocutors in the local plants are subject to noise inherent to industrial activity. Despite the phraseology, relaxed speech is identified quite regularly. Based on the work of [29], the following improvements were made:

1. signal amplification: aiming to ensure that all files were audible, by people and models, gains of 2 dB were applied;
2. normalization: guarantees homogeneity in the amplitude of signals from all interlocutors;
3. high-pass filters: filters noise at frequencies lower than 120 Hz;
4. low-pass filters: filters noise at frequencies above 4 kHz.

In addition to audio processing, the call processing configuration has changed. Current files are stored at rates of 16 kHz. Until then, the ENGIE standard was 8 kHz. Also, the audios are recording in stereo format, where each of the speakers is saved in one channel. Both actions increase audio fidelity. During early development tests, the models were not performing well with an average WER of 0.65. Applying these audio procedures, the average WER dropped to 0.47, an improvement of about 28%.

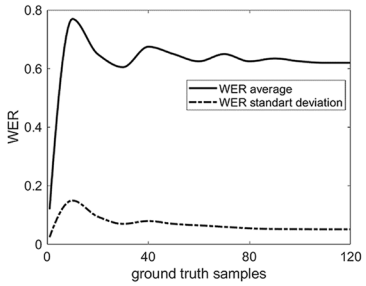
### 2.3 The ground truth (GT) size definition

In projects related to data science, it is necessary to define the minimum number of samples that allow the performance in an environment that reasonably represents reality [38]. The WER is a quality metric. For few samples, the standard deviation is likely to be high. The convergence and the standard deviation decrease are expected with more words. The stability point occurs when the addition of more samples does not interfere with the WER metric.

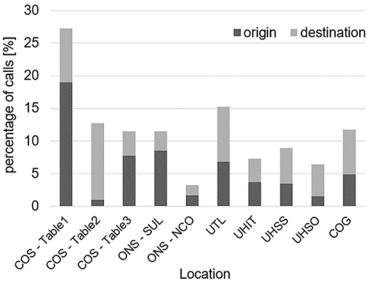
Initially, a GT with 50 random calls of manual transcripts was used. When checking the WER mean, it was observed that it showed a tendency to change, while the standard deviation showed a downward trend, suggesting the need for new samples. Figure 2a shows that the values stabilize around 120 samples. More transcripts have little influence on the mean and standard deviation values of WER. Thus, 120 samples are the minimum value necessary to represent the scenario of our work.

In addition to quantity, samples must effectively represent the entire data set. Beyond the information intrinsic to audio calls (~550,000 mp3 files) from ENGIE's

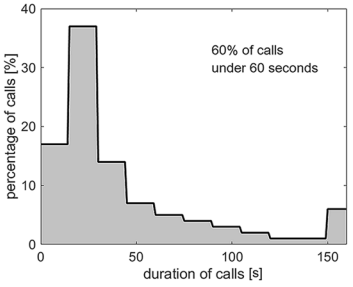
**Fig. 2** Ground truth (GT) features from ENGIE's control center



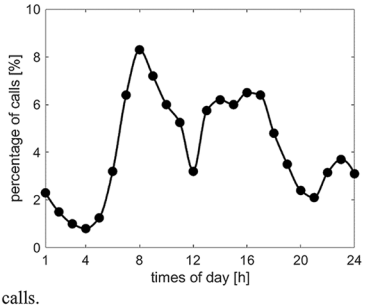

(a) WER versus ground truth samples.




(b) location of calls.




(c) duration of calls.




(d) time occurrence of calls.

power operation control center, each file has metadata such as origin and destination of calls, duration of calls, and time occurrence of calls. Figure 2b shows the percentage of calls from ten more actives locations [systems operation center (COS); generation operation center (COG); National Electric System Operator (ONS); Jorge Lacerda thermoelectric power plant (UTL), Itá hydroelectric power plant (UHIT); Salto Santiago hydroelectric power plant (UHSS); Salto Osório hydroelectric power plant (UHSO)]. We identify 4335 origin and 3786 destination locations in total. The ten more active locations represent 61% of the origin and 67% of the destination of all calls (~550,000). Therefore, ~0.25% of all locations are responsible for over two-thirds of calls.

Figure 2c shows the duration of calls. Typical calls from power control centers are short and direct, with approximately 77% of calls lasting up to 60 s. Thus, this group is the part that most faithfully represents the calls by the operation centers.

Regarding the time occurrence of calls (Fig. 2d), it is possible to observe that the distribution follows a similar pattern to the load modulation over a typical day. It is curious to observe features about sizing teams and defining shift change times. However, concerning the choice of audios to create the GT representation, Fig. 2d allows us to conclude that the best option can be to choose calls at random so that the GT samples reproduce the behavior of the data set, according to this criterion.

After analyzing the metadata and studying the convergence of the WER average, we defined that this paper uses as GT a mass of 200 transcripts, whose origin and destination belong to the set of the ten most recurrent locations. In addition, only audios with a duration of less than one minute are used, randomly chosen according to the time of the call.

### 2.4 STT transcription tool

We study five consolidated tools: Amazon Transcribe [39]; Google Speech-to-Text [40]; Retextoar [41]; IBM Watson Speech to Text [42] and Microsoft Azure Cognitive Services [43]. Working with market solutions has advantages and drawbacks. It can be a drawback that the technology is a black box. Legal and security issues must be considered when controlling critical infrastructure systems by cloud service. On the other hand, being used by multiple customers, and constantly patched and updated may be beneficial.

It is not the focus of this work to develop or to contribute to the state of the art of StT or automatic speech recognition methodologies but to explore the technical features of power operation control centers to contribute to a structured specific vocabulary. Therefore, we use the five tools in their free and open access format. The most relevant features regarding audio transcriptions apply to the power operation centers application are:

*AWS* It is possible to create a custom vocabulary. It allows you to add new words to the basic vocabulary or adapt your language type to generate accurate transcriptions for domain-specific words and phrases such as place names, technical terminology, or the names of individuals.

*Google* The asynchronous recognition module analyzes the audio data and initiates a long-duration operation. It has pre-trained libraries for several applications, such as generic call centers. Furthermore, the boost tool allows you to assign numerical weights to words or phrases according to how often they are recognized.

*Digitro* For companies with needs such as specialized vocabulary dictionaries and a large volume of transcription. It has an algorithm specially developed for the Brazilian Portuguese language.

*IBM* In its free version, IBM Watson does not allow the use of custom vocabulary.

*Azure* In its free version, the MACS service only supports audio files in WAV format, not supporting consolidated standards such as MP3. Batch audio transcription allows asynchronous transcription of large volumes of audio data and was the option used for this comparison.

No software solution is better in every aspect. Nevertheless, the IBM Watson solution did not perform well on any evaluated criteria. Retextoar, a Brazilian–Portuguese language natively developed, presented some intermediate results, although nonexceptional. These results confirm that both tools are considered, for now, insufficient for the audio transcription service in the power operation centers. The Azure solution was the best place for word identification, presenting solid results for NER, WER, or KER. Noteworthy, however, there is a limitation on the MP3 files used for the free versions. For solutions that do not store audio in this format or that may have a specific conversion step, this Azure feature is not a problem. The Google solutions were the most balanced, performing well in all criteria. The use of the boost tool did not substantially improve the transcript quality, and in some cases, the standard version of the solution showed better results than after customization. This behavior attests to the solution quality, but it is insufficient in terms of the exploitation by advanced users, willing to invest more time to optimize the transcription models. On the other hand, the AWS solution had natively one of the least interesting results. However, after setting up the custom vocabulary, the entity recognition performance was the best among all the solutions. The tool is not natively adapted to the Brazilian–Portuguese language, nor for the power operation centers. However, its customization has many possibilities. It allows advanced users to reach good results, as long as they are willing to invest in specialized vocabularies.

In conclusion, based on the studies reported here, we decided to use the AWS solution thanks to its high performance due to its custom dictionary transcripts. The StT transcription tool studies are not the main focus of the paper. Thus, comparisons must be cautious. The strength is in the power systems application.

### 2.5 Creating custom vocabularies

This section explores the customized vocabularies to add value to the StT tool, adapting it for the power operation control center as main contribution of this paper. It means correctly transcribing technical expressions, proper names, and considering variations of pronunciations. AWS allows specifying custom vocabularies in a simple word list format or by a four-column table [44]. In this work,

we explore table one. In this format, each row in the table represents a word or phrase and consists of four fields:

1. Phrase represents the word or phrase that must be recognized.
2. DisplayAs is an optional feature that defines how the word or phrase will be displayed in text format after transcription. It is possible to map a Phrase as 'Itá Hydroelectric Power Plant' to be displayed only as IHPP, for example. This feature can be used to standardize the representation, especially of terms that present some variation, so that keyword searches, call classification or any other type of traceability can be easily implemented in the transcripts.
3. The International Phonetic Alphabet (IPA) is a feature and can be used to specify the pronunciation of the word or phrase. Although quite useful for other more consolidated languages, this functionality is not complete for the Brazilian Portuguese language, because of its inability to support dialect-level diversity. Likewise, IPA can be difficult to input data accurately without expertise in phonetics and linguistics. It is why this field will not be used in the personalized vocabularies of this work.
4. SoundsLike describes a phoneme, however, SoundsLike allows you to split a word or phrase into smaller parts and provide pronunciation for each fraction, using the language's standard spelling to mimic the sound of the word. The pronunciation of the word "ANEEL", for example, can be described as "á-nê-él". As they are concurrent options, if the SoundsLike field is specified, it is not possible to use the AFI field and vice versa. In the vocabularies developed here, preference will be given to the SoundsLike feature. In addition to being able to be used to improve audio identification and word transcription, SoundsLike is critical for creating accent and pronunciation variations. To do so, it is necessary to create more than one line, with the same Phrase, but with different SoundsLike features, representing different pronunciations.

It was possible to identify the words to create a specialized vocabulary. It is our main objective. In its first version, the vocabulary already had more than 300 terms between specific words and phrases, except for pronunciation variations. Contrary to common sense, broad vocabularies do not necessarily represent an increase in the quality of transcriptions. In practice, the excess of specific terms ends up confusing the model, which uses similar words during transcription. Cases were found in which the standard version of AWS transcribed audio with higher quality than the model with an overpopulated vocabulary. Although the 300 terms are pertinent to the power operation context, many did not apply to most calls. Terms such as 'spillway', 'flow rate', and 'downstream', for example, are only relevant in the context of calls related to hydroelectric plants. In cases of connections to wind or thermal power plants, these expressions only end up polluting the vocabulary, attenuating the relevance of other pertinent words, and increasing the incidence of errors. To keep the expertise accumulated in an extensive vocabulary, but without causing errors due to information overload, it was

decided to divide the customized vocabulary into smaller and more specialized versions. The criterion is the nature of the assets that receive the call.

In this way, as the results of this paper, it was possible to create five personalized vocabularies for general operation (Appendix A), ENGIE's operation centers (Appendix B), hydroelectric (Appendix C), thermoelectric (Appendix D), and photovoltaic (Appendix E) power plants. See also next section.

To define the vocabulary used for transcription, the metadata of the destination phone of each audio was consulted. The solution proved to be efficient. It was possible to choose custom vocabulary options for each call without the risk of adding complexity to the process. For each kind, in addition to the proper names of the assets, some operational activity expression was listed, as well as some components of each type of plant.

## 3 Methodology summary and results

Figure 3 provides a visual aid of the developed methodology architecture. In the first layer, the audio files recorded during the power sector operation are stored according to the legislation (Sect. 2). The contribution at this stage concerns the recording of calls in high quality, with a sampling rate of 16 kHz or more.

The next step is the audio files pre-processing, seeking to reduce their signal-to-noise ratio. The treatments used showed significant gains (Sect. 2.2). The transcription occurs at this point. Researching for solutions, AWS was elected (Sect. 2.4). It was chosen thanks to the performance of the transcriptions and the high customization capacity of the vocabularies, even supporting pronunciation variations.

Also, during the transcription tests, we noticed that large vocabularies tend to reduce the quality of the transcriptions. Therefore, a solution was chosen that splits the vocabularies, according to the type of power plant.

At this point, custom vocabularies were created for operation centers, photovoltaic, hydroelectric, thermoelectric, and general power plants (Sect. 2.5). It is the main result and contribution of this work. The criteria for using each vocabulary are the destination phones for each call, stored as audio file metadata.

We made five custom vocabularies. It is relevant to highlight that they all share a single core composed of terms common to all vocabularies. These terms are present in all contexts, and to avoid repetition, they are presented only once in Appendix A. The other four are:

- Appendix B displays custom vocabulary extracted from calls destined to ENGIE's Generation Operations Center (COG). It has the names of the assets operated by the COG. Also, it has some jargon from the activities of hydroelectric and wind power plants, ENGIE's predominant assets.
- Appendix C displays custom vocabulary terms used in calls to hydropower plants. In addition to proper names and procedures, components of this type of asset are listed.

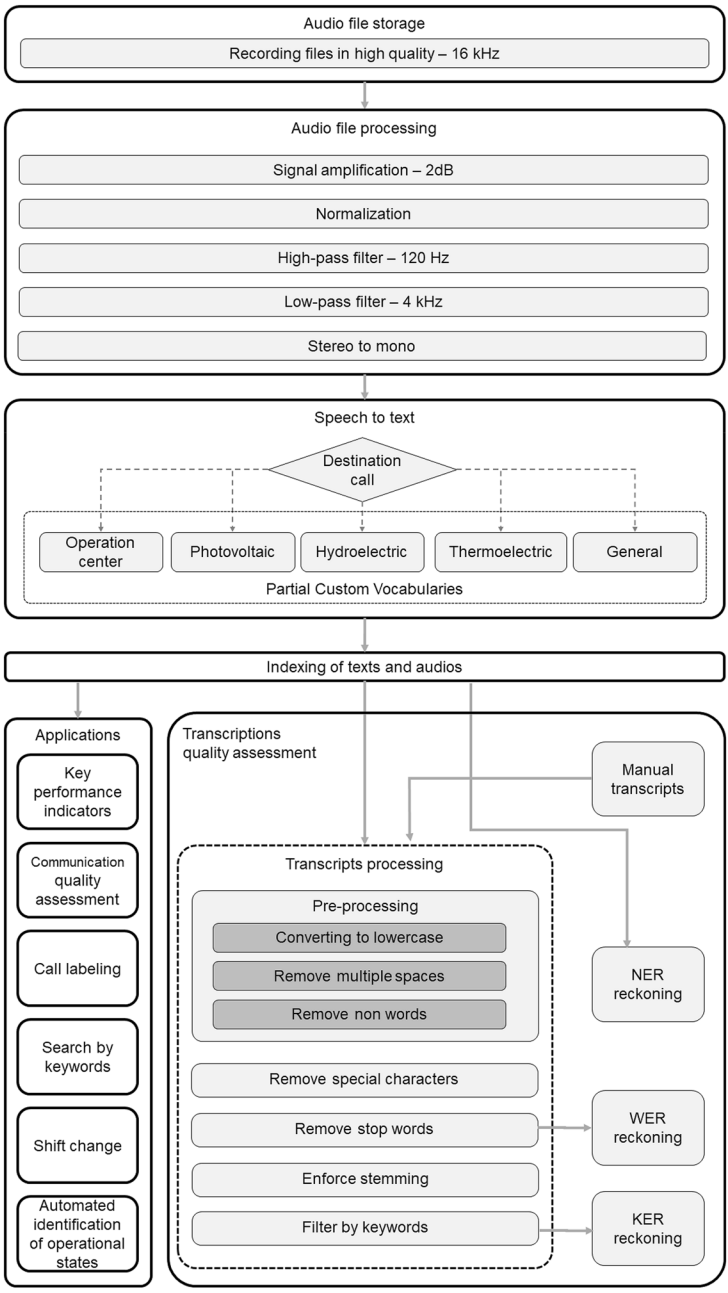



**Fig. 3** The architecture of the proposed methodology

- Similarly, Appendix D displays the terms of the custom vocabulary used in calls destined for thermoelectric plants. Some components and parts of this type of plant are listed in addition to the names of the thermoelectric assets.
- Appendix E displays the custom vocabulary used by calls to photovoltaic units. It is possible to observe that, for a type of generation still at an early stage within ENGIE, it is represented by a smaller vocabulary, considering few asset names and only a few specific technical terms.

Finally, the methodology for evaluating the quality of the transcripts is required (Sect. 2.1). The objective is to use the methodology to the model quality periodically check. Using new manual transcripts and monitoring the quality of the model, ensuring that it is not deteriorated over time and by any updates to the custom vocabularies.

After the transcriptions, the texts are saved and indexed with their respective audios. Many applications can use this database. Next section discusses some challenges faced by ENGIE's operation team. The motivation is to inspire future related works.

## 4 Audio transcription potential applications for power operation control centers

The methodology proposed here creates custom vocabularies for power operation control centers, as well as the transcription of their audio calls. Correlating the transcripts with their respective metadata (place of origin and destination, duration and time of audio calls as shown in Fig, 2), a database is structured capable of providing more assertive decision-making. Some possibilities are:

*Keyword search* Perhaps the most obvious and useful of applications, keyword-indexed search is relatively simple to implement and has an immediate value. The possibility of selecting audio calls for relevant terms increases the efficiency of post-operation teams, who are responsible, among other activities, for finding and sending evidence whenever requested and creating reports [8, 45]. In addition, other activities that require confirmation via audio may benefit. The value is even greater if we increment the filters with audio file metadata, such as the date and duration of calls and source and destination phones.

*Call classification* Although the amount of information about the power sector operation that must be reported via audio channels is large, most of them are well standardized and structured [28, 46], so that a solution capable of analyzing each of the transcriptions and classifying them into preexisting types would possibly have a high assertiveness rate. A suggestion for future works would be to classify the links in labels, according to the theme addressed in the link: tension control; generation adjustment; change of operating status of generators; reservoir levels; the amount of rain; etc.

*Assessment of communication quality* ONS establishes the rules, basic procedures, and standard phraseology that apply to the real-time teams of the ONS`s and companies' operation centers, in any and every operational voice communication,

aiming at objectivity, clarity, and effectiveness of the messages transmitted. Agents are responsible for periodically evaluating whether the quality of communication of their operators complies with the criteria listed in the Operation Procedure Manual [28]. Currently, at ENGIE, this task is performed manually by the regulation team, which randomly selects a sample of some audios from each operator, usually every quarter. Future work is the feasibility study of a computational intelligence solution that identifies the presence or absence of some criteria such as operator and center identification; message confirmation; clarity and objectivity; assertiveness; proper phraseology; and friendliness.

*Change* The agents' operation centers must have operation teams in uninterrupted shifts. The daily operation of ENGIE consists of three 8-h shifts. In each alternation of operators, the shift change takes place. The professionals who are being replaced contextualize their substitutes and provide them with relevant information to the operation of the assets. Another popular application that makes use of computational intelligence is automatic text summarization. In short, these applications can condense a lot of information, from several different sources, into small paragraphs [47]. A promising technological solution would be to explore the possibility of developing a tool that synthesizes the occurrences of each asset, mixing transcribed data, information from the supervisory system, and internal documentation in a given period, helping the shift change activity and highlighting conditions that demand more attention, contributing to the situational awareness of operators.

*Automatic chance of operating states* An electrical generator connected to the electrical system can transition between several operating states. Some of the most common and self-explanatory ones are: powered on as a generator; connected as a synchronous compensator, at the request of the ONS, for voltage control; automatically turned off by actuation of a protection or control system; turned off for operational convenience of the ONS or due to insufficiency of useful fall or inflow to hydroelectric plants; among dozens of other options. These events of changes in the operating state must be registered in the ONS database, through a specific application [28]. It is the responsibility of the ONS to provide generation companies, with data on changes in operating status and the availability of centralized plants [48]. When discrepancies are found, the audio recordings of the control centers are consulted to confirm the existence and timing of changes in operating states. In this context, a possible area of study concerns the automation of registries of change of operating status of generators, through the identification of transcripts, structuring of information, and automatic recording of calls that deal with this subject. In addition to the associated financial impact and the operational efficiency gain of the process, solutions of this type would be replicable for most agents connected to the electrical system.

*Key performance indicators (KPI)* KPIs can be created by treating transcripts as relational data. It is possible to find non-obvious relationships by crossing the information of calls with other data, such as operator schedule, issued documentation, and supervisory system data. Features such as calls seasonality, number of calls by type, recurrence of commands, and types of requests can be efficient inputs for dimensioning control center teams. Another indicators category can be related to the quality of the transcription tool itself. Average WER per asset, for example,

can identify phone quality, noise, or other pickup issues. Variations in the quality of transcripts may mean the degradation of the model. A high incidence of calls with long durations can point out that there are many non-standard calls, such as conversations with maintenance teams during interventions. There are endless possibilities, and the exploration and definition of KPIs is a topic to be explored in the future.

## 5 Conclusion

The way of operating power systems has been changing rapidly. Such dynamism is necessary to maintain operational reliability regardless of new challenges, such as the teleassistance of power plants and substations, energy planning with several intermittent sources, and the search for efficiency and competitiveness. These changes affect real-time operation teams, who need to adjust generation more frequently.

In this context, it is possible to recognize the potential of the proposed methodology: a structured vocabulary specific to power operation control centers. Regarding inputs, the current legislation determines that communication via voice channels. The availability of data allows their transformation into useful information to assist in making more assertive decisions for the operation of assets.

Concerning technology, this paper demonstrates transcription assertiveness considering the natural language and technical vocabulary. It is due to the development of specialized vocabularies with expressions, terms, phraseology, and proper names of power operation control centers. The results obtained allow the application of intelligence to the automatically transcribed texts, conserving the possibility for improvements and evolutions, depending on the users' feedback, ensuring that the quality of the transcriptions is incremental.

Our methodology creates a structured vocabulary specific to power operation centers. The natural way forward is to approach an automatic speech recognition tool as an end-to-end strategy [15–20]. We are interested in testing the bidirectional encoder representations from transformers (BERT) [49]. Unlike recent language representation models, the BERT model can be fine-tuned for a wide range of tasks, such as question answering, without substantial task-specific architecture modifications. Question answering is a relevant feature for power operation communication centers.

Whether simplifying the execution or increasing the quality of tasks already performed by the operation teams or even providing analyses that were unavailable, the applications presented potential and also would be objects of future work (Sect. 4). Using computational intelligence techniques, it is possible to cross-reference the transcribed data with information from other operating systems, such as supervisory systems and occurrence management, to generate valuable relationships for the power operation sector.

Whether simplifying the execution or increasing the quality of tasks earlier performed by the operation teams or even providing unavailable analyses, the applications presented potential and it would be future works (Sect. 4). It is possible to cross-reference the transcribed data with information from other operating systems

using computational intelligence techniques, such as supervisory systems and occurrence management, to generate valuable relationships for the power operation sector.

## Appendix A: Keywords for power operation control centers. Brazilian–Portuguese Language

| | | | |
|---|---|---|---|
| ONS | Partida a Frio | Subcorrente | Aerogerador |
| IHM | Tomadad'Agua | Subfrequência | Cana Brava |
| ENGIE | Fuga Terra | Montante | Estreito |
| PGI | Comando | Jusante | Itá |
| SGI | Evento | Compensador | Jaguara |
| ANEEL | Proteção | Perturbação | Machadinho |
| CNOS | Transformador | Sistêmica | Miranda |
| COSR | Geração | Umburanas | Passo Fundo |
| COS | Programação | Black start | Ponte de Pedra |
| COG | Válvula | Termelétrica | São Salvador |
| UHE | Partida | Pararaio | Salto Osório |
| PCH | Tubulação | Disjuntor | Salto Santiago |
| MUST | Indisponibilidade | Energizada | Lacerda A |
| Partida Quente | Comporta | Normalizada | Lacerda B |
| Sincronismo | Seccionadora | Restrição | Lacerda C |
| Oscilação | Inversor | Turbina | José Gelázio |
| Rotação | PotênciaAtiva | Caldeira | Rondonópolis |
| Turbina | PotênciaReativa | Precipitador | Cacimbas |
| Instabilidade | Fator de Potência | Vertimento | Estrela |
| Manutenção | Manobra | Vertedouro | Ouro Verde |
| Alarme | Retificador | Deplecionar | Santa Mônica |
| Gerador | Sobretemperatura | Amperes | Flecheiras |
| Falha | Sobretensão | Volts | Guajirú |
| Subtensão | Sobrecorrente | Watts | Mundaú |

## Appendix B: Personalized vocabulary example for ENGIE's operation centers common calls. Brazilian–Portuguese Language

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| Assú | a-sú | Assú |
| Aurora | au-ro-ra | Aurora |
| Biomassa | | biomassa |
| black-start | | black start |

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| C.O.G | cê-ó-gê | COG |
| Cacimbas | ca-cim-bas | Cacimbas |
| Campo-Largo | cam-po-lar-go | Campo Largo |
| Cana-Brava | ca-na-bra-va | Cana-Brava |
| Co-geração | co-ge-ra-ção | Co-geração |
| Complexo-termelétrico | com-ple-xo-ter-me-lé-tri-co | Complexo Termelétrico |
| conjunto-eólico | com-jun-to-ê-ó-li-co | conjunto eólico |
| Depleciona | | depleciona |
| Deplecionar | | deplecionar |
| emissão-de-particulado | | emissão de particulado |
| Eólico | | Eólico |
| Estreito | es-trei-to | Estreito |
| Estrela | es-tre-la | Estrela |
| Flecheiras | fle-xê-ras | Flecheiras |
| Fotovoltaica | | fotovoltaica |
| Grid | gri-di | Grid |
| Guajirú | gua-ji-rú | Guajirú |
| Hidrelétrica | | hidrelétrica |
| Ibitiúva | i-bi-ti-ú-va | Ibitiúva |
| Itá | i-tá | Itá |
| Jaguara | já-gua-ra | Jaguara |
| Jorge-Lacerda | jor-ge-la-cer-da | Jorge Lacerda |
| José-Gelázio | jo-zé-gê-lá-zio | José Gelázio |
| Jusante | | jusante |
| Lacerda | la-cer-da | Lacerda |
| Lages | la-ges | Lages |
| limpeza-do-condensador | | limpeza do condensador |
| Machadinho | ma-cha-di-nho | Machadinho |
| Miranda | mi-ram-da | Miranda |
| Montante | | montante |
| Mundaú | mun-da-ú | Mundaú |
| nível-do-reservatório | | nível-do-reservatório |
| Nova-Aurora | no-va-au-ro-ra | Nova Aurora |
| O.N.S | ó-enê-ésse | ONS |
| Ouro-Verde | ou-ro-ver-de | Ouro Verde |
| P.C.H | pê-cê-ágá | PCH |
| Pampa | pam-pa | Pampa |
| pampa-sul | pam-pa-sul | Pampa Sul |
| parque-eólico | par-kê-ê-ó-li-co | parque eólico |
| Passo-Fundo | pa-so-fun-do | Passo Fundo |
| Ponte-de-Pedra | pon-te-de-pe-dra | Ponte de Pedra |
| Precipitador | | precipitador |
| Reservatório | | reservatório |

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| Rondonópolis | ron-do-nó-po-lis | Rondonópolis |
| Salto-Osório | sal-to-o-só-rio | Salto Osório |
| Salto-Santiago | sal-to-san-ti-a-go | Salto Santiago |
| Santa-Mônica | san-ta-mô-ni-cá | Santa Mônica |
| São-Salvador | são-sal-va-dor | São Salvador |
| Síncrono | | síncrono |
| Sobreposição | so-bre-po-si-ção | sobreposição |
| Termelétrica | | termelétrica |
| Trairi | tra-í-ri | Trairi |
| Tubarão | tu-ba-rão | Tubarão |
| Turbina | | turbina |
| U.T.L.A | ú-tê-éle-á | UTLA |
| U.T.L.B | ú-tê-éle-bê | UTLB |
| U.T.L.C | ú-tê-éle-cê | UTLC |
| Umburanas | um-bu-râ-nas | Umburanas |
| Vertido | | vertido |
| Vertimento | | vertimento |

## Appendix C: Personalized vocabulary example for hydroelectric power plant operation calls Brazilian–Portuguese language

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| Reestabelecimento | res-ta-be-le-ci-men-to | reestabelecimento |
| Compensador | | Compensador |
| Compensar | | Compensar |
| Depleciona | | Depleciona |
| Deplecionar | | Deplecionar |
| Estreito | es-trei-to | Estreito |
| Hidrelétrica | | Hidrelétrica |
| Itá | i-tá | Itá |
| Jusante | | Jusante |
| Montante | | Montante |
| nível-do-reservatório | | nível-do-reservatório |
| P.C.H | pê-cê-ágá | PCH |
| reduz-o-vertimento | | reduz o vertimento |
| Reservatório | | Reservatório |
| Salto-Osório | sal-to-o-só-rio | Salto Osório |
| Síncrono | | Síncrono |
| Sobreposição | so-bre-po-si-ção | Sobreposição |
| Turbine | | Turbine |
| Vertimento | | Vertimento |

## Appendix D: Personalized vocabulary example for thermoelectric power plant operation calls. Brazilian–Portuguese language

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| Biomassa | | Biomassa |
| Caldeira | | Caldeira |
| Co-geração | co-ge-ra-ção | Co-geração |
| Complexo-termelétrico | com-ple-xo-ter-me-lé-tri-co | Complexo Termelétrico |
| emissão-de-particulado | | emissão de particulado |
| Ibitiúva | i-bi-ti-ú-va | Ibitiúva |
| Jorge-Lacerda | jor-ge-la-cer-da | Jorge Lacerda |
| Lacerda | la-cer-da | Lacerda |
| Lages | la-ges | Lages |
| limpeza-do-condensador | | limpeza do condensador |
| Pampa | pam-pa | Pampa |
| pampa-sul | pam-pa-sul | Pampa Sul |
| Sobreposição | so-bre-po-si-ção | Sobreposição |
| Termelétrica | | Termelétrica |
| Turbina | | Turbine |
| U.T.L.A | ú-tê-éle-á | UTLA |
| U.T.L.B | ú-tê-éle-bê | UTLB |
| U.T.L.C | ú-tê-éle-cê | UTLC |

## Appendix E: Personalized vocabulary example for photovoltaic power plant operation calls Brazilian–Portuguese language

| Phrase | SoundsLike | DisplayAs |
|---|---|---|
| Aurora | au-ro-ra | Aurora |
| conjunto-eólico | com-jun-to-ê-ó-li-co | conjunto eólico |
| Eólico | | eólico |
| Grid | gri-di | Grid |
| Nova-Aurora | no-va-au-ro-ra | Nova Aurora |
| Painél | | painél |
| Tubarão | tu-ba-rão | Tubarão |

## References


1. ONS (2020) PR 2.16, Procedimentos de Rede: Submódulo 2.16 Requisitos operacionais para os centros de operação e instalações da Rede de Operação. http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%202.16-RQ_2020.12.pdf. Accessed 25 Mar 2022
2. Cuisinier, E., Bourasseau, C., Ruby, A., Lemaire, P., Penz, B.: Techno-economic planning of local energy systems through optimization models: a survey of current methods. Int J Energy Res (2021). https://doi.org/10.1002/er.6208
3. Bjarghov, S., et al.: Developments and challenges in local electricity markets: a comprehensive review. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3071830
4. ONS (2020) DESSEM – Entenda o que muda no ONS com a entrada em operação do DESSEM em janeiro de 2020. http://www.ons.org.br/Paginas/Noticias/20190807_oquemudacomodessem.aspx. Accessed 25 Mar 2022
5. ANEEL (2022) Geração Distribuída. https://bit.ly/3x8W8At. Accessed 25 Mar 2022
6. Yang, B., Yu, L., Chen, Y., Ye, H., Shao, R., Shu, H., Yu, T., Zhang, X., Sun, L.: Modelling, applications, and evaluations of optimal sizing and placement of distributed generations: a critical state-of-the-art survey. Int J Energy Res (2021). https://doi.org/10.1002/er.6104
7. Chen, B., Wang, J., Lu, X., Chen, C., Zhao, S.: Networked microgrids for grid resilience, robustness, and efficiency: a review. IEEE Trans Smart Grid (2021). https://doi.org/10.1109/TSG.2020.3010570
8. ONS (2020) PR 6.2 Procedimentos de Rede: Análise da operação, ocorrências e perturbações e acompanhamento das providências, http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.2-OP_2020.12.pdf. Accessed 25 Mar 2022
9. Pinzón JD, Arrieta L (2021) Data analytics for power grid control centers management. In: IEEE PES innovative smart grid technologies conference. https://doi.org/10.1109/ISGTLatinAmerica52371.2021.9543087
10. Valencia A, Pinzón JD (2021) Asset management model of SCADA infrastructure of power control centers based on indicators. In: IEEE PES innovative smart grid technologies conference.https://doi.org/10.1109/ISGTLatinAmerica52371.2021.9543089
11. Sabzi, H.Z., Abudu, S., Alizadeh, R., Soltanisehat, L., Dilekli, N., King, J.K.: Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources. Energy Sources Part A Recov Util Environ Eff (2018). https://doi.org/10.1080/15567036.2018.1476934
12. Wood, D.A.: Feature averaging of historical meteorological data with machine and deep learning assist wind farm power performance analysis and forecasts. Energy Syst (2022). https://doi.org/10.1007/s12667-022-00502-x
13. Lisi, F., Shah, I.: Forecasting next-day electricity demand and prices based on functional models. Energy Syst (2020). https://doi.org/10.1007/s12667-019-00356-w
14. Roy, A., Bai, L.: Estimating utilities of price-responsive electricity consumers. Energy Syst (2022). https://doi.org/10.1007/s12667-021-00496-y
15. López-Espejo, I., Tan, Z.-H., Hansen, J.H.L., Jensen, J.: Deep spoken keyword spotting: an overview. IEEE Access (2022). https://doi.org/10.1109/ACCESS.2021.3139508
16. Audhkhasi, K., Rosenberg, A., Saon, G., Sethy, A., Ramabhadran, B., Chen, S., Picheny, M.: Recent progress in deep end-to-end models for spoken language processing. IBM J Res Dev (2017). https://doi.org/10.1147/JRD.2017.2701207
17. Bhable SG, Kayte CN (2020) Review: multilingual acoustic modeling of automatic speech recognition (ASR) for low resource languages. In: IEEE international conference on advent trends in multidisciplinary research and innovation.https://doi.org/10.1109/ICATMRI51801.2020.9398431
18. Jorge, L., Giménez, A., Silvestre-Cerdà, J.A., Civera, J., Sanchis, A., Juan, A.: Live streaming speech recognition using deep bidirectional LSTM acoustic models and interpolated language models. IEEE/ACM Trans Audio Speech Lang Process (2022). https://doi.org/10.1109/TASLP.2021.3133216
19. Zhang, W., Zhai, M., Huang, Z., Liu, C., Li, W., Cao, Y.: Towards end-to-end speech recognition with deep multipath convolutional neural networks. In: Yu, H., Liu, J., Liu, L., Ju, Z., Liu, Y., Zhou, D. (eds.) Intelligent robotics and applications. Lecture notes in computer science, vol. 11745. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27529-7_29

20. Płaza, M., Pawlik, L., Deniziak, S.: Call transcription methodology for contact center systems. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3102502
21. Sheikh, I., Fohr, D., Illina, I., Linarès, J.: Modelling semantic context of OOV words in large vocabulary continuous speech recognition. IEEE/ACM Trans Audio Speech Lang Process (2017). https://doi.org/10.1109/TASLP.2017.2651361
22. Tian, J., Yu, J., Weng, C., Zou, Y., Yu, D.: Improving mandarin end-to-end speech recognition with word N-gram language model. IEEE Signal Process Lett (2022). https://doi.org/10.1109/LSP.2022.3154241
23. Pakoci ET, Popović BZ (2021) Recurrent neural networks and morphological features in language modeling for serbian. In 29th telecommunications forum. https://doi.org/10.1109/TELFOR52709.2021.9653410
24. Escobar-Grisales D et al (2021) Colombian dialect recognition based on information extracted from speech and text signals. In: IEEE automatic speech recognition and understanding workshop.https://doi.org/10.1109/ASRU51503.2021.9687890
25. NILC (2016) Núcleo Interinstitucional de Linguística Computacional, USP. http://nilc.icmc.usp.br/macmorpho. Accessed 25 Mar 2022
26. Kashevnik, A., et al.: Multimodal corpus design for audio-visual speech recognition in vehicle cabin. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3062752
27. Rui, L.: Does game-based vocabulary learning APP influence chinese EFL learners' vocabulary achievement, motivation, and self-confidence? SAGE Open (2021). https://doi.org/10.1177/21582440211003092
28. ONS (2022) MPO 5.13 Manual de Procedimentos da Operação: comunicação verbal na operação. http://www.ons.org.br/%2FMPO%2FDocumento%20Normativo%2F4.%20Rotinas%20Operacionais%20-%20SM%205.13%2F4.1.%20Rotinas%20Gerais%2F4.1.7.%20Relacionamento%20Operacional%2FRO-RO.BR.01_Rev.12.pdf. Accessed 25 Mar 2022
29. ONS (2022) PR 2.15. Procedimentos de Rede: Requisitos mínimos para telecomunicações. http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%202.15-RQ_2020.12.pdf. Accessed 25 Mar 2022
30. Torcoli, M., Kastner, T., Herre, J.: Objective measures of perceptual audio quality reviewed: an evaluation of their application domain dependence. IEEE/ACM Trans Audio Speech Lang Process (2021). https://doi.org/10.1109/TASLP.2021.3069302
31. Errattahi, R., El Hannani, A., Ouahmane, H.: Automatic speech recognition errors detection and correction: a review. Procedia Comput Sci (2018). https://doi.org/10.1016/j.procs.2018.03.005
32. Morris AC, Maier V, Green P (2004) From WER and RIL to MER and WIL: improved evaluation. In: International conference on spoken language processing. Corpus ID: 18880375
33. Liu, Y., Lu, Y., He, Y., Zhao, X., Zhang, C.: Marker codes using the decoding based on weighted Levenshtein distance in the presence of insertions/deletions. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.3037169
34. Sari, L., Hasegawa-Johnson, M., Yoo, C.D.: Counterfactually fair automatic speech recognition. IEEE/ACM Trans Audio Speech Lang Process (2021). https://doi.org/10.1109/TASLP.2021.3126949
35. Park Y et al (2018) An empirical analysis of word error rate and keyword error rate. In: International conference on spoken language processing. Corpus ID: 1966719
36. NLTK: Natural Language Toolkit (2022). https://www.nltk.org/. Accessed 25 Mar 2022
37. Tran, O.T., Bui, V.T.: Neural text normalization in speech-to-text systems with rich features. Appl Artif Intell (2021). https://doi.org/10.1080/08839514.2020.1842108
38. Strobel PB, Clematide MVS (2020) How much data do you need? About the creation of a ground truth for black letter and the effectiveness of neural OCR. In: Conference on language resources and evaluation. Corpus ID: 218974414. https://doi.org/10.5167/UZH-197209
39. Amazon: AWS Transcribe (2022). https://aws.amazon.com/transcribe. Accessed 25 Mar 2022
40. Google (2022) Speech-to-text reconhecimento de fala automático. https://cloud.google.com/speech-to-text. Accessed 25 Mar 2022
41. DÍGITRO (2022) Retextoar transcrição rápida e eficiente de conteúdos. https://www.digitro.com/solucao/retextoar-transcricao-rapida-e-eficiente-de-conteudos. Accessed 25 Mar 2022
42. IBM (2022) Watson speech to text. https://www.ibm.com/cloud/watson-speech-to-text. Accessed 25 Mar 2022.
43. Microsoft (2022) AZURE speech to text. https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/. Accessed 25 Mar 2022

44. Amazon (2021) AWS Improving domain-specific transcription accuracy with custom language models. https://docs.aws.amazon.com/transcribe/latest/dg/custom-language-models.html. Accessed 25 Mar 2022
45. ONS (2020) PR 6.3 Elaboração do Relatório de Análise de Perturbação. http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.3-RS_2020.12.pdf. Accessed 25 Mar 2022
46. ONS (2020) PR 5.1 Procedimentos de Rede: Operação do sistema e das instalações da rede de operação. http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%205.1-RS_2020.12.pdf. Accessed 25 Mar 2022
47. Gambhir, M., Gupta, V.: Recent automatic text summarization techniques: a survey. Artif Intell Rev (2017). https://doi.org/10.1007/s10462-016-9475-9
48. ONS (2020) PR 6.5 Apuração da geração e de indisponibilidade de geração. http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.5-RS_2020.12.pdf. Accessed 25 Mar 2022
49. Deping L, Hongjuan W, Mengyang L, Pei L (2020) News text classification based on bidirectional encoder representation from transformers. In: Conference on artificial intelligence, big data and algorithms. https://doi.org/10.1109/CAIBDA53561.2021.00036

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.