Neurocomputing 167 (2015) 83–93

Contents lists available at ScienceDirect

Neurocomputing

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

Phonation biomechanic analysis of Alzheimer's Disease cases

Pedro Gómez-Vilda [a,*], Victoria Rodellar-Biarge [a], Víctor Nieto-Lluis [a], Karmele López de Ipiña [b], Agustín Álvarez-Marquina [a], Rafael Martínez-Olalla [a], Miriam Ecay-Torres [c], Pablo Martínez-Lage [c]

[a] Biomedical Technology Center, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid, Spain
[b] Department of Systems Engineering and Automation, University of the Basque Country, Donostia, Gipuzkoa, Spain
[c] Neurology Department/CITA-Alzheimer Foundation, Donostia, Gipuzkoa, Spain

ARTICLE INFO

Article history:
Received 2 December 2014
Received in revised form
4 March 2015
Accepted 6 March 2015
Communicated by: Carlos Manuel Travieso-González
Available online 22 May 2015

Keywords:
Speech processing
Cognitive disease monitoring
Neurodegenerative speech modeling
Aging-care
e-Health systems

ABSTRACT

Speech production in patients suffering of dementias of Alzheimer's type is known to experience noticeable changes with respect to normative speakers. Classically this kind of speech has been described as presenting altered prosody, rhythmic pace, anomy, or impaired semantics. Phonation, conceived as the production of voice in voiced speech fragments remains as an unexplored field. The aim of the present paper is to open a preliminary study presenting biomechanical estimates from phonation produced by two patients (male and female) suffering Alzheimer's Disease (AD), contrasted on two controls of both genders (CS: control speakers). A vocal fold biomechanical model is inverted to facilitate estimates of the vocal fold stiffness to analyze significant segments of phonated speech as long vowels and fillers. The estimates of both the AD patients and CS subjects are contrasted on a database of phonation features from a normative speaker population of both genders, as well as in paired tests contrasting AD and CS subjects. Results show the possibility of establishing significant discrimination between AD and CS when using *f*0, as well as vocal fold body stiffness, although this last feature seems to be more relevant and shows larger statistical significance.

1. Introduction

Since the pioneering study of Auguste Deter's case by Alois Alzheimer, where a first description of what we nowadays recognize to be the main dramatic cause of dementia severely impairing quality of third-age life in developed societies, one hundred and ten years have passed, and a definitive cure of the disease is still waiting. A first difficulty comes from the fact that AD seems to be the result of severe biochemical degradation of neural tissues in the brain cortex, eventually extending to other deeper structures, whose nature is still the subject of intensive studies, which are far from concluding [1]. Thus, the difficulty of finding pharmacological treatments is still high, and strong research efforts have to be devoted to this task. A related problem is that establishing a clear differential diagnose of AD type of dementias is very complicate, according to neurologist experts. The prevalence of AD is around 30 million people worldwide [2], with an incidence of 4.6 million of new cases per year (one case every 7 s), so as to expect the number of AD cases to climb up to 81.1 million cases by 2040 (nearly doubling every 20 y). With this panorama in mind it is of most importance that applied technology contributes to medicine, pharmacology and biochemistry in helping to diagnose and monitor AD facing patient treatment. It is in this scenario where speech sciences can contribute with inexpensive methodologies and procedures in establishing an early diagnose, and a continued longitudinal monitoring of patients. The anticipated diagnose is especially relevant, as it can contribute to a better treatment slowing the progression of the disease. Speech technologies may offer a simple way of assessing early symptoms by the analysis of AD speech characteristics based on the “dry-lab” concept, in which on-line objective analysis may be available to the neurologist in real time, not depending in other more expensive tests with long turn-over time lags. This fact is already considered by neurologists, as tests where the patient is asked to produce speech are currently among the diagnose protocol. Nevertheless the evaluation of AD speech using that methodology may become rather subjective and depending on the specific rating judge opinion. Therefore, under this point of view a good description of AD-affected speech is of most interest. It is well documented that AD speech is characterized by a decline in semantic abilities, as well as by deficits in discourse and prosody, known with the global name of non-fluent aphasia. Anomia (the difficulty of assigning names to things or persons) and single word comprehension deficits are also other related symptoms. Some studies also mention deficits in production, repetition and comprehension. Other evidence suggests a

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.03.087
0925-2312/

degree of vulnerability in articulatory and phonological levels as well. Some related descriptions can be found in [3–6]. The main observable behavioral trends in AD speech can be summarized under the terms of dysfluency (aphasic, or anomic), dysprosody (emotional decay, production, repetition and comprehension) and agrammatism (deficient syntax). The following is a brief description of the behavior observables associated to these descriptions:

- Dysfluency may result in broken speech rhythm, therefore, the control of speech presence will be crucial. Intervals with no speech and with speech will have to be carefully detected and measured.
- Dysprosody may be traced from the fundamental frequency *f*0 and the intensity of speech production. Repetition and comprehension are left for other kinds of studies.
- Agrammatism does not leave important acoustical marks and is to be studied using other techniques, as automatic linguistic speech recognition [7].

The most important referents under the speech processing point of view are the acoustic correlates which can be associated to each behavioral observable. These are some of the most relevant ones

- Sub-segmental intervals: syllables defined as associations of consonant–vowel structures (V, CV, VC, CVC or VCV, where V stands for vowel and C for consonant), and inter-syllabic pauses. Verbal rate as the number of syllables produced per unit time is an important feature.
- Fillers, as prolonged vowels in syllable endings or independent insertions (such/uh/,/ah/,/eh/and/hmmm/-like phonations).
- Supra-segmental intervals: groups of continuous voiced speech with no separation (phonation groups), and pause intervals separating these groups. Mean duration of pauses is another important feature.
- *f*0 or fundamental frequency profile, estimated for phonation groups, defining the prosodic contour of the sub- and supra-segmental intervals. Cepstral peak [8], SWIPE [9] and DYPSA [10] are popular methods for *f*0 estimation.
- Intensity profile, estimated for both the voiced and unvoiced segments of speech, defining the energy envelope of sub- and supra-segmental intervals. Low-pass energy envelope filtering and Teager–Kaiser algorithms are popular estimation methods currently used [11].
- Emotional arousal and temperature are related concepts which are being currently defined, based on complexity studies as well as in non-linear methods to establish an average estimation of the patient's emotional state [12].

Researchers have resorted to study the degradation of speech at different linguistic levels, ranging from the phonological and articulatory[4,6]. The syntactic and lexical-semantic levels have been also studied in [13–17], although these are out of the scope of the present work. The *f*0 contour and intensity profile, as well as temporal interval estimations are usual features at the acoustic level, which can be found in the same studies.

More recently, language and speech researchers have concentrated in finding markers in order to early diagnose and monitor the progression of the disease ([16–19]). In [20] five temporal measures are explored, finding that verbal rate, mean duration of pauses, and standardized phonation time are significant markers in the differentiation between AD and healthy subjects, concluding that the length of pauses is more relevant than their number. Following this study, some researchers have used these indices, combining them with other measures, to develop algorithms for automating feature extraction and analysis processes. Several other researchers have investigated the prosodic production and recognition of emotions in AD ([5], and [21,22]), finding that these markers were significantly altered in AD patients. The inclusion of voice quality defined as a vehicular index for emotional expression [23,24], instead of taking into account only *f*0 and intensity profiles, can be considered a relevant contribution.

The present paper is intended to carry on an exploratory study about the possible influence of AD in basic phonation. It is well known that diseases of neuro-motor origin affect vocal fold biomechanics [25]. As the nature of AD is mainly cognitive, it cannot be inferred that this type of diseases could influence biomechanics. Nevertheless, it is well documented in the literature that AD speech presents distortions in *f*0 and energy, as commented above. These correlates can be related closely to vocal fold biomechanics. Therefore, it would be possible to conclude that AD affects also vocal fold biomechanics. It is well known that *f*0 is sustained by the biomechanical parameters of the vocal folds, mainly by the mechanical stiffness of the vocal fold body. This factor is controlled by the laryngeal nerves, activated by the neuromotor speech planner through the bulbar system. Although secondary neurons do not seem to be affected by AD progress (at least at early stages), it is true that in late stages of AD many patients succumb to pneumonic infections when the laryngeal reflex is impaired. These two facts, i.e., progressive dystonic phonation and loss of laryngeal reflex in AD may indicate that some kind of laryngeal biomechanical deterioration may be expected from AD as well [26].

The intention of this research is to give some hints on the study of biomechanical analysis of phonation in AD patients (Section 2) to determine possible markers showing these effects in their statistical distributions, as an extension to [27]. The ultimate objective will be to launch a massive study on a large database, which is to be built based on these premises. Section 3 discusses the materials and methods used in this exploratory study, necessarily concentrated in a few study cases. Preliminary results are discussed in Section 4. Concluding observations and remarks are given in Section 5.

2. Biomechanical analysis of phonation

Studies in AD speech production have concentrated their scope in phonological and articulatory, syntactic, lexical-semantic, and acoustic levels, including both temporal measurements and *f*0 production, as discussed in Section 4. Not much interest has been devoted to phonation in itself (glottal source production in voiced speech). Apparently this line has not been taken into account, allegedly because AD being a disease of cognitive origin, phonation is considered mainly a feature controlled by neuromotor activity (peripheral to the central nervous system) [28]. However, this scope cannot be supported without confirmation or refutation.

The key technique used for the analysis of voice quality in the present work is adaptive vocal tract inversion to produce an estimate of the glottal source. From this signal a possible approach to vocal fold biomechanics is granted. Specific behavior of vocal folds, as muscular tension, is directly related to neuro-motor activity. Whether the irregular behavior of this activity be considered peripheral or central, is a matter of investigation.

This work proposes the use of acurate spectral domain techniques [29] allow the estimation of a set of biomechanical parameters associated to a 2-mass model of the vocal folds [30] as the one depicted in Fig. 1. Template (a) shows the classical body-cover structure of the vocal folds. Template (b) on its turn gives a biomechanical model of the structures depicted in template (a). The average dynamic mass of the body contributing to vibration of the body is represented by masses M_{br} and M_{bl}, for the right and