How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?

Alain Ghio, Gilles Pouchoulin, Bernard Teston, Serge Pinto, Corinne Fredouille, Céline de Looze, Danièle Robert, François Viallet, Antoine Giovanni

▶ To cite this version:

Alain Ghio, Gilles Pouchoulin, Bernard Teston, Serge Pinto, Corinne Fredouille, et al.. How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?. Speech Communication, 2012, 54 (5), pp.664-679. 10.1016/j.specom.2011.04.002 . hal-01317198

HAL Id: hal-01317198

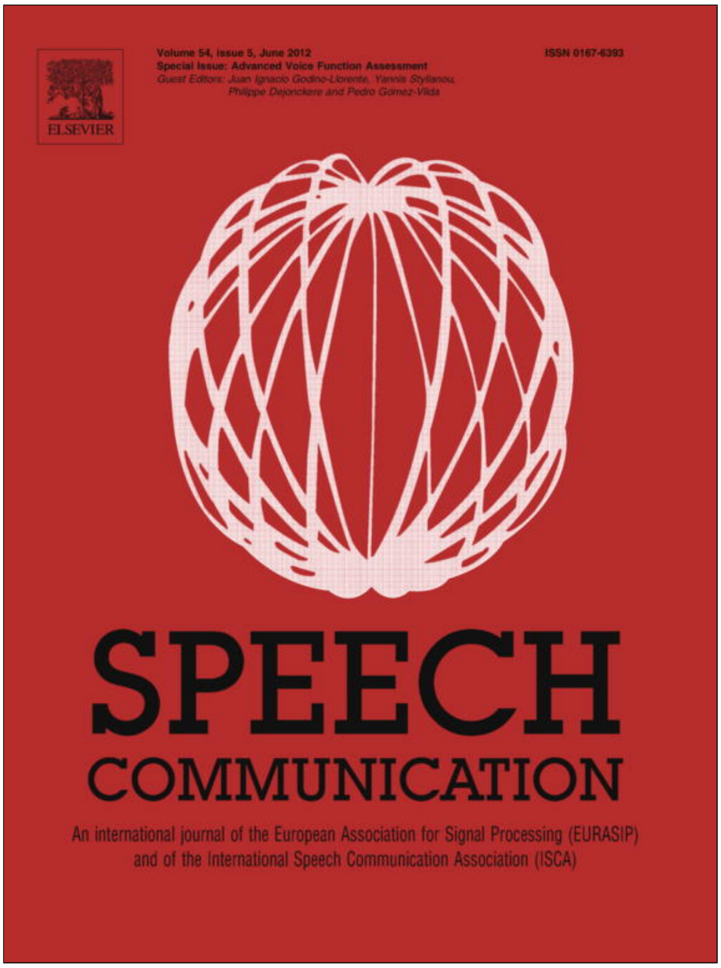
https://hal.science/hal-01317198

Submitted on 19 Apr 2018

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

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

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Speech Communication 54 (2012) 664–679

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?

A. Ghio [a,*], G. Pouchoulin [a,b], B. Teston [a], S. Pinto [a], C. Fredouille [b], C. De Looze [a], D. Robert [a,d], F. Viallet [a,c], A. Giovanni [a,d]

[a] *LPL, Laboratoire Parole et Langage, CNRS UMR 6057, Aix-Marseille University, France*
[b] *LIA, Laboratoire d'Informatique d'Avignon, Avignon University, France*
[c] *Service de neurologie, Centre Hospitalier du Pays d'Aix, Aix-en-Provence, France*
[d] *Service ORL, Centre Hospitalier Universitaire de la Timone, Marseille, France*

Available online 6 May 2011

Abstract

The aim of this contribution is to propose a database model designed for the storage and accessibility of various speech disorder data including signals, clinical evaluations and patients' information. This model is the result of 15 years of experience in the management and the analysis of this type of data. We present two important French corpora of voice and speech disorders that we have been recording in hospitals in Marseilles (MTO corpus) and Aix-en-Provence (AHN corpus). The population consists of 2500 dysphonic, dysarthric and control subjects, a number of speakers which, as far as we know, constitutes currently one of the largest corpora of "pathological" speech. The originality of this data lies in the presence of physiological data (such as oral airflow or estimated sub-glottal pressure) associated with acoustic recordings. This activity led us to raise the question of how we can manage the sound, physiological and clinical data of such a large quantity of data. Consequently, we developed a database model that we present here. Recommendations and technical solutions based on MySQL, a relational database management system, are discussed.
© 2011 Elsevier B.V. All rights reserved.

Keywords: Voice/speech disorders; Dysphonia; Dysarthria; Database; Clinical phonetics

1. Introduction

Advanced research on the assessment of voice and speech disorders requires the structuring and organisation of a large set of data. Speech is highly variable even in "normal" conditions, and disorders can manifest in various ways from one patient to another. Therefore, a very large data set is necessary to obtain significant statistical results which are clinically reliable, valid and generalizable to other clinical cases. Research also requires that information relative to the speakers and their environment be archived with the speech data in a transparent manner and easily accessible. In fact, for the purposes of research, the speech signal is useless if it is not linked to the speaker's clinical state or the context of speech production.

1.1. Voice and speech disorders assessment: an interdisciplinary challenge

For about fifteen years, interest in studies on voice and speech disorders has extended beyond the simple framework of clinical research, with dysfunctional speech corpora attracting the attention of researchers from speech sciences laboratories and the field of communication and computer sciences. Comparison of "normal speech"

* Corresponding author. Address: CNRS – Université de Provence, Laboratoire Parole & Langage, 5 avenue Pasteur, BP 80975, 13604 Aix en Provence Cedex 1, France. Tel.: +33 (0)4 42 95 36 30/34; fax: +33 (0)4 42 95 37 88.
E-mail address: alain.ghio@lpl-aix.fr (A. Ghio).
URL: http://www.lpl-aix.fr/~ghio (A. Ghio).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.04.002

Author's personal copy

corpora, models or tools with their counterparts in dysfunction situations enriches clinicians', linguists' and engineers' knowledge and understanding of speech phenomena.

The question of variations and variability must be considered from different points of view: What is a "normal" variation? What is an abnormal one? Where is the boundary between "standard" variability and pathological deviation? Investigation of these questions calls for interdisciplinary co-operation, as in the work conducted by Hardcastle and Gibbon, at Queen Margaret College in Edinburgh (UK), where the electropalatography technique, originally developed for phonetic studies, is now used with speakers suffering from articulation disorders. This work opened a new field of investigation leading to numerous publications (Dent et al., 1995; Gibbon et al., 1998) as well as a regional health network, CleftNet (see http://www.qmu.ac.uk/ssrc/cleftnet/epghome.htm). Similarly, the European COST Action 2103 "Advanced Voice Function Assessment" is a joint initiative of speech processing engineers, laryngologists and phoniatricians to foster progress in the clinical assessment and enhancement of voice quality (see http://www.cost2103.eu).

The need for such a multidisciplinary approach stems from the characteristics and constraints induced by "pathological speech" data.

1.2. Characteristics and constraints induced by "pathological speech"

We use the expression "pathological speech" to refer to speech produced by speakers suffering from voice and speech dysfunctions, such as dysphonia and dysarthria. Dysphonia is a voice disorder, resulting from an organic lesion of the larynx and/or a phonation dysfunction. Dysarthria is a speech disorder resulting from neurological injury: any speech subsystem (respiration, phonation, articulation and movements of organs, prosody...) may be affected (Darley et al., 1975). Voice assessment is an important activity in the follow-up of dysarthric speakers.

Studying pathological speech requires:

(1) High-quality signals, so that distortions and noise may not be attributed to voice or speech dysfunctions. For instance, the harmonic-to-noise ratio (Yumoto et al., 1982), which measures the periodic vs. noisy components of voice, is consistent only if applied to a signal with low electronic or environmental noises.
(2) Linguistic material (utterances produced by the speakers) which provides sufficient relevant information for research. Sustained vowels are necessary to assess phonation mechanism but continuous speech is indisputably more natural from a spoken communication point of view (Parsa and Donald, 2001).
(3) Sufficiently accurate clinical information in order to manage various sets of speakers and conditions (pathology, therapy, clinical context of recordings, etc.).
(4) A large number of speakers. Any generalization of a specific clinical population requires data from many speakers because of the very-high inter-speaker variability encountered (different pathologies, individual compensation strategies, the severity and specificity of diseases, etc.).

Because some pathologies are rare and because it is not always easy to record some patients, the acquisition of pathological speech data is difficult. For these reasons, it is important to capitalise on existing recordings. However, to be useable, these recordings must satisfy the requirements mentioned above. Most often, the difficulties faced in voice and speech dysfunction studies are relative to data access.

1.3. Dissemination and loss of data

First, because of the heterogeneity of format and metadata, data may be hard to share. The information on the conditions, instructions for the recordings, the phonetic material used (e.g. sustained vowels, reading text, spontaneous speech) may be insufficient to easily run analyses: it is often necessary to browse the whole corpus in order to extract useful sequences required for a specific study. The local management of metadata (i.e., the information related to the speaker, the pathology, treatments...) may not be available. Consequently, it is difficult to obtain clear results as the homogeneity of the tested population is not confirmed. Lastly, some results obtained using perceptual judgements or instrumental assessment cannot be compared because of the differences in the methods used for data collection and analysis and especially because of the lack of information regarding the protocols employed in these tasks. The difficulties in accessing and distributing data are also due to the variety of formats used to record and store data. In France, for example, different centres store pathological speech corpora in various ways, including analog audio tapes, digital tapes, Iomega© storage, Zip disk, floppies, CD-Roms, DVDs, and hard disks. Some of these media have become or are likely to become obsolete and some of their reading devices are becoming rare (e.g. Revox © tape player). As some corpora become old and deteriorate over time (particularly analog recordings), and as their reading devices no longer exist, data may be unreadable (Fougeron et al., 2010). In addition, "physiological" signals such as electroglottography, oral airflow, nasal airflow or intra-oral pressure may also be associated with the sound tracks. These data, which are frequently stored in a system-specific format, may require physical access from their site of production, with attendant difficulties. In light of these factors, it is clear that cumulative progress in research in this area requires that corpora be durable, easily disseminated and shared.

Author's personal copy

1.4. Our objectives

We present two important French corpora of voice and speech disorders, recorded over the last 15 years in Marseilles (MTO corpus) and Aix-en-Provence (AHN corpus) hospitals. We also propose a database model designed for storage and provision of access to various speech disorder data. We aim to present the concepts relating to the database design in terms understandable to all and not only computer scientists.

2. Corpora of voice and speech disorders

2.1. Current situation

Several committees and programs have been established to encourage and promote interaction and cooperation in the areas of Spoken Language Processing, including ARPA (Klatt, 1980), SAM (Speech Assessment Methodology; Fourcin et al., 1989), ESPRIT-ACCOR (Marchal and Hardcastle, 1993), EAGLES (Expert Advisory Group on Language Engineering Standards; Gibbon et al., 1997), COCOSDA (**C**ommittee for the **Co**-ordination and **S**tandardisation of Speech **D**atabases and **A**ssesment Techniques; http://www.cocosda.org), Linguistic Data Consortium (http://www.ldc.upenn.edu), ELRA (European Language Resources Association; http://www.elra.info), and more recently for French, the PFC project (Phonologie du Français Contemporain; Durand et al., 2002).

A wide range of speech databases have been collected. However, these databases were originally meant for the development of speech synthesis/recognition and possibly for linguistic research. Also, available databases of voice and speech disorders are extremely rare. Databases of voice and speech disorders fall into two categories. The first type are produced in speech technology laboratories. In general, they are designed to test human-machine interfaces and include very few speakers. An example of this type is the "Whitaker" database of dysarthric speech (Deller et al., 1993), a collection of 19,275 isolated-word utterances spoken by six cerebral palsy patients whose speech spans a broad spectrum of dysarthria. Another example is the "Nemours database of dysarthric speech" (Menendez-Pidal et al., 1996), a collection of 814 short nonsense sentences, 74 sentences and two connected-speech paragraphs produced by eleven male speakers with varying degrees of dysarthria. This database was designed to test the intelligibility of dysarthric speech before and after samples were enhanced using various signal processing methods. A recent initiative by Kim et al. (2008) proposes a "Dysarthric Speech Database for Universal Access Research". This resource is designed for automatic speech recognition development for people with neuromotor disability. Speech materials consist of 765 isolated words produced by nineteen speakers with cerebral palsy: 300 distinct uncommon words and three repetitions of digits, computer commands, radio alphabet and common words.

The second type of disordered speech database, using a wide range of speakers, is found in clinical institutions but the origin and organisation of data is rarely detailed. An example of this type of database use can be seen in Kim et al. (2011):

> "*One hundred seven subjects with dysarthria …were selected for the present study from the UW-Madison–Mayo Clinic dysarthria database, which consists of digital speech recordings obtained at the Mayo Clinic in Rochester, MN. Parkinson Disease, Stroke, and Traumatic Brain Injury groups were chosen for this study*".

However, to our knowledge, no precise information is available on this dysarthria database.

A widely disseminated database of audio signals from a large set of speakers is "The Disordered Voice Database", developed by the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab., which includes more than 1400 voice samples (sustained /a/ and the first 12 s of the Rainbow Passage) from approximately 700 subjects. This database has been developed for the acoustical and perceptual analysis of disordered voices for either clinical or research applications. It includes samples from patients with a wide variety of organic, neurological, traumatic, psychogenic, and other voice disorders. Although it is the most widely available of all the voice quality databases, Saenz-Lechon et al. (2006) outlined key points that should be carefully taken into account when using this database: normal and pathological voices were recorded at different locations; files are already edited to include only the stable part of the phonation; there is only one phonation per patient and visit. At present, no comparable disordered voice database is available for languages other than American English. This situation has led us to develop a comparable database for French, as described below.

2.2. Dysphonic patients: MTO (Marseille Timone ORL) corpus

For over fifteen years, recordings have been made of dysphonic patients at the ENT department of the Timone University Hospital in Marseille. For logistics reasons, the patients' information is handwritten into notebooks in which relevant data including the speakers' identity, pathology, examination date, the pre/post-operatory context, etc. are indicated. The resulting recordings and notebooks were digitized and indexed to build a database of 1530 dysphonic patients producing sustained vowels, reading a text, singing a song. The database contains a total of 1953 recording sessions (some speakers are recorded several times), with data from 504 men and 1026 women, of whom 332 were recorded several times (e.g. before and after surgery). The main pathologies are represented in Fig. 1.

Most of the voice productions (1766 sessions) were perceptually evaluated using the GRBAS scale (Hirano, 1981),

Author's personal copy

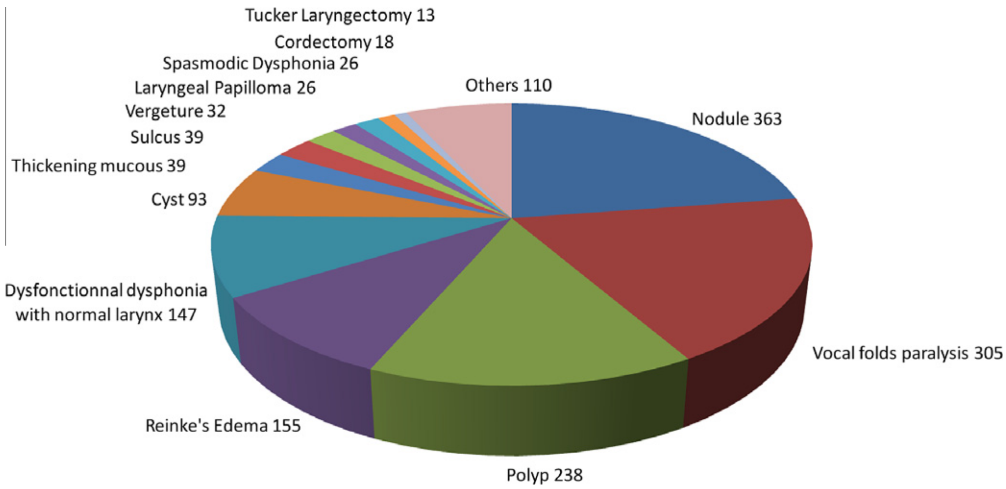


Fig. 1. Vocal pathologies represented in the MTO corpus.

where five dimensions are rated on a scale of four levels (0 = normal, 1 = light, 2 = medium, 3 = severe). These dimensions are:

1. Hoarseness (G) – the amount of noise in the produced sound. This dimension is considered the Global or General level of dysphonia.
2. Roughness (R), in relation to the irregular fluctuation of the fundamental frequency.
3. Breathiness (B), the level of additive turbulence noise in the produced sound.
4. Asthenia (A), the overall weakness of voice.
5. Strain (S): the tenseness of voice, overall muscular tension (S).

In the MTO corpus, dimension G is evaluated by one speech therapist during the recording session. As an isolated perceptual assessment, this value must be considered as an approximate level of the dysphonia (Fig. 2).

A significant number of publications on voice assessment are based on the MTO corpus (Giovanni et al., 1999a,b, 2002; Revis et al., 2002; Robert et al., 1999; Yu et al., 2001, 2007). The corpus is currently being used for advanced voice assessment using automatic techniques (Fredouille et al., 2005, 2009; Bonastre et al., 2007; Pouchoulin et al., 2007).

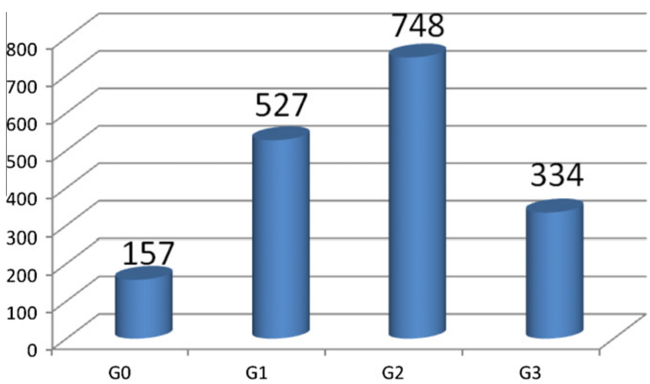


Fig. 2. Dysphonia severity repartition in the MTO corpus.

2.3. Dysarthric patients: AHN (Aix Hospital Neurology) corpus

Recordings have also been made over the last 15 years of dysarthric patients coming for medical consultations at the Neurology Department in Pays d'Aix Hospital. A computerised form is used to store clinical data. We have collected the sound and aerodynamic recordings of 990 patients and 160 age-matched control subjects. It is important for the purposes of voice assessment to match the control population with the patients especially if we want to take into account the effect of age on voice characteristics (Baken and Orlikoff, 2000; Hixon et al., 2008). Recordings include patients presenting various neuromotor disorders: stroke, amyotrophic lateral sclerosis (ALS), Friedreich's disease, Huntington's disease... The corpus is largely composed of recordings of patients with Parkinson's disease (601), and Parkinsonian syndromes (98) – group of neurodegenerative diseases displaying the classical features of Parkinson's disease (e.g. tremor, rigidity, akinesia) with additional characteristics distinct from simple idiopathic Parkinson's disease. Parkinsonian syndromes include multiple system atrophy (MSA), progressive supranuclear palsy (PSP), and corticobasal degeneration (CBD). Parkinson's patients typically have a low-volume voice (hypophonia) with monotone expressionless prosody.

Voice assessment in such a neurological disease is an important challenge (Pinto et al., 2010). The speech production disorders associated with Parkinson's disease represent a major factor of handicap, for which medical correction becomes more and more difficult as the disease progresses. Surgical treatment by subthalamic nucleus stimulation gives variable results on voice quality and speech intelligibility, whereas other motor disorders are remarkably improved. Rehabilitation, which constitutes a significant social cost, is regarded as useful, but its

Author's personal copy

effectiveness remains unknown and the techniques used suffer from a deficiency of objective evaluation.

Our work aims to objectively analyze speech production disorders (hypophonia, dysprosody, dysarthria) and evaluate the effects of L-dopa, subthalamic nucleus stimulation and speech therapy on patients. The creation of the AHN corpus has enabled a significant list of publications (Duez, 2006; Duez et al., 2009; Sarr et al., 2009; Viallet et al., 2004, 2002; Viallet and Teston, 2003).

2.4. Salient features of the MTO and AHN corpora

These corpora are useful for research because of the following features:

(1) Corpora include signals, complementary to sound signals, such as SPL intensity, oral airflow, subglottal air pressure (see 2.5).
(2) Recordings of 601 Parkinson's patients in a range of contexts (with/without medication, with/without subthalamic stimulation...), spanning 1616 recording sessions, with 332 of the MTO corpus dysphonic speakers recorded several times (before and after treatment).
(3) Collection of precise information about the speakers (date and place of birth, mother tongue...) and the clinical conditions (date of disease detection, symptoms location, regular treatment and actual treatment while recording, results of clinical exams...).
(4) Range of different tasks produced by the speakers:
   - Sustained vowel /a/ (with oral airflow in order to measure a glottal leakage).
   - Maximum Phonation Time (together with oral airflow to measure the air volume exhaled during this phonation task).
   - syllable repetition ("pa pa pa pa") or ad-hoc sentences ("papa ne m'a pas parlé de beau papa") in order to estimate the sub-glottal pressure by using the airway interrupted method (see 2.5).
   - Glissandos.
   - Repeating sentences (e.g. with/without nasal phonemes to study possible abnormal movements of the velum).
   - Reading texts (normal, fast and slow rate).
   - Describing a scene (semi-spontaneous).
   - Diadochokinesia (e.g. /pataka/...).

These tasks were not executed by all patients.

2.5. Multiparametric recording system used for MTO and AHN corpora

Acoustic and physiological data are recorded with the EVA© workstation (SQLab-LPL, Aix en Provence, France; Teston and Galindo, 1995), which was designed to record and measure sound waves, pitch, SPL intensity, airflow, and pressure, for the evaluation of speech production. Multi-parametric data are recorded using SESANE software (Software Environment for Speech ANalysis and Evaluation – see http://www.sqlab.fr). A range of acoustic, aerodynamic and electrophysiological sensor devices are connected to a PC (Fig. 3). Additional devices, such as electroglottography equipment (Fabre, 1957), can be connected and recorded as auxiliary data.

Oral airflow is measured with a flow meter based on a resistive grid (pneumotachograph principle) with a small dead volume and specific linearization for the inhaled and exhaled airflow (Ghio and Teston, 2004). A soft silicone rubber mask, pressed against the speaker's face prevents air leakage, without hindering articulatory movements.

For voice assessment, the subject is asked to pronounce sustained /a/ vowels, which are analysed using the fundamental frequency (F_0 in Hz), the intensity curve (SPL dB) and oral airflow data collected (Fig. 4a). Several measurements can be computed as proposed by Baken and Orlikoff (2000): mean F_0, mean intensity, mean oral airflow, coefficients of variation (ratio between standard deviation and mean), jitter, shimmer, harmonic to noise ratio. Several graphical outputs (e.g. histograms, long term spectra) allow visual reporting of these parameters (Fig. 4b). For measurement purposes the entire sustained vowel is used after removal of the 250 ms at the beginning of the utterance (to avoid unsteady onset) and the 250 ms at the end (to avoid unsteady release).

Subglottal pressure is estimated with the airway interrupted method (Smitheran and Hixon,1981) using a PVC

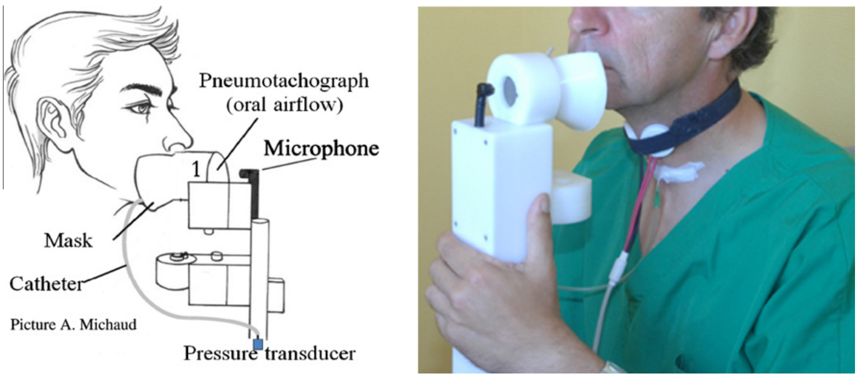


Fig. 3. EVA aerophonometer to record simultaneously the speech, airflows and pressure. Additional devices can be added as EGG (right picture).

Author's personal copy

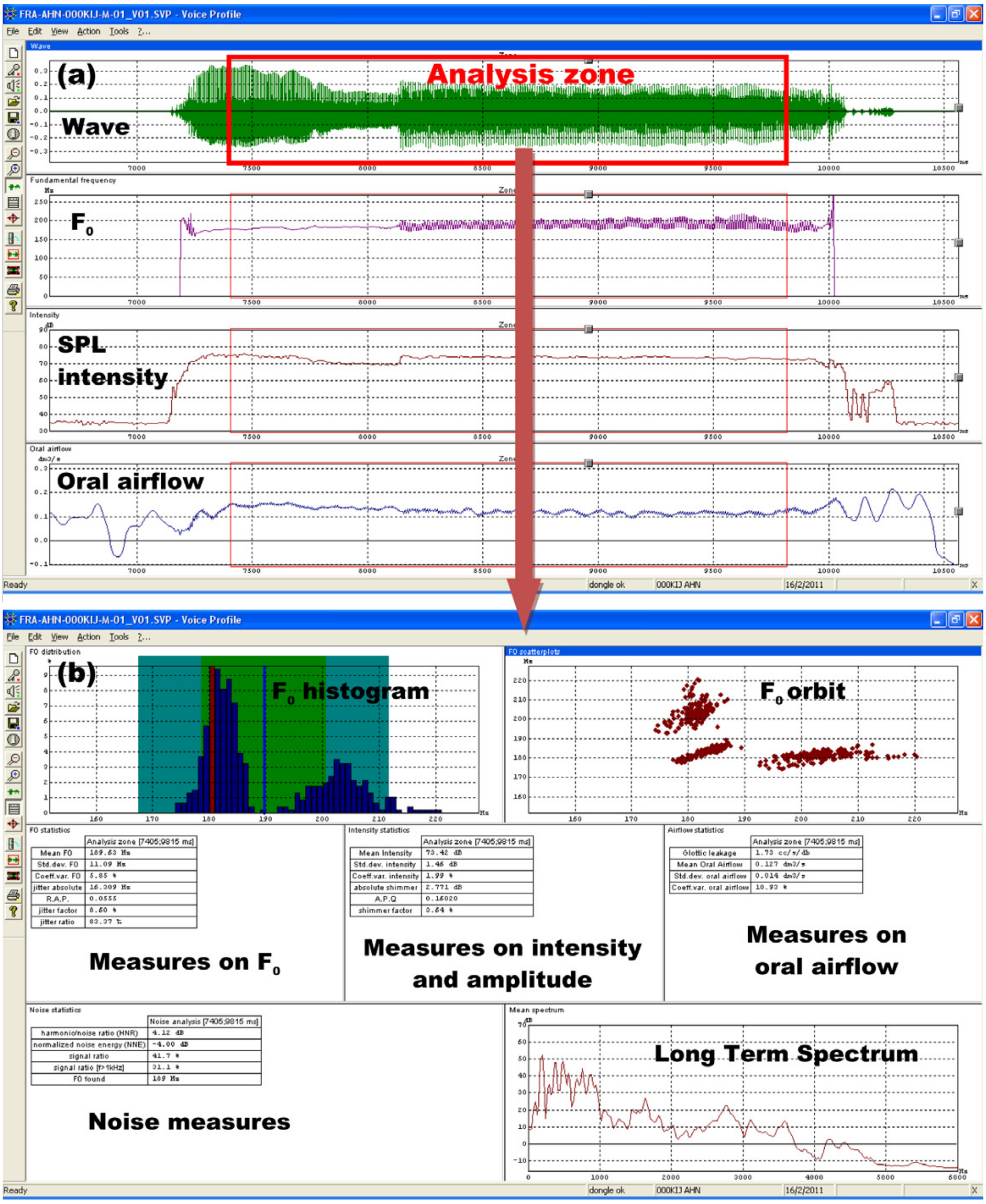


Fig. 4. Multiparametric analysis of sustained vowel with EVA device (SESANE software): example of a bitonal voice. In (a) are displayed wave signal, fundamental frequency (F_0), SPL intensity and Oral Airflow. In (b) are displayed the results obtained in the analysis zone: mean F_0, mean Intensity, mean Oral Airflow, coefficients of variation (ratio between standard deviation and mean), jitter, shimmer…

probe located in the subject's mouth and connected to the pressure sensor device of the workstation (Fig. 3) while the subject is instructed to pronounce consecutive /pa/ or ad-hoc sentences ("papa ne m'a pas parlé de beau papa") at normal pitch, rate and loudness. During the occlusion of /p/, the lips are closed, the glottis is opened, and the vocal tract can be considered as a single air volume from the lungs to the lips: pressure in the oral cavity is the same as in the sub-glottal cavity. Thus, sub-glottal pressure (ESGP) can be estimated using intra oral pressure (Fig. 5). ESGP is useful for voice assessment and relevant in distinguishing normal and dysphonic subjects (Yu et al., 2007; Ketelslagers et al., 2006). It is also a good marker to test L-dopa and STN stimulation effects on pneumophonic coordination in Parkinsonian dysarthria (Sarr et al., 2009).

Finally, the voice range can be assessed by measuring the maximal and minimal fundamental frequency using glissando tasks. Maximum phonation time is also a very robust index and can be obtained by asking the speakers to produce the longest sustained vowel they can.

Author's personal copy

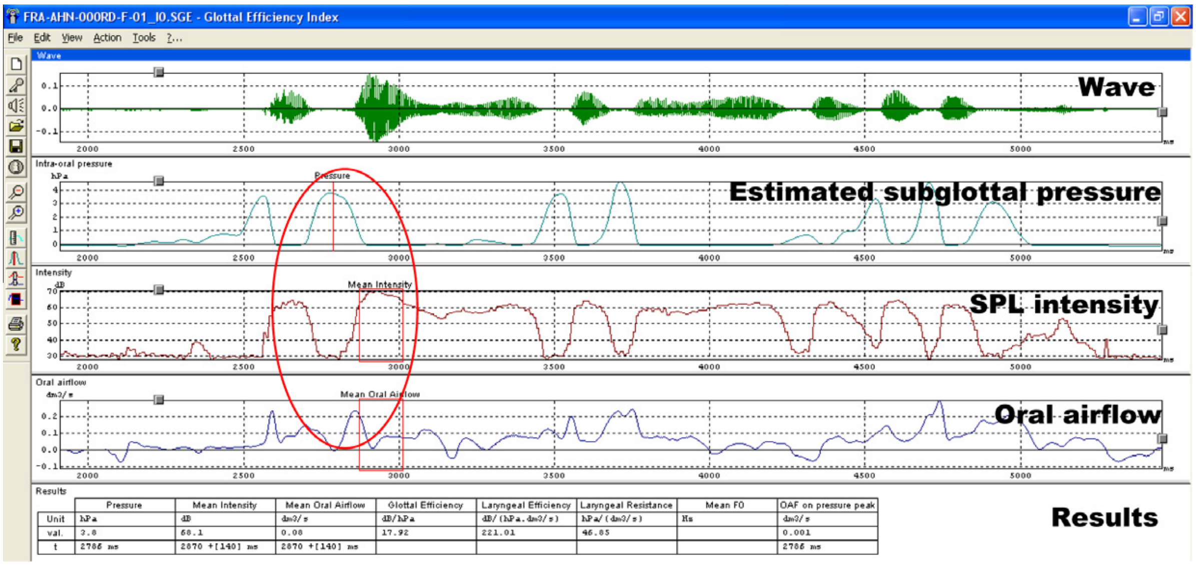


Fig. 5. Estimated Sub-Glottal Pressure for the sentence («papa ne m'a pas parlé de beau papa») From top t bottom are displayed wave signal, estimated sub-glottal pressure, SPL intensity and oral airflow. An array of results provides on the selected /pa/ (circled in the figure) the pressure (ESPG) on /p/, the SPL intensity and oral airflow (OAF) for the consecutive /a/. Several ratio are also proposed to explore the relations between ESGP, OAF and SPL intensity: glottal efficiency = SPL/ESGP; laryngeal efficiency = SPL/(ESGP * OAF) ; laryngeal resistance = ESGP/OAF (Smitheran et al., 1981).

Advances have been made in the quality and capacity of media formats, improved training of staff, and improved experimental design. However normalisation and structuring of data on speakers and their linguistic production remains the weak link in the creation and use of disordered speech corpora. Therefore, careful database design and management are prerequisites for effective exploitation of disordered speech corpora.

3. Database organisation

Even though large sets of sound or physiological corpora are available, analyzing them may be uninformative if the recordings cannot be linked to the speakers' clinical data. This raises the issue of the fundamental differences between data storage and databases.

3.1. Current state of French computerised databases

In France, the first computerised database of spoken French sounds to be described was in Carre et al. (1984). Other databases such as BDSONS (Descout et al., 1986), EUROM (Zeiliger et al., 1992), PSH (Marchal and Meunier, 1993), BDBRUIT (Zeiliger et al., 1994) were also built. They contain isolated utterances, sentences, digits lists, logatoms – essentially for the evaluation and the training of recognition and synthesis systems, and were managed by the now obsolete GERSONS software.

The international EMU Speech Database System (McVeigh and Harrington, 1992; Bombien et al., 2006) is an integrated set of tools for creating, querying and analysing annotated speech corpora. It allows hierarchical and autosegmental annotations and the possibility of providing a query language able to extract annotations and their associated signal files (see http://www.ipds.uni-kiel.de/forschung/emu_splus.de.html). This system is very powerful for linguistic analyses based on sound signals and annotated information. However, this is not sufficient for voice disorders database management, which requires very complex queries. For instance, we may want to query the database for all male Parkinson's patients between 50- and 60-years-old, right-handed, recorded in Marseilles, whose jitter is higher than 1%. It is clear that any database must be constructed to store a wide variety of demographic and clinical speaker information in addition to sound recordings and annotations, and, above all, must have inbuilt flexibility and power to create complex queries of disparate data fields.

3.2. Database concepts

A well designed database (DB) allows different information to be safely and accessibly archived, and permits users to update data and improve data management protocols via a data server. The design of our DB was therefore based on a functional analysis. It was created in a clinical environment, based on empirical corpora, such as those presented in Section 2.

While DB design and management fundamentals are familiar to computer scientists, researchers from other fields may not be *au fait* with these concepts. In fact, it often seems that corpora or collections of data are regarded as databases – pointing to a popular confusion between a collection of data (such as a collection of sound recordings) and a DB. Databases are distinct from simple collections of stored data because in a DB information is organised and structured consistently follow a data model. This allows data to be stored electronically, accessed using precise

Author's personal copy

criteria and shared by a group of people. The cornerstone of DB construction is a database management system (DBMS) which implements the protocols of the data model. For our purposes, the DBMS must (1) make data sharing clear between the different users, (2) protect data confidentiality where necessary, (3) provide a framework for data to queried and (4) make available different interface languages according to the user's profile.

In our case, we opted for a relational model, which we considered the simplest and the most elegant DB model. In a relational DB data are stored in linked tables, generating a minimalist and intuitive data architecture.

Fig. 6a gives an example of the organisation of the data within a conventional flat model with its attendant duplication of the same information in many columns (example on Fig. 6: two diagnosis A and B, three sessions of recordings 1–3). Such a structure can generate extremely large and difficult to manage matrices.

The relational model (Fig. 6b) we propose reduces complexity: one table contains the speaker demographic information, a second one contains the speaker's diagnosis and clinical information, while a third one contains information related to the several recorded productions. Records can be regularly added and updated without modifying the model (the "flat model" version would require the addition of columns: Session 4, 5, 6…).

As summarised in Fig. 7, the DB is composed of about fifty tables which give the speakers' civil (date and place of birth, place of residence…), sociolinguistic information (mother-tongue, professions…), medical data (symptoms, diagnosis, usual treatments), recording sessions (date, location, operator…), recording context (with/without treatment), experimental protocol (task, instructions to the participant, linguistic content, devices used…), associated documents (wave, EGG…) and possibly assessment scores (perceptual, instrumental…). We also propose a system whereby data used in a particular study can be tagged to allow later recovery of large datasets using a simple query.

As information is collected from various sources, there is a need for normalisation, which is achieved using a set of closed lists of identifiers linked to terms related to professions, languages, countries/regions, symptoms, therapies, diagnoses, risk factors, localizing of pathologies, experimental contexts, and evaluation methods. The use of these closed lists avoids the increase of versions of the same term. For instance, a "Parkinson disease" diagnosis can be noted as PD, Parkinson's disease, Parkinson, Park… It is more efficient to propose a closed list where "Parkinson's disease" is assigned as diagnosis n° 11 (a fixed arbitrary value, see Fig. 8). All Parkinsonian patients will then be referred to using this identifier. An important advantage of such a coding is international compatibility. If all the items of the lists are translated, the whole content of the database is operational and adapted to the new language. A list of common diagnoses relative to voice and speech disorders is proposed which can be added to as required.

3.3. Storage of clinical data

As mentioned above, the study of pathological speech specifically requires the collection and storage of precise information – personal as well as medical – relative to the speakers and the medical contexts in which they were recorded. Maximum information should be collected:

(1) *Sociolinguistic information*:
- Gender, date of birth.
- Places of birth and of successive residence.
- Mother-tongue, languages spoken, competence.
- Professional status or level of study.
- Dominant hand.

(a) Flat model (wrong solution)

Speaker Identifier	Birth date	Birthplace	Diagnosis A	Diagnosis B	…	Filename Session 1	Filename Session 2	Filename Session 3	…
164	02/12/1968	Marseille	Nodule			B01957			
165	06/03/1961	Paris	Parkinson	Stroke		A0678	A2524	A2536	

(b) Relational model (best solution)

Speakers Table

Speaker Identifier	Birth date	Birthplace
164	02/12/1968	Marseille
165	06/03/1961	Paris

Diagnosis Table

Diagnosis counter	Speaker Identifier	Diagnosis	Date
645	164	Nodule	2001
646	165	Parkinson	2005
647	165	Stroke	2007

Recordings session Table

Recording session counter	Speaker Identifier	Filename Session	Date
953	164	B01957	24/04/2001
954	165	A0678	10/02/2005
955	165	A2524	12/05/2007
956	165	A2536	20/11/2008

Fig. 6. Data organised as flat model (a) or with a relational structure (b).

Author's personal copy

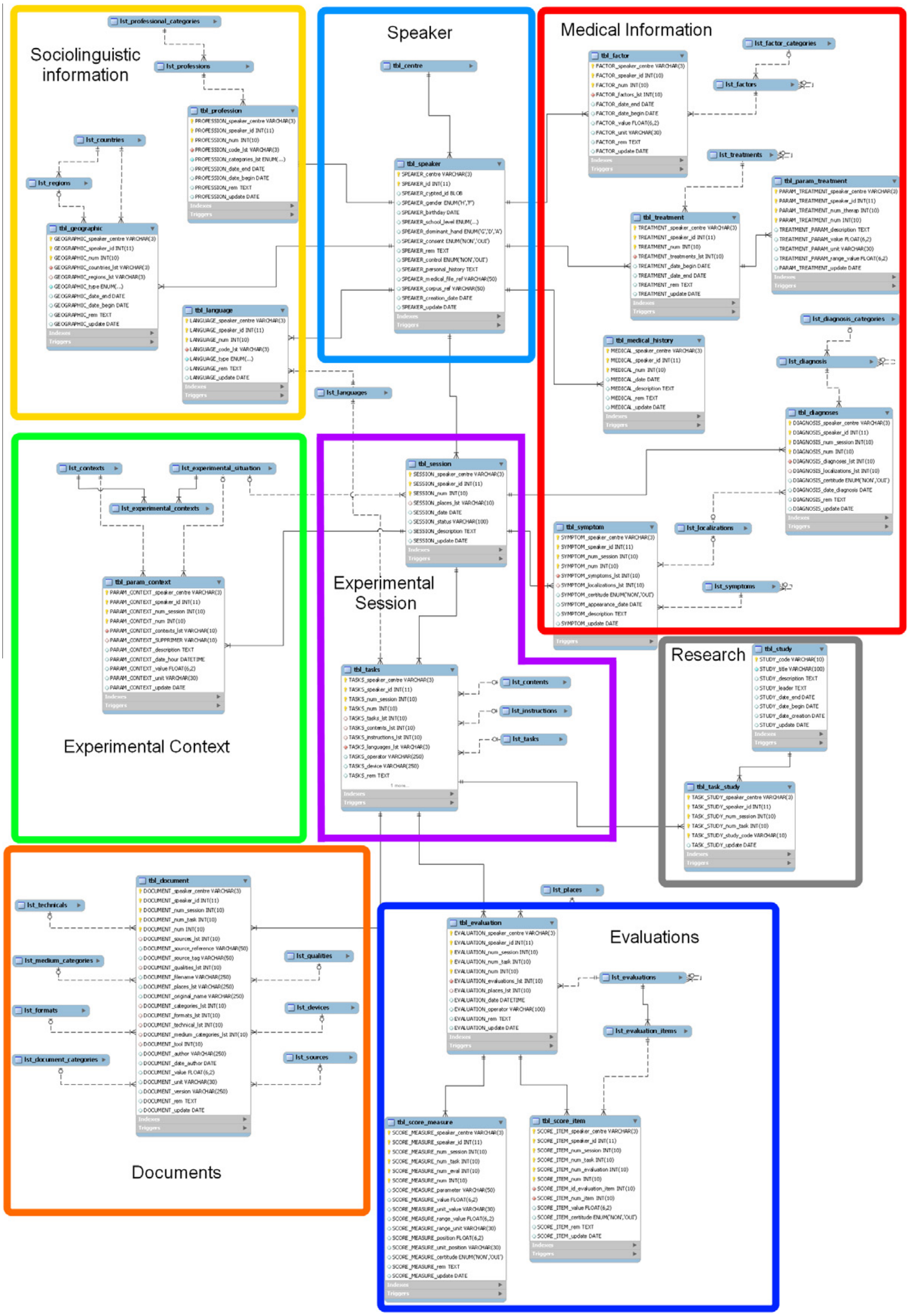


Fig. 7. Conceptual model of a database of pathological speech.

Author's personal copy

Neurological disorder

1. Stroke
2. Cerebral arteriopathy
3. Cerebellar ataxia
4. Degenerative encephalopathy
5. Evolutive encephalopathy
6. Progressive myoclonic epilepsy
7. Alzheimer's disease
8. Amyotrophic lateral sclerosis (ALS)
9. Friedreich's disease
10. Huntington's disease
11. Parkinson's disease
12. Strumpel-Lorrain disease
13. Wilson's disease
14. Lysosomal disease
15. ___Gangliosidose GM1
16. ___Niemann-Pick type A et B
17. ___Niemann-Pick type C
18. Mitochondriopathy
19. Myastheny
20. Myopathy
21. ___Muscular dystrophy - Duchenne
22. ___Muscular dystrophy - Ocular Pharynx
23. ___Steinert's disease
24. Myositis - sporadic inclusion body
25. Multiple Sclerosis Atrophy
26. Cerebellar syndrom
27. ___vascular accident
28. ___genetic degenerative atrophy
29. Guillain Barre Syndrome
30. Little Syndrome
31. Steele-Richardson Syndrome
32. Wallenberg Syndrome
33. Lacunar syndrome
34. Parkinsonian syndromes
35. ___Multiple System Atrophy (MSA)
36. ___Cortico-Basal Degeneration (CBD)
37. ___Dementia with Lewy Bodies (DLB)
38. ___Progressive Supranuclear Palsy (PSP)
39. ___Vascular Parkinson Syndrome
40. Pseudobulbar Syndrome
41. Cranial trauma
42. Essential tremor

.....

99. Other neurological disorder

ENT disorder

101. Amydalectomy
102. Angioma
103. Arytenoidectomy
104. Bioplasty
105. Cleft palate
106. Cord atrophy
107. Cordectomy
108. Cyst
109. Dysfonctionnal dysphonia with normal larynx
110. Epiglotectomy
111. Granuloma
112. Horton disease
113. Laryngeal asymetry
114. Laryngeal cancer
115. Laryngeal ectasia
116. Laryngeal irritation
117. Laryngeal leukoplakia
118. Laryngeal mucosity
119. Laryngeal Papilloma
120. Laryngeal Paralysis
121. ___bilateral
122. ___unilateral
123. Laryngeal sarcoidosis
124. Laryngeal trauma
125. Laryngectomy
126. ___Partial Laryngectomy
127. ______Labayle
128. ______Mayer Piquet
129. ______Tucker
130. ___Total Laryngectomy
131. Laryngitis
132. Laryngocele
133. Larynx biopsy
134. Lesion
135. Nodule
136. Oral cancer
137. Oral surgery
138. Oval glottis
139. Polyp
140. Post intubation injury
141. Post surgery cancer
142. Presbyphony
143. Reinke's Edema
144. Scar
145. Spasmodic Dysphonia
146. Sulcus
147. Surgery of the oropharynx
148. Thickening mucous
149. Thyroïdectomy
150. tracheal stenosis
151. Tumor
152. Vergeture
153. Vocal abuse
154. Vocal essential tremor
155. Vocal particularity linked to transexuality
156. Voice breaking

.....

199. Other ENT disorder

Fig. 8. Example of closed list content (diagnosis).

– General remarks (e.g. difficulty in reading, illiteracy, deafness, short/long sightedness, stuttering, singing experience, physical fitness).

To illustrate the importance of this type of information: we have been confronted in some cases of dysarthria with the phenomenon of /r/ elision which may be similar to the one found in "Creole" accents; only knowledge of the speaker's places of birth and of residence has allowed us to know if this phenomenon was pathological or sociolinguistic.

(2) General medical information:

– medical history (e.g. psychological state, depressive syndrome, hallucination, behavior and/or cognitive disorder, other disorders,...
– treatments received (e.g. surgery, medicine, speech therapy, electrophysiology),
– contributory factors to disease (e.g. alcohol and tobacco, sound and air pollution, respiratory allergy, vocal abuse, stress, intubation).

This information allows patient data to be excluded or included for the purposes of particular research. For example, a study on Parkinsonian's speech can select only non-smoker patients who have not received any treatment and without cognitive disorders.

(3) *Symptomatic information*: the patient's symptoms and the signs observed by the doctor should also be indicated (e.g. dysphonia, dysarthria, tremor, glottal leakage, cognitive disorder, auditory processing

Author's personal copy

disorder), as well as the date when they were seen, possibly giving an indication of certainty and if necessary, the anatomical location (e.g. jaws, right/left-upper limb, right/left-lower limb).

(4) *Pathological information*: the diagnoses given by the doctor (e.g. nodule, polyp, Parkinson's disease, Charcot's disease, brain trauma), the date when they were established, with a possible indication of certainty, and if necessary, their anatomical location (e.g. left/right vocal fold, frontal lobe, parietal lobe) should be given too.

(5) *Contextual information*: as indicated above, the clinical context the patient is recorded in represents an important piece of information to collect in order to run rigorous and significant analyses. Here are some of experimental contexts to collect:

- Pharmacological status (e.g. the date and hour of the last medication, usual medication's nature and amount and medication while the patient is recorded.
- Activated and deactivated neurostimulation state.
- Pre/post-operatory situation (e.g. the operation date).
- Complementary information (e.g. "the patient has bronchitis, wears a brace, had his medicine 4 h ago, forgot his glasses").

(6) *Protocol*: because of the diversity of acoustic features linked to voice and speech disorders, we propose to distinguish on the one hand the tasks of vocal elocution produced by the speakers (e.g. singing, sustained vowel, reading a text, repetition, image description, spontaneous speech) and on the other hand the linguistic content (e.g. /a/ vowel, the days of the week, Rainbow Passage). In addition, it is interesting and relevant to store the instructions given for the different tasks (e.g. fast, slow, usual rate). If the use of a database management system is recommended for the traceability and the exploitation of metadata, the standardization of the protocol for collecting sound or physiological data is hardly compatible with the clinical context. In fact, a complete protocol including the production of sustained vowels, vocal efforts, sentences, repetitions, read texts, spontaneous speech, is hardly feasible because of the complete series of tests the patient is having and the exhaustion caused by too many long efforts. It is therefore better to adapt elocution tasks to the speaker's dysfunction state. For example, a study on nasality is particularly interesting in the case of paralytic dysarthria because of the immobility of the soft palate but is not important in Parkinson's disease for which phonatory exercises can be preferred because of hypophonia.

(7) *Document*: in the "document" table, recording filenames, characteristics (e.g. sampling rate, format, quality for a signal file), as well as the experimenter's name are stored. A document can be a signal file (recording) but can also consist of orthographic transcriptions, annotations or images associated to the task. Regarding the filenames, it is not necessary to code all of the speaker's information, the session, the context, the task while naming a file as this may generate extremely complex names. However, it is necessary to normalize these names and the principle in use here can be demonstrated using the example below FRA-MTO-000052-M-03_L02.wav where the data comes from France (FRA) and is part of the MTO corpus, with a male (M) speaker identified as number 52. The document is the wave file of the third recording session for this speaker, executing the second reading task (L) during the session. A list of the current tasks is available in Fig. 9. The DB must be queried to get further information (not contained in filename) such as context, pathology, age, geographical origin, socioprofessional category, treatments, pharmacological context.

(8) *Evaluations*: perceptual or instrumental evaluations are informative resources that should be stored, as in the following examples:

- for dysphonic patients: Hammarberg Scheme (Hammarberg et al., 1980), Vocal Profile Analysis Scheme (Laver,

Task	Code	Example
Spontaneous Speech	P	
Guided Description	D	Picture « chute dans la boue »
Reading text	L	Chèvre de M. Seguin, Cordonnier
Sentence	S	C'est une affaire intéressante
Word	W	Liste {Bonjour, Femme…}
Non-sense word	O	acha, ibu, issou
Repetition	R	
Automatic Serie	E	Months, days of the week
Diadochokinesy	K	Pataka, badaga
Maximum Phonation Time	T	/a/
Sustained vowel	V	/a/, /i/
Singing	C	Au clair de la lune, glissando
Airway Interrupted Method	I	Papapa, Papa ne m'a pas parlé de beau papa
Breathing	B	Normal, forced, tussometry

Fig. 9. Speech tasks, associated code and examples.

Author's personal copy

1980), GRBAS (Hirano, 1981), Buffalo Voice Profile (Wilson, 1987),

- for dysarthric patients: Frenchay Dysarthria Assessment (Enderby, 1983), BECD (Auzou et al., 2006), UPDRS (Fahn et al., 1987) particularly for patients that suffer from Parkinson's disease...
- neuropsychological and cognitive scales of evaluation such as MMS (Folstein et al., 1975), MATTIS (1988), Wisconsin (Heaton et al., 1993), test on verbal fluency (Cardebat et al., 1990)...

The results obtained from instrumental measurements can also be stored in the DB and used as criteria in the queries to select speakers. For instance, it could be necessary to extract all dysphonic speakers assessed as G1, G2, G3 (light, moderate, severe disorder) according to the GRBAS perceptual scale with a maximum phonation time greater than 15 s.

As proposed by Duez (2006), prosodic evaluation should also be taken into account as well as classifications from automatic methods such as those developed by Wester (1998), Fredouille et al. (2005) or Saenz-Lechon et al. (2006).

3.4. Data format and data description

As mentioned in the introduction, data related to pathological speech corpora may be stored in various different formats. The disappearance of the associated tools and software may have a dramatic effect in terms of long-term availability and accessibility.

First, these various formats may have consequences on the type of associated information available (e.g. clinical state, elocution context) as not all formats enable the storage of such contextual information. For multimedia resources, the RIFF format (Resource Interchanges File Format) may be used. RIFF is not a single file format, but a file architecture principle. Such files are composed of various blocks (called Chunks), some blocks defining technical features (e.g. number of channels, sampling rate), and others listing information (e.g. comments, contents, speakers). RIFF files include formats such as the Wave format (.wav) used for audio. In multi-parametric applications (e.g. simultaneous audio, airflows, EGG), as with the EVA device (cf. 2.5), the data are provided in a particular RIFF format in order to include information on calibration (e.g. values in l/s for the flows, hPa for pressures) or other metadata. Most of the data we have been recording are coded using the RIFF standard. However, experience has proven that this type of coding causes problems for short- and long-term exploitation, and multi-parametric data require transformation before they can be used with analysis software or media players.

Moreover, it is not very likely that this complete but complex coding format will be maintained by standard media players. We think that the best solution remains the one recommended by the SAM projects (Speech Assessment Methodology, Fourcin et al, 1989) where data

```
[SETTINGS]
DATA_TYPE=temporal_serie
DATA_CLASS=Waveform
DATA_FORMAT=int16_intel_data
SAMPLING_RATE=22050
DATA_COUNT=403881
BITS_PER_SAMPLE=16
N_CHANNELS=1
FORMAT_TAG=PCM

[SCALE]
H_UNIT=ms
V_UNIT=
FLOAT_VALUE_AT_MAX=1.000000
SIGNAL_DYNAMIC=32767.000000
SIGNAL_ZERO=0

[SOURCE]
SOURCE=bicf1.wav
SRC_BITSPERSAMPLE=16
SRC_NCHANNELS=1
…
SOFTWARE=CoolEdit v.96 by Syntrillium Soft. Corp.
TECHNICIAN=deniaud
LOCATION=LPL-Schuman, Aix-en-Provence
CREATION_DATE=2003-11-13

[SPEAKER]
CODE=BICF0035
GENDER=F
BIRTH_YEAR=1949
….
[CONTEXT]
THERAPY01=Surgery
….
```

```
[SETTINGS]
DATA_TYPE=temporal_serie
DATA_CLASS=oral airflow
DATA_FORMAT= int16_intel_data
SAMPLING_RATE=6250
DATA_COUNT=64000
BITS_PER_SAMPLE=16

[SCALE]
H_UNIT=ms
V_UNIT=dm3/s
FLOAT_VALUE_AT_MAX=2.000000
SIGNAL_DYNAMIC=32767.000000
SIGNAL_ZERO=0

[SOURCE]
SOURCE= CGMM0000.oaf
SRC_BITSPERSAMPLE=16
SRC_NCHANNELS=16
…
SOFTWARE=EVA Recorder SQLab
TECHNICIAN=ghio
LOCATION=Neuro CH Aix, Aix-en-Provence
CREATION_DATE=1999-10-1

[SPEAKER]
CODE= CGMM0000
GENDER=M
BIRTH_YEAR=1963
….
[CONTEXT]
THERAPY01=L-Dopa ON
THERAPY02=Subthalamic stimulation OFF
….
```

Fig. 10. Description Files associated to raw data: the left column gives a description of audio data; the right column gives a description of oral air flow data and includes information on the absolute calibration of such signal (see [SCALE]). Information on the speaker and on the context are stored in this description in order to perpetuate these meta data even if the database manager is not available.

Author's personal copy

themselves are stored in files in a raw binary format without headers, and where descriptive data (technical specifications) are saved, along with data themselves, in an organized text file (Fig. 10). In addition to technical information, we added clinical information such as the patients' age, gender or nationality, as well as their pathology and treatment. Fig. 10 shows text examples of description files. This format was adopted as the native format of data in the "PHONEDIT Signaix" software[1] for voice and speech analysis developed in "Laboratoire Parole et Langage", Aix-en-Provence. This format of coding appears perennial, as the exploitable data of the SAM corpora of the 90's. We are also investigating the possibilities of coding descriptive data in XML format.

In addition to the specific attention we have to pay about to organisation of pathological speech data, there are technical considerations in the implementation of the system which we describe below.

4. Technical realisations

The pathological speech database was created using a PHP/MySQL development environment, on an Apache Server using SSL (Secure Sockets Layer) for communication encryption. For the purposes of security and access it is important to manage privileges/roles given to users and to encrypt confidential data.

Two distinct applications allow data access:

1. An MS Access user interface which updates the MySQL DB via an ODBC connection; Microsoft Access software is used to easily create simple forms in order to fill the necessary fields associated to the data (Fig. 11).
2. A Web interface developed in PHP/JS allowing complex queries to download sound and physiological data.

Our objective is to deliver the database management system under a free public licence, at the end of the project ANR BLAN08-0125 DESPHO-APADY Fougeron et al. (2010). Part of the data will also be available on the the project website[2] or on the Centre de Ressources pour la Description de l'Oral[3] which offers labs and scholars a free-of-charge service to share and archive their oral data using procedures compliant with the OAIS model for long-term preservation.

It could also be interesting to connect this database management system with other medical applications such as a video workstation or with a hospital information system. HL7[4] (Health Level Seven) provides a framework and standards for the exchange, integration, sharing, and retrieval of electronic health information such as clinical practice, management, delivery, and evaluation. As far as we know, there is no such application for speech sample collections. Moreover, the HL7 specification itself does not define canonical rules that allow a "standard" implementation. While HL7 implies a data model and management rules, it is not based on explicit reference architecture, so compliance testing is not practical.

Finally, there are issues relating to data dissemination and legal matters that must be addressed if databases are to be used for research.

5. Dissemination and legal aspects

When data is collected and stored in a database for dissemination to a wider group of researchers, it is important to consider questions of confidentiality and data protection. A system for formal licensing is needed to ensure that data sources are acknowledged – the corpus should be viewed as the producer's intellectual property; so when exploiting a corpus or publishing results from a corpus or even a part of it, users must quote this reference. In addition, protocols for updating and adding data by new researchers should be established to ensure the integrity of the corpus. The creation of a corpus and a DB to manage it is a significant investment of time and effort, and the question must be asked: What is the benefit to the producer of the corpus? When data is accessed and used in research by other groups it adds value and increases visibility for the producer of the corpus as their data is used more widely in the field. Sharing of data can make the original researchers known to other teams and lead to new collaborations and further research. This highlights the importance of agreements with users on the necessity for data sources to be quoted in any work generated from the data.

These necessary recommendations do not exempt researchers from declaring to the "Commision nationale de l'Informatique et des Libertés (CNIL)" – board which enforces law on data protection – the collecting and the storing of clinical information as well as the statistical analyses carried out from the data during the research. For instance, in the document "Methodology of reference for the processing of personal data involved within the framework of biomedical research" (CNIL, 2006), the authorized data relating to the participant to biomedical research can exclusively concern the following categories:

Identity: number and/or code alphanumeric,[5] other than the complete name and of the PIN of social security;

- *Health: therapy followed within the framework of research and concomitant, results of examinations, events undesirable, previous personal or family, associated diseases or events.*

[1] www.lpl-aix.fr/~lpldev/phonedit/.
[2] http://despho-apady.univ-avignon.fr/.
[3] http://crdo.up.univ-aix.fr.
[4] http://www.hl7.org.

[5] the alphanumeric code can correspond to the first three letters of the name. It is however recommended to only limit itself to initial, i.e. with the first letter of the name and the first letter of the first name since a number is also allotted to the inclusion of the person.

Author's personal copy

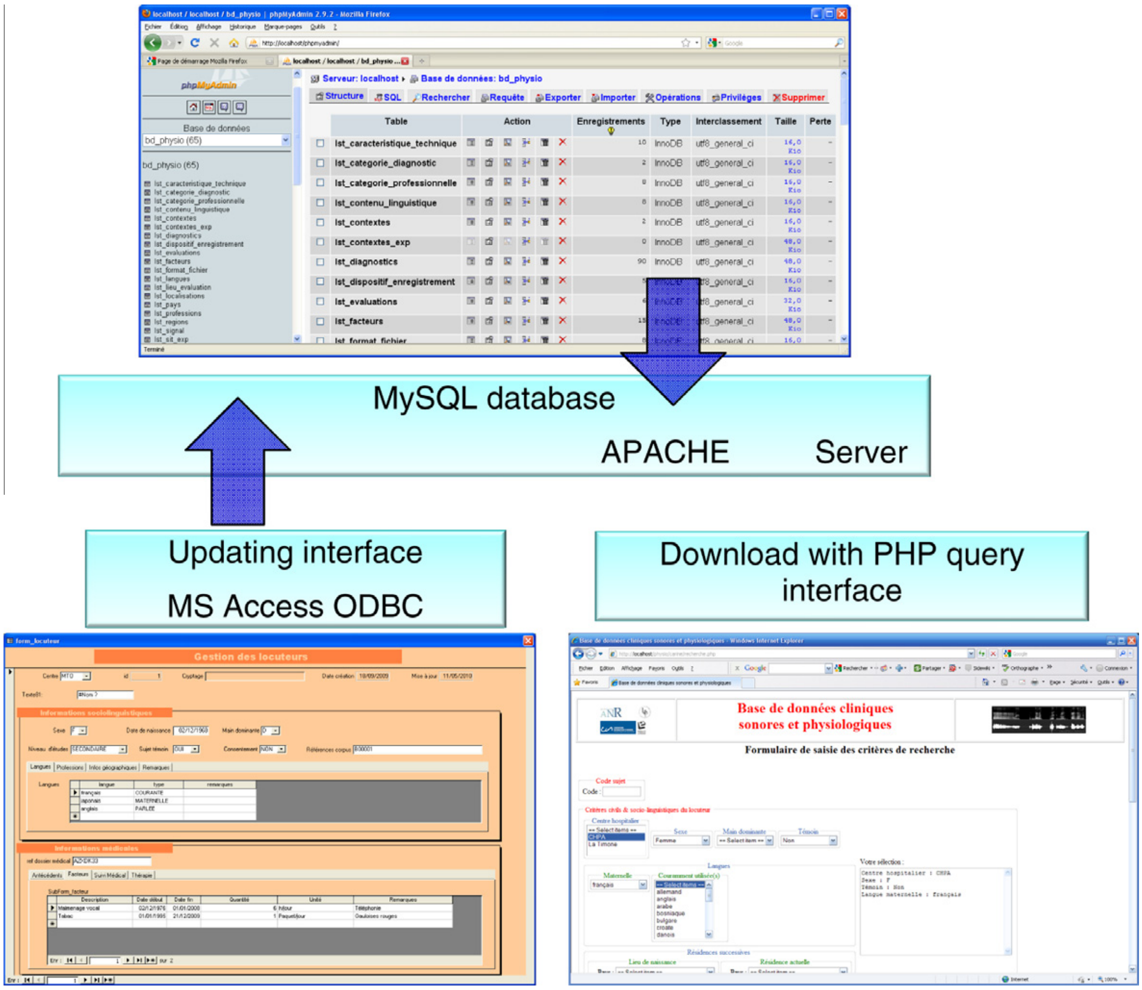


Fig. 11. Technical realisation of "pathological" speech database.

- *Descriptive information: age or date of birth, birthplace, sex, weight, size.*
- *Date of inclusion in research.*
- *Ethnic origin only if it is justified by the aim of the research.*
- *Genetic variations including genetic polymorphisms and/or variations of the form of genes, in relation to the response to a drug or a product (within the framework of research pharmaco-genetics).*
- *Marital status, number of children.*
- *Educational level (primary, secondary, higher).*
- *Social and economic category (categories INSEE).*
- *Economic and financial situation: mode of social protection, existence of a complementary insurance (mutual, private insurance).*
- *Affiliation with a social security system other than the number of social security.*
- *Perceived allowance*s.
- *Participation in other research.*
- *Professional life: current profession, historical, unemployment, professional displacements.*
- *Displacement (towards the place of care: mode, duration, distance).*
- *Consumption of tobacco, alcohol, drugs.*
- *Life habits and behaviors: dependence (single, in institution, autonomous, bedridden), assistance (domestic, family assistance), physical exercise (intensity, frequency, duration), mode and food behavior.*
- *Way of life: urban, semi-urban, wandering, sedentary; habitat (house or apartment, stage, elevator).*
- *Sexual life (only in the cases it is justified by the aim of the research).*
- *Scale of quality of validated life.*

We emphasize that the data we proposed to include in the database follow these recommendations. The most critical point is anonimization. It is a key point since the anonimization of the documents is necessary but a table of correspondence between the speaker code and his/her identity must be maintained safely especially if speakers are regularly recorded. This protected table must be safely available in the recording site to enrich the database with new recordings or information.

Finally, ethical aspects must be considered. Consent forms must be signed by all the speakers in order to inform them about the possible use of their voice in the context of

Author's personal copy

clinical or scientific studies. We must distinguish regular clinical activity from a precise research program. In a research context, the principal investigator must bring it into line with the local ethics committee according to the Helsinki Declaration (2004).

6. Conclusion

Although the state of the art reports significant achievements in understanding the voice production mechanism and in assessing voice quality, there is a continuous need to improve the analysis of healthy and pathological voices. Large scale collection of data is required to take into account the "normal" and "pathological" variability of speech. A structured database of pathological speech represents a milestone in progress towards these goals.

Such a database can provide developers and users of clinical software with reference data to form the basis on which different methods may be compared. Databases have been central to the development of robust automatic speech and speaker recognition devices. A speech disorder database can be to provide a similar stimulus for clinical applications.

Our current objective is to use this database management system to facilitate and extend studies on the perceptual assessment of voice (Revis et al., 2002), multiparametric instrumental assessment of voice (Yu et al., 2007), Automatic Speaker Recognition techniques applied to pathological voice assessment (Fredouille et al., 2009), phonation disorders linked to neurological diseases (Viallet and Teston, 2003; Sarr et al., 2009) and dysprosody (Duez et al.; 2009).

Acknowledgments

The authors would like to thank the financial supports: PHRC (projet hospitalier de recherche clinique), ANR BLAN08-0125 of the French National Research Agency, COST Action 2103 "Advanced Voice Function Assessment" and "France Parkinson Association".

References

Auzou, P., Rolland-Monnoury, V., 2006. Batterie d'Evaluation Clinique de la Dysarthrie. Ortho-édition, Isbergues, France.

Baken, R., Orlikoff, R., 2000. Clinical Measurement of Speech and Voice. Singular Thomson Learning, San Diego, USA.

Bombien, L., Cassidy, S., Harrington, J., John, T., Palethorpe, S., 2006. Recent developments in the Emu speech database system. In: Proceedings of the Australian Speech Science and Technology Conference, Auckland, New Zealand, pp. 313–316.

Bonastre, J., Fredouille, C., Ghio, A., Giovanni, A., Pouchoulin, G., Révis, J., Teston, B., Yu, P., 2007. Complementary approaches for voice disorder assessment. Proc. Interspeech. Antwerp, Belgium, pp. 1194–1197.

Cardebat, D., Doyon, B., Puel, M., Goulet, P., Joanette, Y., 1990. Evocation lexicale formelle et sémantique chez des sujets normaux. Performances et dynamiques de production en fonction du sexe, de l'âge et du niveau d'étude. Acta Neurol Belg. 90 (4), 207–217.

Carre, R., Descout, R., Eskenazi, M., Mariani, J., Rossi, M., 1984. The French language database: defining, planning, and recording a large database. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, San Diego, CA, USA, pp. 324–327.

CNIL, 2006. Méthodologie de référence pour les traitements de données personnelles opérés dans le cadre des recherches biomédicales. Décision d'homologation de la CNIL du 5 janvier 2006.

Darley, FL., Aronson, AE., Brown, JR., 1975. Motor Speech Disorders. W.B. Saunders & Co.

Deller, J.R., Liu, M.S., Ferrier, L., Robichaud, P., 1993. The Whitaker database of dysarthric (cerebral palsy) speech. J. Acoust. Soc. Am. 93 (6), 3516–3518.

Dent, H., Gibbon, F., Hardcastle, B., 1995. The application of electropalatography (EPG) to the remediation of speech disorders in school-aged children and young adults. Eur. J. Disord. Comm. 30 (2), 264–277.

Descout, R., Serignat, J., Cervantes, O., Carre, R., 1986. BDSONS: une base de données des sons du français. In: Proc. 12th Internat. Congress on Acoustic, Toronto, Canada, Vol. A, pp. 4–7.

Duez, D., Legou, T., Viallet, F., 2009. Final lengthening in Parkinsonian French speech: Effects of position in phrase on the duration of CV syllables and speech segments. Clin. Linguist. Phon. 23 (11), 781–793.

Duez, D., 2006. Syllable structure, syllable duration and final lengthening in Parkinsonian French speech. J. Multilingual Comm. Disord. 4 (1), 45–57.

Durand, J., Laks, B., Lyche, C., 2002. Protocole, conventions et directions d'analyse. Phonologie Du Français Contemporain, usages, Variétés Et structures. Bull. PFC 1, 1–92.

Enderby, P., 1983. Frenchay Dysarthria Assessment. Pro-Ed, Austin, Texas.

Fabre, P., 1957. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. Premiers résultats. Bull. 'Acad. Nationale Méd. 141, 66–69.

Fahn, S., Elton, R.L., UPDRS Development Committee, 1987. Unified Parkinson's Disease Rating Scale. In: Fahn, S., Marsden, C.D., Calne, D.M., Lieberman, A. (Eds.), Recent Development in Parkinson's Disease, MacMillan Health Care Information, Florham Park, NJ, pp. 153-163.

Fougeron, C., Crevier-Buchman, L., Fredouille, C., Ghio, A., Meunier, C., Chevrie-Muller, C., Audibert, N., Bonastre, J., Colazo Simon, A., De Looze, C., Duez, D., Gendrot, C., Legou, T., Levèque, N., Pillot-Loiseau, C., Pinto, S., Pouchoulin, G., Robert, D., Vaissière, J., Viallet, F., Vincent, C., 2010. The DesPho-APaDy Project: Developing an acoustic-phonetic characterization of dysarthric speech in French. In: Proc. Language Resources and Evaluation (LREC), Valetta, Malta, pp. 2831–2838.

Fourcin, A., Harland, G., Barry, W., Hazan, N., 1989. Speech Input and Output Assessment: Multilingual Methods and Standards. Ellis Horwood, Chichester.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. «Mini-mental state». A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 12 (3), 189-198.

Fredouille, C., Pouchoulin, G., Bonastre, J., Azzarello, M., Giovanni, A., Ghio, A., 2005. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). In: Proc. Eurospeech, Lisboa, Portugal, pp. 149–152.

Fredouille, C., Pouchoulin, G., Ghio, A., Revis, J., Bonastre, J., Giovanni, A., 2009. Back-and-forth methodology for objective voice quality assessment: from/to expert knowledge to/from automatic classification of dysphonia. EURASIP J. Adv. Signal Process., 1–14.

Ghio, A., Teston, B., 2004. Evaluation of the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. In: Proc. Internat. Conf. on Voice Physiology and Biomechanics, Marseille, France, pp. 55–58.

Gibbon, D., Moore, R., Winski, R., 1997. Handbook of Standards and Resources for Spoken Language Systems. Mouton de Gruyter, Berlin, Germany.

Author's personal copy

Gibbon, F., Crampin, L., Hardcastle, B., Nairn, M., Razzell, R., Harvey, L., Reynolds, B., 1998. CleftNet (Scotland): a network for the treatment of cleft palate speech using EPG. Internat. J. Lang. Comm. Disord. 33 Suppl., 44–49.

Giovanni, A., Ouaknine, M., Triglia, J., 1999a. Determination of largest lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. J. Voice 13 (3), 341–354.

Giovanni, A., Revis, J., Triglia, J.M., 1999b. Objective aerodynamic and acoustic measurement of voice improvement after phonosurgery. Laryngoscope 109 (4), 656–660.

Giovanni, A., Guelfucci, B., Yu, P., Robert, D., Zanaret, M., 2002. Acoustic and aerodynamic measurements of voice production after near-total laryngectomy with epiglottoplasty. Folia Phoniatr. Logop. 54 (6), 304–311.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., Wedin, L., 1980. Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolaryngol. 90 (5–6), 441–451.

Heaton, R.K., Chelune, G., Talley, J., Kay, G., Curtiss, G., 1993. Wisconsin Card Sorting Test. Psychological Assessment Resources, Odessa, Fla.

Helsinki, 2004. World Medical Association General Assembly. In: Declaration of Helsinki. Amemdment, Tokyo.

Hirano, M., 1981. Clinical Examination of Voice. Springer-Verlag, Wien.

Hixon, T., Weismer, G., Hoit, J., 2008. Preclinical Speech Science. Plural Publishing, San Diego, USA.

Ketelslagers, K., Bodt, M.S., Wuyts, F.L., Heyning, P., 2006. Relevance of subglottic pressure in normal and dysphonic subjects. Eur. Arch. Otorhinolaryngol. 264 (5), 519–523.

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T.S., Watkin, K., Frame, S., 2008. Dysarthric speech database for universal access research. Proc. Interspeech '08. Brisbane, Australia, pp. 1741–1744.

Kim, Y., Kent, R.D., Weismer, G., 2011. An acoustic study of the relationships among neurologic disease, dysarthria type and severity of dysarthria. J. Speech Lang. Hear. Res. 54 (2), 417–429.

Klatt, D., 1980. Overview of the Arpa speech understanding project, Part 3: Im pact of the Arpa Sur project. In: Lea, W.A. (Ed.), Trends in Speech Recognition. Prentice Hall, Upper Saddle River, NJ, USA, pp. 49–271.

Laver, J., 1980. The Phonetic Description of Voice Quality. Cambridge University Press, Cambridge; New York.

Marchal, A., Hardcastle, W.J., 1993. ACCOR: instrumentation and database for cross-language study of coarticulation. Lang. Speech 36 (Pt. 2–3), 137–153.

Marchal, A., Meunier, C., 1993. A database of subaquatic and hyperbaric speech: PSH/DISPE. J. Acoust. Soc. Am. 93 (5), 2990–2993.

Mattis, S., 1988. Dementia Rating Scale Professional Manual. Psychological Assessment Resources, Odessa, Fla.

McVeigh, A., Harrington, J., 1992. The mu+ system for speech database analysis. In: Proc. 4th Australian Internat. Conf. on Speech Science and Technology, Brisbane, Australia, pp. 548–553.

Menendez-Pidal, X., Polikoff, J., Peters, S., Leonzio, J., Bunnell, H., 1996. The Nemours database of dysarthric speech. In: Proc. ICSLP '96, Philadelphia, PA, USA, pp. 1962–1965.

Parsa, V., Donald, G.J., 2001. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. J. Speech Hear. Res. 44 (2), 327–339.

Pinto, S., Ghio, A., Teston, B., Viallet, F., 2010. La dysarthrie au cours de la maladie de Parkinson Histoire naturelle de ses composantes: dysphonie, dysprosodie et dysarthrie. Rev. Neurol. 166 (10), 800–810.

Pouchoulin, G., Fredouille, C., Bonastre, J., Ghio, A., Giovanni, A., 2007. Frequency study for the characterization of the dysphonic voices. Proc. Interspeech. Antwerp, Belgique, pp. 1198–1201.

Revis, J., Giovanni, A., Triglia, J.M., 2002. Influence of voice onset on the perceptual analysis of dysphonia. Folia Phoniatr. Logop. 54 (1), 19–25.

Robert, D., Pouget, J., Giovanni, A., Azulay, J.P., Triglia, J.M., 1999. Quantitative voice analysis in the assessment of bulbar involvement in amyotrophic lateral sclerosis. Acta Otolaryngol. 119 (6), 724–731.

Saenz-Lechon, N., Godino-Llorente, J., Osma-Ruiz, V., Gomez-Vilda, P., 2006. Methodological issues in the development of automatic systems for voice pathology detection. Biomed. Signal Process. Contr. 1 (2), 120–128.

Sarr, M.M., Pinto, S., Jankowski, L., Teston, B., Purson, A., Ghio, A., Régis, J., Peragut, J., Viallet, F., 2009. Contribution de la mesure de la pression intra-orale pour la compréhension des troubles de la coordination pneumophonique dans la dysarthrie parkinsonienne. Rev. Neurol. 165 (12), 1055–1061.

Smitheran, J.R., Hixon, T.J., 1981. A clinical method for estimating laryngeal airway resistance during vowel production. J. Speech Hear. Disord. 46 (2), 138–146.

Teston, B., Galindo, B., 1995. A diagnostic and rehabilitation aid workstation for speech and voice pathologies. In: Proc. Eurospeech, Madrid, Spain, pp. 1883–1886.

Viallet, F., Jankowski, L., Purson, A., Teston, B., 2004. Dopa effects on laryngeal dysfunction in Parkinson's disease: an acoustic and aerodynamic study abstract. Mov. Disord. 19(S9), S237.

Viallet, F., Teston, B., Jankowski, L., Purson, A., Peragut, J., Regis, J., Witjas, T., 2002. Effects of pharmacological versus electrophysiological treatments on Parkinsonian dysprosody. Proc. Speech Prosody. Aix-en-Provence, France, pp. 679–682.

Viallet, F., Teston, B., 2003. Objective evaluation of dysphonia: acoustic and aerodynamic methods. In: Proc. Around Dysphasia, Dysarthria, Dysphonia, pp. 53–55.

Wilson, D.K., 1987. Voice Problems of Children. Williams & Wilkins, Baltimore, MD.

Wester, M., 1998. Automatic Classification of Voice Quality: Comparing Regression Models and Hidden Markov Models. In: Proc. VOICEDATA98, Utrecht, pp. 92–97.

Yu, P., Ouaknine, M., Revis, J., Giovanni, A., 2001. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. J. Voice 15 (4), 529–542.

Yu, P., Garrel, R., Nicollas, R., Ouaknine, M., Giovanni, A., 2007. Objective voice analysis in dysphonic patients: new data including nonlinear measurements. Folia Phoniatr. Logop. 59 (1), 20–30.

Yumoto, E., Gould, W.J., Baer, T., 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. J. Acoust. Soc. Am. 71 (6), 1544–1549.

Zeiliger, J., Serignat, J., Autesserre, D., Dolmazon, J., 1992. EUROM1: une base de données «parole» multilingue partie française. In: Proc. Journées d'Étude sur la Parole (JEP), Brussels, Belgium, pp. 213–216.

Zeiliger, J., Serignat, J., Autesserre, D., Meunier, C., 1994. BDBRUIT, une base de données «Parole» de locuteurs soumis à du bruit. In: Proc. Journées d'Étude sur la Parole (JEP), Lannion, France, pp. 287–290.