Evaluation of Phonatory Behavior and Voice Quality in Patients with Multiple Sclerosis Treated with Deep Brain Stimulation

***Manfred Pützer, †Wolfgang Wokurek, and ‡Jean Richard Moringlane,** **‡Saarbrücken and †Stuttgart, Germany*

Summary: Objectives. The effect of deep brain stimulation (DBS) on phonatory behavior and voice quality in eight patients with multiple sclerosis (MS) was examined instrumentally and perceptually. The acoustic signals of vowel productions obtained from patients (produced with and without stimulation) and from a group of 16 healthy control speakers were analyzed to prove statistically the changes of phonatory behavior and voice quality.

Study Design. This is a randomized study.

Methods. Firstly, a new parametrization was used to determine phonatory behavior. Secondly, a perceptual evaluation of voice quality of the same speech material was performed.

Results. With stimulation, phonation has a greater tendency to be strained. The results of perceptual evaluation support this strained phonation behavior under stimulation, resulting in a smaller degree of breathiness ratings of all raters. Without stimulation, an impaired and partly disturbed adduction of the vocal folds can be shown. These findings are also supported in the perceptual experiment providing a higher degree of hoarseness ratings of all raters for these signals.

Conclusions. High-frequency electrical impulses to the thalamus in patients with MS influence the phonatory behavior of their vocal folds. The results suggest the need for long-term monitoring of phonatory behavior during DBS to initiate adequate treatments without delay.

Key Words: Multiple sclerosis–Deep brain stimulation–Acoustic analysis–Perceptual evaluation–Correlation analysis.

INTRODUCTION

Multiple sclerosis (MS) leads to diverse disturbances because of axon decline in disseminated areas of the central nervous system as well as because of demyelinization. Symptoms like tremor of the upper extremities, head tremor, and ataxia are frequent complications of this pathology.[1] The extremity tremor is primarily an intention tremor with a significant action and postural component. Further, changes of phonatory behavior and glottal-supraglottal articulation disturbances may also appear for patients as accompanying symptoms to these movement disorders. In this regard, they can show different dysarthric symptoms, for example, in cases of cerebellar ataxia, patients may have rough voice quality with pitch and loudness perturbations and, additionally, reduced precision of articulation behavior.[2] These symptoms often already exist before treatment, but they are also found as undesirable symptoms after deep brain stimulation (DBS) for particular kinds of tremor.[3]

Surgery for patients with MS tremor has been aimed at the ventral intermediate nucleus (v.i.m.) of the thalamus including the subthalamic area. Stereotactic thalamotomy provides good control at first but the benefit declines over time.[4] For this reason, and also because of the excellent results obtained with chronic DBS, this neurosurgical method has been used to treat patients with MS, Parkinson's disease, and other neurological pathologies like dystonia and essential and cerebellar tremor.[5–7]

During the application of DBS, the effect on glottal phonation and on other speech subsystems has been studied in several papers.[8–16]

The purpose of the present study is to investigate, instrumentally and perceptually, characteristics of phonatory behavior and voice quality, respectively, in patients with MS treated with DBS. Consequently, the aim is to evaluate the effects of DBS itself rather than compare the preoperative and postoperative status of the patients' phonatory behavior and voice quality. Although we carefully reviewed the recent literature, we could not find a methodological approach comparable to the one used again by our research group in the present study.[17] Thus, to the best of our knowledge this procedure has not been applied before.

In addition to this main purpose, the study has been undertaken to validate findings of a previous investigation of our research group, in which the acoustic signals were analyzed using parameters of the Multidimensional Voice Program (Kay Elemetrics model 4338).[17] For this reason, a new parametrization has been applied to characterize the phonatory behavior of the patients under the two conditions and to compare it to the one of a healthy control group. This parametrization has been successfully used in our previous studies to demonstrate phonatory behavior and voice quality properties.[18,19]

MATERIALS AND METHODS

Patients and control group

Eight patients suffering from MS (two males and six females) were studied. They were treated with high-frequency electrical impulses to the left and/or right ventrolateral area of the thalamus. The impulse generator model ITREL II (Medtronic, Medtronic GmbH, Earl-Bakken-Platz 1, D-40670 Meerbusch) was implanted in the subclavicular region. Comprehensive details of the surgical technique have been published elsewhere.[20,21]

Accepted for publication October 27, 2016.
From the *Language Science and Technology, Saarland University, Saarbrücken, Germany; †Institute for Natural Language Processing, University of Stuttgart, Stuttgart, Germany; and the ‡Department of Neurosurgery, Saarland University, Saarbrücken, Germany.
Address correspondence and reprint requests to Manfred Pützer, Language Science and Technology, Saarland University, Postfach 15 11 50, D-66123 Saarbrücken, Germany. E-mail: puetzer@coli.uni-saarland.de
Journal of Voice, Vol. 31, No. 4, pp. 483–489
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.10.022

Informed consent had been obtained in line with the Declaration of Helsinki and the Institutional Review Board of the Saarland University. The participants provided informed verbal consent. All patients were diagnosed as having definite MS following the criteria of Poser et al.[22] They exhibited a severe combined upper and lower limb, truncal, and gait ataxia, which could not be sufficiently suppressed pharmacologically. In two male patients, severe head and shoulder tremor was particularly accentuated. The patients who were not able to use their upper extremities efficiently were dependent on foreign aid for all activities of daily living: eating, drinking, brushing teeth, combing hair, dressing, hygienic care, etc. Dialing a telephone number and writing were impossible.

The two male patients were aged 42 and 45 years at the time of speech registration. The six female patients ranged in age from 35 to 47 years. Mean (SD) age was 37 (3) years.

The expanded disability status scale (EDSS[23]) scores for the two male patients were 7.0 and 8.0. They were predominantly wheelchair bound. Two female patients had an EDSS score of 6.5 and the other four female patients had 7.0 (also predominantly wheelchair bound). Patients were being treated for upper limb ataxia and intention tremor and were therefore available for recording without additional inconvenience to themselves or their carers. Because upper limb ataxia and intention tremor, not voice and speech quality, were the reasons for undertaking DBS, there had been no preoperative dysarthria or voice quality classification. Sixteen healthy subjects (eight males and eight females) with no known speaking and hearing problems and a mean (SD) age of 36 (4) years served as control subjects.

Speech material and recording procedure

Subjects were required to produce the continuous vowels [i:], [a:], and [u:] in normal pitch. Patients produced the vowels with and without stimulation. The microphone signal was recorded in a sound-treated room, using a neckband condenser microphone (NEM 192.15, Beyerdynamic, Heilbronn, Deutschland). The signal was fed directly into a Computerized Speech Lab (CSL station, model 4300B) at a sampling rate of 50 kHz. Amplitude resolution was 16 bit. For each signal, a portion of more than half a second between positive zero-crossing was selected, starting 0.5 s after the beginning of phonation.

Acoustic measurements

Phonatory behavior is parameterized in this study by the bandwidth of the first formant and by spectral gradient features of the source spectrum.[24] The original parameters were defined by Stevens and Hanson.[25] They noticed that different aspects of phonation quality dominate certain parts of the glottal excitation spectrum. Five gradient parameters capture these aspects of the source spectrum. One bandwidth parameter captures the damping of the first formant introduced by the incomplete closed glottis.

Thus, our approach differs from Stevens and Hanson by using spectral slope parameters. As a consequence, we can account for a more complete compensation of the vocal tract filter. Finally, a frequency normalization makes one of the parameters (incompleteness of closure (IC)) more vowel independent. The more complete compensation of the vocal tract filter uses the first four

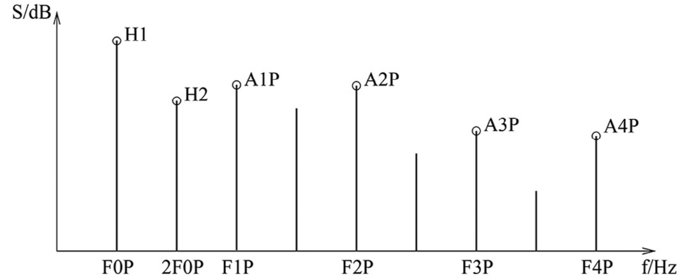

FIGURE 1. Schematic harmonic speech spectrum.

formant frequencies F1, . . ., F4 and formant bandwidths B1, . . ., B4 from Linear Predictive Coding (LPC) estimation. All four formants will be compensated to closely approach the source spectrum. Using spectral slope parameters instead of spectral amplitude decibel differences makes the parameters more independent of both changes in fundamental frequency (FO) and changes of vowel quality. Usually the spectrum of the voiced source decays with increased frequency. Hence an increased formant frequency has lower spectral amplitude and the spectral distance to the reference point, for example H1~ - A1P~, increases. The spectral gradient uses the decay rate instead of the amplitude difference and is less dependent on the formant frequency. The spectral slope better models the source spectrum envelope and will be less changed by these frequency shifts.

First, a spectrum is calculated using a 25 ms Hamming window, long enough to show the harmonic structure.

Figure 1 shows a schematic harmonic speech spectrum. Denoted are the amplitudes and frequencies that are used for the calculation of voice quality parameters: the first harmonic or fundamental oscillation H1 at F0P, the second harmonic H2 at 2F0P, and the harmonic next to each of the first four formants A1P,. . ., A4P at F1P,. . ., F4P.

Now the gain of the four formants is subtracted (in decibel amplitude scale) to estimate the source spectrum shown schematically in Figure 2.

Figure 2 also shows the triangle slopes that are our voice quality parameters. We keep the parameter names of Stevens and Hanson. OQ stands for opening quotient and the spectral slope parameter is OQG with the appended G for gradient. The unit is decibels per barks because the perceptual bark scale is used. GO stands for glottal opening and defines the voice quality parameter GOG. SK stands for skewness; RC stands for rate of closure. To our experience, these names do not correlate for every speaker what

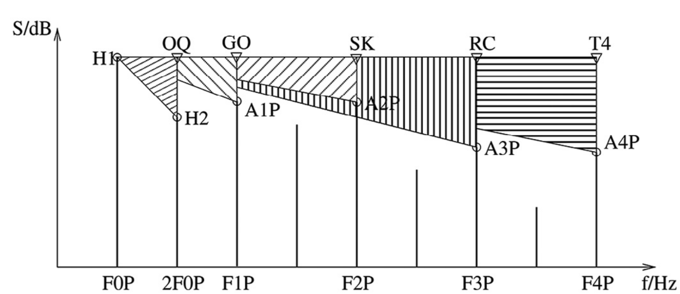

FIGURE 2. Schematic harmonic voice spectrum and parameter definitions.