The *DesPho-APaDy* Project: Developing an acoustic-phonetic characterization of dysarthric speech in French.

C. Fougeron[1], *L. Crevier-Buchman*[1], *C. Fredouille*[2], *A. Ghio*[3], *C. Meunier*[3], *C. Chevrie-Muller*[4], *N. Audibert*[2], *J.-F. Bonastre*[2], *A. Colazo Simon*[1], *C. Delooze*[3], *D. Duez*[3], *C. Gendrot*[1], *T. Legou*[3], *N. Levèque*[1], *C. Pillot-Loiseau*[1], *S. Pinto*[3], *G. Pouchoulin*[2], *D. Robert*[3], *J. Vaissiere*[1], *F. Viallet*[3], *C. Vincent*[1].

[1] Lab. de Phonétique et Phonologie, UMR 7018 CNRS-Paris3/Sorbonne Nouvelle, Paris, France
[2] University of Avignon, CERI/LIA, Avignon, France
[3] Lab. Parole et Langage, UMR 6057 CNRS Aix-Marseille Univ., Aix-en-Provence, France
[4] Lab. MoDyCo, UMR 7114, CNRS- Université Paris 10, Paris, France
cecile.fougeron@univ-paris3.fr, corinne.fredouille@univ-avignon.fr, alain.ghio@lpl-aix.fr

Abstract

This paper presents the rationale, objectives and advances of an on-going project that aims to characterize the acoustic-phonetic features of French dysarthric speech, covering a large amount of patients and three dysarthria types. The two French corpora of dysarthric patients used for the selection of the speech data are described, as well as the design of a multi-field query computer interface built to manage recordings and associated clinical meta-data. Advances of the project related to the selection of the population, the pre-processing of the speech files, their transcription and their automatic alignment are also presented.

Index Terms: dysarthria, speech disorders, database, phonetic and acoustic description.

1. Introduction

Dysarthria refers to neurologically-based speech disturbances. It results from damage to the central and/or peripheral nervous system that impairs the transmission of neural messages to the muscles involved in speech production. Dysarthria is therefore the expression of a deficit in the motor execution of speech movements, and thus a motoric speech disorder. Strength, speed, range, rigidity, coordination and precision of speech gestures can be altered at any level of the speech production system (respiratory, phonatory, supralaryngeal).

Dysarthria is one of the most frequent disorders of verbal communication associated with damage of the nervous system. Indeed, it can appear in the clinical profile of a large number of neurological disorders, including cerebellar diseases, stroke, Parkinson's disease, Amyotrophic Lateral Sclerosis (ALS), multiple sclerosis, cerebral palsy, and traumatic brain injury (see e.g. Duffy, 1995; McNeil, 1997; Peacher, 1950).

The clinical manifestation of dysarthria and the characteristics of the patients' speech depend on its cause and the disease associated with it. Therefore, a classification of dysarthria as a unitary condition is inaccurate, and dysarthria has rather to be considered as a label for a group of disorders (Peacher, 1950; Grewel, 1957; Darley et al., 1969a). Several classification schemes have been proposed in the literature to characterize different groups of dysarthrias. They are either based on salient auditory-perceptual features (phonatory, articulatory, prosodic…) that are used to characterize specific articulatory or kinematic behaviors (e.g. ataxic, hypokinetic dysarthrias - Darley et al., 1969a; Darley et al., 1969b; Darley et al., 1975) or based on etiological and/or neuroanatomical criteria (localization of lesion site) (see Grewel, 1957; Auzou et al., 2007 for a review). Although the main features that differentiate 'typical' patients affected by different dysarthria types have been identified, the study of dysarthrias needs more comprehensive phonetic descriptions to overcome the great diversity observed in patients' speech patterns.

In the following section, we will present the rationale and the main objectives of our on-going research project on the acoustic-phonetic characteristics of the speech of dysarthric French patients. Section 3 describes two dysarthric speech corpora (with a focus on the Claude Chevrie-Muller corpus) and the design of a multi-field query computer interface developed to facilitate the management and storage of the recordings. Section 4 presents the advances of the project with a description of the selection procedure of the patients to be analyzed, and the method developed for the pre-processing of the speech files. Finally, section 5 concludes this paper by discussing some theoretical issues related to this long-term project.

2. Rationale and Objectives of the Project: Characterizing Dysarthric Speech

2.1. Challenges

One major challenge to overcome when trying to characterize dysarthric speech is that dysarthrias are complex disorders. All dysarthrias stem from defined neuropathological conditions with a deficit in the spatio-temporal execution of speech movements. However, muscular weakness, spasticity, coordination disorder, involuntary movements, or altered muscle tonus will have varied consequences on the articulatory movements (articulatory target undershoot or overshoot, reduced control of movement amplitude and speed or over time,

uncoordinated speech gestures…). Moreover, all dysarthrias involve disturbances, at some varying degrees, affecting different levels of speech production: respiratory, laryngeal, velopharyngeal (resonance), and articulatory (oro-facial) (Auzou et al., 2007; Kent et al., 1998). Thus dysarthria not only refers to a deficit in articulation per se, but encompasses disturbances in the control of voice quality, speech rhythm, loudness, segmental articulation, pitch, fluency, etc.

A second challenge stems from the vast amount of inter- and intra-speaker variability. As mentioned above, different types of dysarthria sharing common features have to be considered. While these types can be defined by shared features (reduced pitch modulation, speech rate perturbation, impaired coordination, nasal resonance…), they are not well defined by a distinctive and exclusive set of features. Individual speaker idiosyncrasies, differences in the severity of the disease, speaker-specific impairments and compensatory strategies are among the different sources of variability that have to be taken into account.

Given these challenges, the search for relevant and stable criteria in order to describe dysarthric speech patterns needs to include multiple deviant speech dimensions, both at the segmental and the suprasegmental levels, and to be applied to a large population of patients for intra- and inter-group comparison as well as longitudinal observation.

2.2. Limitations

Even though associations between deviant acoustic-phonetic dimensions and certain types of dysarthria have been made in clinical practice and in the clinical literature, descriptions of dysarthria are often based on perceptual assessments as done in the precursory studies of Darley et al. (1969; 1975). It is true that perceptual analysis is still considered as the "gold standard" and a patient is declared dysarthric because he is perceived dysarthric (Duffy, 2005). However, instrumental analysis is more and more recommended to provide complementary information for the assessment and to objectively quantify descriptions of the speech patterns (Collins, 1984 ; Kent et al., 1999). A review of acoustic studies of dysarthric speech is available in (Kent et al., 1999). It reports that "the great majority (of studies) focuses on a small set of measures and typically a very small number of subjects". We can add that most studies focus on a single subsystem (laryngeal, velopharyngeal, labial articulation…) and are based on ad hoc task of speech production (sustained vowel, isolated sentences, diadochokinesis…). In the review done by Murdoch et al. (1998) of 17 acoustic studies, most studies were based on word and sentence reading, one study looked at read texts, and only two studies used spontaneous speech. Acoustic analysis of continuous speech is thus scarce except in the case of prosodic studies as in (Schlenk et al., 1993; Viallet et al., 2002; Mori et al., 2004; Duez, 2006). Finally, very few comparisons between existing studies have been made, and there is no overall characterization of dysarthric speech patterns. This lack of a comprehensive phonetic description of dysarthric speech patterns can be partly explained by the following facts:

(a) Dysarthric speech can be very impaired and information in the speech signal is difficult to obtain and analyze. Consequently, studies are often restricted to a limited set of acoustic measures, and attention is usually focused on a few specific impaired aspects of the speech production system. Since all studies have not been concentrated on the same acoustic cues and on the same patient population, comparisons are rare. As a further consequence, studies are usually restricted to small cohorts of dysarthric speakers and limited to a small variety of speech material.

(b) The absence of a comprehensive picture of dysarthric speech features can also be explained by the fact that the majority of studies is limited to the analysis of one type of dysarthria, or the comparison of at most two types of dysarthria. Although the acoustic features of the major types of dysarthria have been fairly well documented, most of the acoustic studies have focused on dysarthrias associated with Parkinson's disease or ALS.

Furthermore, these studies cover a restricted language area: while significant progress has been made on the description of English dysarthric patients, fewer studies were carried out on French dysarthric speakers (though see Monfrais-Pfauwadel, 1995; Robert et al., 1999; Baudelle et al., 2003; Gentil et al., 2003; Viallet et al., 2004; Pinto, 2005; Duez, 2006).

Finally, different studies have been reported in the literature, based on automatic methods drawing upon the automatic speech processing. Devoted to speech disorders (Gu, 2005; Maier, 2007; Su, 2008; Middag, 2009), the large majority of these methods aims to provide objective measurement of the speech quality in order to cope with the well-known drawbacks of the perceptual assessment like for instance the subjectivity. Based on objective assessment, they do not concentrate their efforts on the characterization of the dysarthric speech by the help of the automatic approaches as proposed in a very few studies like (Teston & al., 1995; Vijayalakshmi et al., 2006).

2.3. Characterizing Dysarthric Speech: Objectives of the Project

2.3.1. A Comprehensive Acoustic-Phonetic Description of Dysarthric Speech

The main objective of this project is to provide a systematic, quantified acoustic description of the speech patterns of French dysarthric speakers. Three major types of dysarthria are examined and a relatively large cohort of patients is included in each type (see 4).

A standardized procedure for the acoustic-phonetic characterization of a patient's production is proposed. The originality of our approach comes from the

combination of methods and analysis procedures drawing upon both phonetics and speech engineering. Thus, the procedure will involve both manual analysis (by human experts) of the acoustic phonetic properties of the productions and automatic acoustic analysis of speech signals. A continuous back and forth between these two techniques should gain from the potential of both approaches.

A large set of acoustic-phonetic dimensions will be investigated to capture the scope of acoustic variations associated with dysarthria and to identify relevant, reliable and robust criteria to characterize patients' speech. Spectral and temporal cues, segmental and suprasegmental criteria, infra and supraglottal dimensions, will be examined via a set of pre-defined measurements that will be used to screen all the selected patients. The relevance of the criteria will be evaluated with respect to their ability to:

- differentiate dysarthric productions from non-dysarthric ones;
- distinguish different (sub-)types of dysarthric speakers;
- monitor the evolution of dysarthria in a longitudinal perspective.

The feasibility and the originality of this project emerge from the collaboration of a team of researchers, specialists of speech but with complementary expertise in phonetics, clinical practice and speech engineering. These partners are located in Paris (Laboratoire de Phonétique et Phonologie - LPP), Aix-en-Provence (Laboratoire Parole et Langage - LPL), and Avignon (Laboratoire Informatique Avignon - LIA).

2.3.2. Development of a Multiple-Field Query Database of Dysarthric Speech

Research on disordered speech is confronted with the difficulty of getting appropriate and sufficiently large quantities of speech data, homogeneous in quality, and sufficiently documented by clinical information on the patients (diagnosis, medical follow-up, medication, symptoms…). Therefore, the second aim of this project (and a preliminary step for our acoustic description) is to design and create a computer database where digitized dysarthric speech corpora and associated patients' clinical information, can be stored, organized and made accessible (for selected and protected usage) through multiple-field queries. The development of this database is motivated by the fact that dysarthric speech recordings are currently disseminated in different locations in France, in different formats, and often without required indexing or clinical documentation. Consequently, their access and handling are difficult, despite the strong demand to use them. Moreover, the development of this database is also motivated by the need to preserve a large speech corpus of French dysarthric speakers recorded from 1967, the CCM database (see section 3.1.1), that must be saved.

While this computer database is designed to manage any clinical content related to speech and voice disorders, it will be firstly designed with the corpora involved in this project and described in section 3.

3. Corpora of French Dysarthric Patients

In the context of our project, the two corpora described below provide us with a large sample of speech data from French dysarthric speakers that can be used for comparisons between speakers, between groups of speakers and in some cases for longitudinal evaluations.

3.1. The CCM Corpus:

Over the past 30 years, Dr Claude Chevrie-Muller (henceforth CCM) with her team recorded at the 'Laboratoire d'étude de la voix et de la parole' (INSERM U3) the patients that were sent to her by different neurologists for the assessment of disordered speech and its relation with neurological pathology. This extensive work has given birth to a unique, highly valuable historical corpus of neurological speech disorders in French, known as "Pathologie de la voix et de la parole en neurologie" or "CCM corpus".

This corpus contains about 1000 hours of disordered speech, produced by 5000 patients (adults and children) approximately, mainly suffering from dysphonia and dysarthria, but also anarthria, aphasia, stuttering, psychiatric disorders and so on. In the population of adult dysarthric speakers, 860 patients were classified according to their neurological diagnosis. Four main types of dysarthria are represented. They include three main groups of neurological syndromes and a group of mixed symptoms:

(1) Disorders related to an impairment of the extrapyramidal system. These disorders are characterized by a modification of initiation and offset of muscle tonus control with rigidity, hypokinesia and hypertonia. This group is represented by Parkinson's disease and related Parkinson's syndromes as well as Choreic disorders.

(2) Disorders related to an impairment of the pyramidal system (principal motor tract) and responsible for paralytic dysarthria. These can be associated with a pseudo bulbar syndrome with a bilateral spastic component or a bulbar syndrome such as in Amyothrophic Lateral Sclerosis (ALS).

(3) Disorders related to an impairment of the cerebellar system which is characterized by an alteration of the ongoing temporal-spatial control of the movement. These can be seen in diseases such as Multiple Sclerosis, Ataxia, Friedreich disease.

(4) A group of mixed dysarthrias related to more diffuse pathologies such as vascular disease, brain injury, etc.

A large variety of speech materials is available in this corpus as listed in Table I. Over the past few years, the protocol has evolved and for the oldest recordings some speech tasks were not recorded: all the items marked with a '*' in table I are present in all recordings, and it is only

after 1980 that the other items were included in the protocol. The production of the whole protocol lasts about 15 minutes per patient.
All the recordings were done in a sound booth with a table-top microphone. Audio and electroglottographic signals were recorded on the two channels of Revox tapes, with indexing in a notebook. Each recording has been analyzed by the CCM team according to specific perceptual and acoustic features. For example, speech rate, word length compared to normative data, segmental description (vowel and consonant realization) and other prosodic variations were reported in the final assessment as well as the oro-pharyngo-laryngeal and praxis clinical examination. The CCM corpus thus contains three types of material stored on different media:

- personal patient information (civil status, tape number, number of recordings—some patients being recorded 4-5 times for longitudinal analysis) and final assessment of the patient's recording were stored as hard-copy;
- medical follow-up (diagnosis, treatments, surgery reports) was stored in patient's charts that consist of typewritten files, letters and reports;
- audio and electroglottographic (EGG) recordings were stored on Revox tapes.

Recordings, notebooks and patient's charts containing all available clinical information are now stored in the Voice and Speech medical lab associated with the Laboratoire de Phonétique et Phonologie (Paris).

Furthermore, a control population of 80 healthy male and female speakers was recorded with the same protocol. In order to continue this activity, Dr L. Crevier-Buchman and her colleagues still record the neurological patients coming to the Voice and Speech Lab of the European Hospital Georges Pompidou (Paris). Recordings are now made on DAT tapes, following the same protocol but with a head mounted microphone to avoid variability in intensity due to patients' movements. EGG is no longer recorded simultaneously.

It is worth noting that there is a huge loss of data in the CCM corpus. Because of the large inter-patient variability, there is a need in updating clinical information about the speaker (score on international scales, precise treatment information, medical states – with/without medication, stimulation, etc). In fact, our experience shows that these requirements are exceptionally satisfied in a retrospective study especially when using old data. It is the reason why we have decided to complete our database with other sources of corpus.

3.2. The Aix-Neurology-Hospital corpus (ANH)

For the past fifteen years, under the impulse of F. Viallet, the department of neurology of Aix-en-Provence Hospital has recorded dysarthric speakers regularly. These patients are recorded with the EVA workstation (Teston et al., 1999) and clinical data are recorded simultaneously on a spreadsheet. Currently, the Aix-Neurology-Hospital (ANH) corpus contains 990 patients (average age = 67,7) and 160 control speakers (average age = 62) with sound, aerodynamic recordings and clinical data (diagnosis, regular and contextual medication, clinical motor evaluation…). The population of patients is mainly composed of Parkinson's disease (601) and Parkinsonian syndromes (98).

The benefit of this corpus is :

(1) the recording of physical (SPL intensity) and physiological signals (oral airflow, estimated sub-glottal pressure, nasal airflow) in addition to of the sound signal (Teston et al., 1999).
(2) the multiple speech tasks : sustained vowels, maximal phonation time, airway interrupted sentences to estimate sub glottal pressure, special sentences to estimate velar leakage, text reading with several speed instructions, spontaneous description of a picture, diadochokinesis and so on. The recorded tasks can vary from a patient to another. For example, estimated sub glottal pressure is now systematically recorded in Parkinsonian hypophonia (Sarr, 2009). On the other hand, velar leakage is mainly recorded for paralytic dysarthria as proposed by Robert et al. (1995).
(3) the multiple clinical contexts of the recording sessions : 601 Parkinson patients recorded with/ without dopa, with/without deep brain stimulator which represents 1616 sessions of recordings;
(4) the collection of a comprehensive set of information on the speaker (date and birthplace, mother tongue, profession…), and the clinical conditions (date of appearance of the disease, localization of the symptoms, medicament dosage, characteristics of possible electro physiological stimulator, scores of the clinical examinations like UPDRS…). Such a precision is necessary for clinical studies (ex: effect of the therapies on the speech production) but also at the linguistic level (search for phonetic-acoustic characterization of homogeneous group of dysarthric speakers).

All the data and information are computerized. This is our main source of Parkinson patients.

4. Advances of the Project

4.1. Getting the audio files

The recordings of the CCM corpus are still on an analog medium (Revox tapes) and, to ensure their safeguarding, need to be urgently digitized. This task is very time consuming. First, each Revox tape contains several patients, and it appeared that digitizing a whole tape at once was a quicker solution than searching for a specific dysarthric patient and digitizing it. Second, during each recording session the speed of the tape was changed according to the speech task (and the need to record EGG). Thus, "real-time" auditory control of the recording has to be done in order to stop the tape at each

change of speed and set the playing speed accordingly. Third, many tapes are in bad conditions, several recordings are of bad quality (mainly due to speaker movements relative to the table top microphone). Thus adjustments have to be made in order to ensure reasonable audio quality in the output files. To date, 180 patients have been digitized.

94 additional patients recorded on DAT tapes were also digitally captured as wav files. Then, all these recordings were segmented per patient and per speech task. The same procedure is applied to the control population. Then the files are renamed for anonymous storage.

In order to get a sufficient amount of speech to be analyzed acoustically, we have chosen to work first on the text reading speech task. It allows to have more than 1 minute of speech, identical for all patients and with segmental, prosodic and fluency variations as well as information on temporal features such as pauses, group phrasing and reading speed through out the text.

4.2. Design of a Database and Multi-Field Query Interface

As mentioned in 2.3.2, the main interest in pooling and organizing clinical resources is to make this information durable, and to allow some exchange and increasing enrichment via an accessible and shared computerized platform.

If the concepts around the databases (DB) are familiar for computer scientists, it can be very different for the non-specialists. It is common to find that a collection of audio recordings or data compose a database. However, a database differs from a collection of recordings/data by a consistent structure and organization based on a model, shareable by a group of people and stored on a numerical support, allowing data selection according to precise criteria. In the literature, these aspects are brought by a DataBase Management System (DBMS), which is responsible for (a) supporting the concepts defined by the data model, (b) ensuring the respect of the consistency rules related to the data, (c) making the sharing of data between several users transparent while ensuring the confidentiality of some parts of the data, (d) replying user queries with a high performance level, and finally, (e) providing different data access languages according to the user profiles.

In this project, a working group has been dedicated to this data structuring task in order to be able to provide users (clinicians, therapists, speech scientists) with a straightforward multi-field query interface capable of responding to their data access needs. It is worth noting that data include here audio and articulatory recordings but also all the information related to them. This information includes patients' information, such as personal and clinical data (diagnosis, medical follow-up, medication, symptoms…), recording protocol information (type of speech, number of sessions, medication state of the patient, ...), material used for the recordings, etc. All this information is necessary for a controlled analysis of the speech data. Before designing this multi-field query interface, this working group chose a relational model to structure data, considered as the most simple and refined models for databases. Its simplicity stems from its tabular but efficient organization, which allows to define a set of objects, their attributes (characteristics) and the relations between objects. This results in an intuitive architecture, efficient in terms of computation access and storage, easily understandable by non-specialists. In this context, a functional analysis has been carried out in order to define a set of objects, attributes and relations related to the clinical environment. This analysis was refined afterward by confronting the relational data model with empirical and "real" clinical data issued from the disorder speech corpora described in section 3.

Finally, the working group is now designing and developing the multi-field query interface, necessary for the data access. This interface is composed of 3 blocks to enter the criteria of the query:

(1) Basic sociolinguistic information (gender, languages, birthplace, address restricted to region);

(2) Clinical information: diagnostics, symptoms, risk factors, therapies;

(3) Recording session information as :

a) the age of the speaker at the recording time,

b) clinical context (ex: ON, OFF, pre-op, post-op…)

c) available assessments (ex: UPDRS, GRBAS, EVA, …)

d) speech tasks (reading, sustained vowels, diadochokinesis,…)

e) linguistic content *([reading]* "La chèvre de M. Seguin", *[sustained vowels]* /a/, *[diadocho-kinesis]* pataka …)

f) studies : the data used by a specific study (ex : ANR, JEP2010, a250, master 2010 Weisz…)

If the query is validated, a tabulated text file is provided including all information chosen by the user. This information can be different from the one used to select the data. For instance, it may be interesting to know the profession of the speaker without being a query criterion. In a second time, the user can refine the selection in excel spreadsheet for instance and can select Parkinson's disease speakers without Deep Brain Stimulation and recorded more than 12 hours of L-dopa withdrawal. When this local selection is done, the user provides a list of target data which are distributed by a secured automaton. For the meantime, as a matter of confidentiality, these operations are not available through network.

4.3. Selection of Patients for the Acoustic Study

In order to include a sufficient number of patients and dysarthria types in our prospective acoustic study, we focused on neurophysiologic alterations of the three main neurological systems: the extrapyramidal system represented by Parkinsonian dysarthria, the cerebellar system represented by ataxic dysarthria and the pyramidal system represented by ALS dysarthria.

For each of these three types of dysarthria, the selection was based on i) the clinical file and information on the disease, the certainty of the diagnosis, the ongoing treatment, ii) the severity of the dysarthria (we are only working on moderate dysarthrias with relatively intelligible speech.). The selection includes:

30 patients with ASL

30 patients with a pure cerebellar alteration

30 patients with Parkinson's disease selected in the ANH corpus. All were out of L-dopa since 12 hours, 15 read the text of the AHN protocol ('La chèvre') and 15 read both the text of the AHN protocol and that of the CCM protocol ('Tic tac').

The recordings of these selected patients are being evaluated perceptually by 3 expert judges. Voice quality, articulation, prosody, intelligibility, naturalness of speech, and severity are rated on a perceptual scale.

4.4. Pre-Processing of the Audio Files

In order to be able to perform the manual and automatic acoustic analysis on the selection of patients described in the previous section, a pre-processing of the audio files is considered as necessary, relying on an automatic text-constrained phonetic alignment. This pre-processing is based on different resources (see below) including an orthographic transcription of the speech production to analyze. Due to the specific nature of the audio files and the quality level of the phonetic alignment expected for the acoustic analysis, individual orthographic transcriptions of each audio file are necessary as they will enable to take into account the possible divergences of speech production (due to difficulties for the patient to speak, disfluencies, …) compared to the expected ones related to the reading tasks (i.e. the texts of "La chèvre" and "Tic tac").

4.4.1. Orthographic transcriptions

Each audio file was listened to and manually transcribed following a set of common transcription rules, especially designed for this clinical context. These rules tend to provide a compromise between the quality level of the phonetic alignment expected and the speech disorders due to dysarthria. The following list provides the main rules defined in this context:

- Rule 1: is considered as a **deletion** the lack of an entire word or one or more syllables (e.g. : the lack of phoneme [R] in the word "pauv**r**e" will not be considered as a deletion);
- Rule 2: is considered as a **substitution** the replacement of at least three successive phonemes by another sequence of phonemes in a word or the replacement of an entire word.
- Rule 3: is considered as an **insertion** all addition of segments of at least one syllable compared to the original text (e.g.: repetition of an entire word or of syllable(s) in the word, hesitations and filled pauses);
- Rule 4: all the speech produced by another speaker (speech therapists for instance) during the recording is transcribed but annotated as some external productions. The same rule is applied for external noise.

Rules 2, 3 and 4 denoting some divergences between the speech production and the expected text to read, the SAMPA alphabet was used to provide a phonetic transcription of added phoneme sequences. Specific tags are added in the transcription to signal these different cases (e.g. for a substitution : *[su=expected_word] pronounced_word_in_sampa [su]*). Finally, a notebook with other remarks about each audio file was also elaborated for the orthographic transcription.

4.4.2. Automatic Text-Constrained Alignment

A text-constrained alignment provides the phoneme time-boundaries of a sequence of words expected in a speech signal[1]. When this alignment is performed by a machine, the automatic system requires as input resources both an orthographic transcription related to the speech production and a text-restricted lexicon of expected words associated with their phonological variants.

Here, the phonetic alignment is performed by an automatic system developed at the LIA laboratory. This system is based on a Viterbi decoding algorithm coupled with a set of 38 French phonemes (in addition to the input resources reported above). Each phoneme model relies on a three state HMM, initially trained on French speech corpora, produced by non-dysarthric speakers. Since the latter has no connection with the dysarthric corpora, classical unsupervised adaptation techniques are applied iteratively on phoneme models for the automatic phonetic alignment to enhance and refine phoneme boundaries.

To deal with the individual orthographic transcriptions (and potential divergences in terms of words pronounced) and the different rules (notably the substitutions and deletions), it is worth noting that the text-restricted lexicon used by the automatic alignment system is dynamically updated for each audio file in order to take new entries (SAMPA-based words or phoneme sequences) pronounced by the speaker into account.

4.4.3. Quality of the Automatic Phonetic Alignment

A subset of productions was selected for a first evaluation of the automatic phonetic alignment. The subset is gender-balanced and includes different degrees of dysarthria severity (2 control speakers, 2 speakers with moderate dysarthria and 2 with severe dysarthria). The automatic alignment of the productions was compared to a manual correction of phonetic labels and boundaries performed by 2 phoneticians. For a given phoneme segmented manually and automatically , the comparison is based on the time shift between the midpoints of the two segments. As defined in (Adda et al.; 2008), the agreement between the automatic and manual alignments

1 as opposed to a non text-constrained alignment, which has to determine the sequences of phonemes as well as their boundaries.

is defined according to a minimum time lag threshold set at 20 ms[2] .

The comparison of the alignments showed a shift above 20 ms for 17% of segments for the control speakers, 24% for the moderately dysarthric patients, and 56% for the heavily dysarthric patients (Audibert et al., 2010).
In order to enhance the quality of the automatic alignment, already quite satisfactory for most speakers (control and moderate), the system was tuned by combining the information of 2 different sets of acoustic models. This optimization improves the overall performance and notably that of heavily dysarthric patients (15% on control speakers, 23% on moderate, and 44% on heavily dysarthric patients). It has to be outlined that the altered productions of the latter set of patients were also hard to segment for the human experts.
A comparison between manual and automatic alignments was also done in terms of their consequences on specific acoustic measurements: segment duration, formant frequency, fricative center of gravity (Fougeron et al., 2010). While temporal measurements extracted from automatic alignment have to be interpreted with caution, spectral measurements (both local in the middle of a vowel, or global over the fricative duration) are comparable with those extracted from a manual alignment. These first results are encouraging regarding the possibility of using automatic alignment for some of the acoustic dimensions to be analyzed in our project.

5. Conclusion and Issues

The understanding of dysarthric speech patterns has evident implications for clinical research on speech disorders, but also for contemporary issues in Speech Science in general.
Recent developments in phonetics and phonology show a trend away from observing the language system towards observing the user of the system. From this perspective, disordered speech is a challenging and promising test case. Basic tenets of our project rely on the assumption that our understanding of speech production proceeds with advances in the study of both normal and disordered speech and that a good model has to unify knowledge from both populations. In that respect, observing the types and range of variation linked to a motoric deficit, such as in dysarthria, is of the utmost interest for a comprehensive model of speech variation. Indeed, it raises challenging issues related to the factors governing variation in speech production in general. Models of variation in phonetics need input from disordered speech patterns, whereas the definition of disordered speech needs references from normal variation. A better understanding of the variation that characterizes dysarthric speech as deviant could thus provide insights into the blurred boundary between normal and pathological speech patterns. In return, dysarthric productions, and their variations, may inform us about normal speaker adaptation to different speech situations. While some progress has been made on the characterization of the acoustic-phonetic properties of dysarthric speech, our knowledge is still limited. The study of disordered speech is at the crossroads between different sub-disciplines of Speech Sciences, and multidisciplinary collaborations, such as the one proposed here, promise progress in this area.

2 20ms corresponds to 2 frames in the automatic alignment system.

Acknowledgments

This project is funded by the ANR BLAN08-0125 of the French National Research Agency. We deeply thank Pierre Clément, Aurélie Nuremberg, and Olavo Panseri who are also collaborating to this project.

References

Adda-Decker, M. & Gendrot, C. & Nguyen, N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français, Traitement Automatique des Langues, 49, n°3, 13-46.

Audibert, N. & Fougeron, C. & Fredouille, C. & Meunier, C. & Panseri, O. (2010). Evaluation d'un alignement automatique sur la parole dysarthrique. 18èmes JEP, Mons, Belgium

Auzou, P. & Ozsancak, C. & Pinto, S. & Rolland, V. (2007). Les dysarthries, Marseille: Solal.

Baudelle, E. & Vaissière, J. & Renard, J. L. & Roubeau, B. & Chevrie-Muller, C. (2003). Caractéristiques vocaliques intrinsèques et co-intrinsèques dans les dysarthries cérébelleuses et parkinsoniennes. Folia Phoniatrica et Logopedica 55, 137-146.

Collins, M. (1984). Integrating perceptual and instrumental procedures in dysarthria assessment. Journal of Communication Disorders, 5, 159–170.

Darley, F. L. & Aronson, A. E. & Brown, J. R. (1969) Clusters of Deviant Speech Dimensions in the Dysarthrias. Journal of Speech and Hearing Research, 12: 462-496.

Darley, F. L. & Aronson, A.E. & Brown, J.R. (1969). Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research, 12: 246-269.

Darley, F. L & Aronson, A. E. & Brown, J. R. (1975). Motor Speech Disorders. Philadelphia: W.B. Saunders.

Duez, D. (2006). Syllable structure, syllable duration and final lengthening in Parkinsonian French speech, Journal of Multilingual Communication Disorders, 4, 1, p. 45-57.

Duffy, J. R. (1995). Motor Speech Disorders: Substrates, differential diagnostics and management. St Louis: Mosby-YearBook, Inc.

Duffy J. R. (2005). Motor speech disorders: substrates,differential diagnosis and management. Mosby-Yearbook. St. Louis.

Fougeron, C. & Audibert, N. & Fredouille, C. & Meunier, C. & Gendrot, C. & Panseri, O. (2010). Comparaison d'analyses phonétiques de parole dysarthrique basées sur un alignement manuel et un alignement automatique. 18èmes JEP, Mons, Belgium

Gentil, M. & Pinto, S. & Pollak, P. & Benadbid, A. L. (2003). Effect of bilateral stimulation of the subthalamic nucleus on Parkinsonian dysarthria. Brain and Language,85, 190-196.

Grewel, F. (1957). Classification of dysarthrias. Acta Psychiatrica Neurologica Scandinavica, 32, 325-337.

Gu, L. & Harris, J. G. & Rahul, S. & Sapienza, C. (2005). Disordered speech evaluation using objective quality measures. In proc. of ICASSP'05. Philadelphia, US.

Kent, R. D. & Kent, J. F. & Duffy, J. R. & Weismer, G. (1998) The dysarthrias: Speech-voice profiles, related dysfunctions, and neuropathologies. Journal of Medical Speech-Language Pathology, 6, 165–211.

Kent, R. D. & Weismer, G. & Kent, J. F. & Vorperian, H. K. & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. The Journal of Communication Disorders 32, 3: 141-186.

Maier, A. & Schuster, M. & Batliner, A. & Nöth, E. & Nkenke, E. (2007). Automatic scoring of the intelligibility in patients with cancer of the oral cavity. In proc. of Interspeech'07, Antwerpen, Belgium.

McNeil, M. R. (1997). Clinical management of sensorimotor speech disorders. New York: Thieme, 1997.

Middag, C. & Martens, J.-P. & Van Nuffelen, G. & De Bodt, M. (2009). Automated intelligibility assessment of pathological speech using phonological features. EURASIP Journal on Advances in Signal Processing. v. 2009.

Monfrais-Pfauwadel, M. C. (1995). Les disfluences autres que celles du bégaiement. Revue de laryngologie, d'otologie et de rhinologie, 116(4), pp. 267-270.

Mori H. & Kobayashi Y. & Kasuya H. & Hirose H. & Kobayashi N. (2004). Prosodic and Segmental Evaluation of Dysarthric Speech, Proc. Speech Prosody, Nara, Japan, 4 p.

Murdoch, B. (1998). Dysarthria - A Physiological Approach to Assessment and Treatment. Nelson Thornes Ltd.

Peacher, W. G. (1950). The etiology and differential diagnosis of dysarthria. Journal of Speech and Hearing Disorders, 15: 252–265, 1950.

Pinto S. & Gentil M. & Krack P. & Sauleau P. & Fraix V. & Benabid A.-L. & Pollak P. (2005). Changes induced by levodopa and subthalamic nucleus stimulation on Parkinsonian speech. Movement Disorders, vol. 20, no. 11. 2005, p. 1507-1515.

Robert D. & Sangla I. & Azulay J.P. & Giovanni A. & Cannoni M. & Pouget J. (1995). Diagnostic et suivi de l'insuffisance vélaire dans les formes bulbaires des maladies du motoneurone. Actes du congrès sur le Voile Pathologique, Société Française de phoniatrie, Lyon, p.63-74.

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 , 119:724-731

Sarr M. & Pinto S. & Jankowksi L. & Purson A. & Ghio A. & Espesser R. & Teston B. & Viallet F. (2009). L-dopa and STN stimulation effects on pneumophonic coordination in Parkinsonian dysarthria: intra-oral pressure measurements. International Congress of Parkinson's Disease and Movement Disorders, Movement Disorders, vol. 24, no. S1. 2009, p. S342.

Schlenck K.-J. & Bettrich R. & Willmes K. (1993). Aspects of disturbed prosody in dysarthria, Clinical Linguistics & Phonetics, Vol. 7, No. 2, Pages 119-128.

Su, H. Y. & Wu, C. H. & Tsai, P. J. (2008). Automatic assessment of articulation disorders using confident unit-based model adaptation. In proc. of ICASSP, Las Vegas, US.

Teston B. & Ghio A. & Galindo B. (1999). A multisensor data acquisition and processing system for speech production investigation. In proc. of ICPHS'99, p.2251-2254.

Teston, B. & Galindo, A. (1995). A Diagnostic and Rehabilitation Aid Workstation for Speech and Voice Pathologies. In proc. of Eurospeech'95, Madrid Spain.

Viallet, F. & Jankowski, L. & Purson, A. & Teston, B. (2004). Dopa effects on laryngeal dysfunction in Parkinson's disease: An acoustic and aerodynamic study, International Congress of Parkinson's Disease and Movement Disorders. Movement Disorders, vol. 19, Suppl. 9, p. S237.

Vijayalakshmi, P. & Reddy, M. R. & O'Shaughnessy, D. (2006). Assessment of articulatory sub-systems of dysarthric speech using an isolated-style phoneme recognition system. In proc. of Interspeech'06, Pittsburgh, US.

Table I : *Speech material recorded in the CCM database. A '*' in the second column indicate whether the material is available for all recordings.*

Speech tasks	
• the production of automatic series (counting from 1 to 10 and/or 1 to 20, month of the years),	*
• two readings of a sentence and its repetitions • ("C'est une affaire intéressante, qu'en pensez-vous? Il faut la faire sans aucun regret")	*
• the reading of two lists of words (Bonjour, Femme, Chasseur, Légat, Exploit, Gargarisme, Voleur, Banane, Coupe, Coupe-papier, Spectacle, Un match de boxe, Jaser, Magique) ; (Bonjour, Jaser, Légat, Banane, Voleur, Coupe-papier, Justice, Zèbre, Magique, Exploit, Chasseur, Carré)	*
• the production of sustained vowels (/a/, /e/, /i/, /o/)	*
• the reading of a text (a fairy tale of 170 words, 'Le cordonnier')	
• a story telling based on a picture support • ("La chute dans la boue", based on a test evaluating language acquisition of children)	
• spontaneous speech (narrating the day's activities)	
• syllable repetition (CV, VC or VCV with V= [a] and C= [p, t, k, S, s, f, b, d, g, Z, z, v, l, R, m, n, j])	