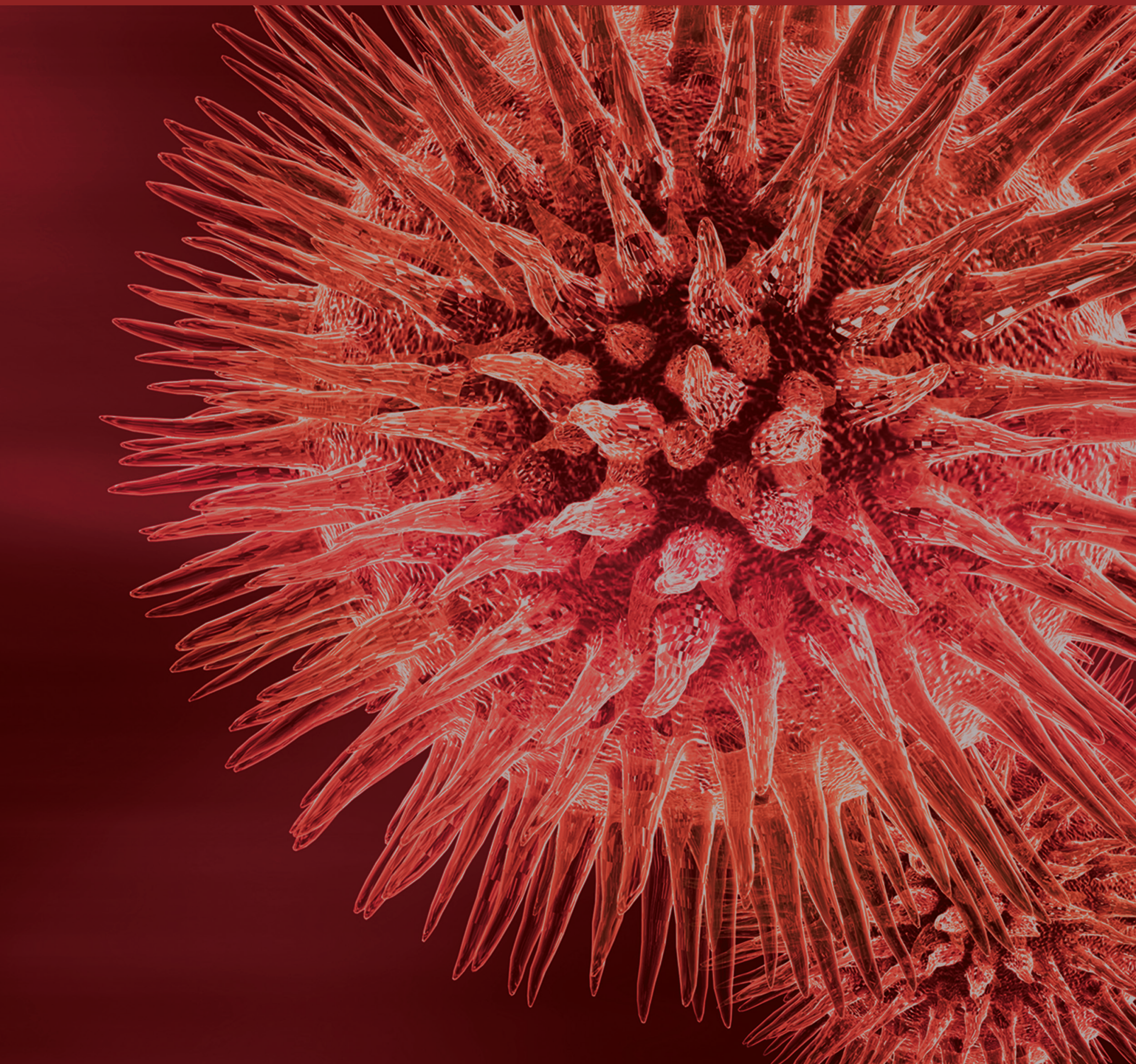
BioMed Research International

Phoniatrics

Guest Editors: Haldun Oguz, Markus Hess, and Adam M. Klein



Phoniatrics

BioMed Research International

Phoniatrics

Guest Editors: Haldun Oguz, Markus Hess,
and Adam M. Klein

Contents

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 156014, 2 pages
http://dx.doi.org/10.1155/2015/156014

Editorial

Phoniatrics

Haldun Oguz,[1] Markus Hess,[2] and Adam M. Klein[3]

[1] *Department of Otolaryngology, Faculty of Medicine, Yüksek İhtisas Üniversitesi, Ankara, Turkey*
[2] *Department of Voice-Speech-Hearing Disorders, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[3] *Voice Center, Emory University, Atlanta, GA, USA*

Correspondence should be addressed to Haldun Oguz; dr@haldunoguz.com

Received 6 August 2015; Accepted 27 August 2015

Phoniatrics is the medical specialty for communicative disorders. It is related with the normal, pathological, and professional processes that involve voice, speech, language, swallowing, and hearing. Besides being a specific discipline, it is also a discipline that combines the accumulation of knowledge from both medical and nonmedical fields of science. Most important medical specialties that contribute to phoniatrics are otorhinolaryngology, neurology, pediatrics, psychiatry, pediatric psychiatry, dentistry, gastroenterology, geriatrics, endocrinology, radiology, genetics, and physical therapy. Otorhinolaryngology (ORL) Section and Board of European Union of Medical Specialists (UEMS) decided "phoniatrics—communication disorders" as one of the seven subspecialities of otorhinolaryngology in 2010 [1]. Many nonmedical specialties also contribute to phoniatrics. These may be named as speech and language pathology, audiology, linguistics, child development, psychology, pediatric psychology, physiotherapy, acoustics, physics, biomedical sciences, pedagogy, and computer information technologies.

The field of *voice* ranges from acoustics and aerodynamics, to the biomechanics of voice, articulation, and resonance, to voice training, vocal ergonomics, and vocal hygiene, to the complex neurologic, psychiatric, and sociological basis of interpersonal communication. The incidence of voice disorders in the general population is considered between 3% and 9% [2, 3]. This incidence is considered much higher for children, being reported up to 24% [4]. These rates only highlight the importance of increasing awareness of voice disorders.

Understanding vocal production is a crucial step in diagnosing and treating voice ailments. In this issue, you will find an article by A. A. Poznyakovskiy et al. that discusses the importance of vocal tract morphology in voice production by using a 3-tesla magnetic resonance imaging model. They showed that the method is suitable for an improved in-detail analysis of the vocal tract morphology during speech and singing. The development of new diagnostic tools continues to improve our ability to properly identify vocal pathologies. To this end, E. Gonzalez-Moreira and colleagues discussed the utilization of computational speech analysis to identify dementia in elderly patients. Also in this issue, L. Moro-Velázquez and colleagues discuss the efficacy of automatized evaluation of subjective voice parameters by modulation spectra morphological parameters. They demonstrate efficiencies of 81.6% and 84.7% for Grade and Roughness, respectively, using modulation spectra parameters. In another article, L. M. T. Jesus et al. pointed out the acoustic correlates of compensatory adjustment in vocal fold paralysis patients. Proper understanding and diagnosis lead to optimization of vocal restoration. This issue shares several modern methods of managing laryngeal pathology. B. Svensson and colleagues continued to work on restoring vocal fold function after injury in an animal study. In their study they have shown that human embryonic stem cells transplanted to injured rabbit vocal folds restored their vibratory characteristics. H. K. Byeon and colleagues reported their experience with pulsed dye laser in treatment of hemorrhagic vocal fold polyps. It was shown that PDL-assisted enucleation laryngeal microsurgery for the treatment of hemorrhagic vocal polyps can be a safe and effective surgical technique, with improvements in subjective voice perception.

Speech and Language Disorders Encompass Articulation Disorders. Diagnostics, evaluation methods, intelligibility, motor examination, nasalance, preventive counselling, speech therapy, and medical and surgical treatment; developmental language disorders: psychomotor, cognitive, auditory, and language development of normal and diseased children, evaluation of verbal and nonverbal communication, psychomotor, vestibular, and kinesthetic development and disorders, prevention, and rehabilitation; acquired language disorders: evaluation of verbal and nonverbal communication, speech perception, speech and motor evaluation, neurological, laboratory and radiological examination, rehabilitation, and treatment are all discussed. The first human speech dates to at least 50,000 years ago [5]. However, diagnosing and treating speech pathologies are still a clinical challenge. The prevalence of speech sound disorders in children is 8-9%. By the first grade, nearly 5% of children have speech disorders. Most of these disorders are idiopathic in origin [2]. The prevalence of combined receptive and expressive language disorders in children below 7 years ranges between 2 and 3% [6]. The prevalence of language impairment in kindergarten children is 8% [7]. Anyone can acquire aphasia at any age. However, most of the patients fall into an age group that involves middle to late adulthood. 0.3% of U.S. population currently suffer from aphasia [2, 3]. Stuttering is seen in all ages but it is most frequent in young children between the ages of 2 and 6, while they are developing language. Boys are more prone to stuttering during this period. The prevalence of stuttering is estimated to be less than 1% in adults [2]. In the following pages, C. Reuterskiöld and M. I. Grigos are discussing the influence of repeating real words and nonwords on speech motor control in children.

Swallowing Disorders. Diagnosis, developmental stages of swallowing, drooling, penetration, retention, regurgitation, aspiration, endoscopic and radiological evaluation, rehabilitation, and medical and surgical treatment are all discussed. Swallowing disorders are seen more frequently in the elderly. The prevalence of solid food dysphagia is 7% in elderly patients [8]. This makes the field of phoniatrics more important in today's ageing society. Swallowing disorders are also more frequent in patients with gastroesophageal reflux and in patients with neurological diseases or with a history of cerebrovascular stroke [9]. Untreated or poorly treated swallowing disorders may result in dehydration, malnutrition, respiratory pathologies, and death [9].

Hearing Disorders. Physiology and pathology of hearing, tympanometry, audiological evaluation, otoacoustic emission, evoked response audiometry, sensorineural, conductive, and mixed hearing loss, diseases of external, middle, and inner ear, diseases of cochlear nerve, brainstem, and auditory cortex, auditory processing disorders, cochlear implants, hearing aids, prevention, medical and surgical treatment, rehabilitation, and tinnitus are all discussed. According to World Health Organization estimates, 5.3% of world's population have disabling hearing loss [10]. 9% of these are children. Approximately one-third of population above 65 years are affected by disabling hearing loss [10]. The article by Z. Y. T. Chan and B. McPherson discusses the appropriateness of the over-the-counter hearing aids for the demand in the presbycusis. In this issue, you will also find an interesting article by S. S. Grasel and colleagues that discusses the use of auditory steady state responses in evaluation of children with severe hearing loss. Another important paper by M. Coene is discussing the spoken word recognition errors in speech audiometry on normal subjects with complete speech intelligibility.

Acknowledgment

Many thanks go to all authors, who put a lot of energy into their studies and writing of manuscripts to share with us their results and conclusions. We thank all reviewers for their effort in enabling these manuscripts to fit into the journals form and focus. We would like to convey our cordial thanks to the editors of BioMed Research International Journal for giving us the opportunity to publish this special issue.

Haldun Oguz
Markus Hess
Adam M. Klein

References

[1] European Union of Medical Specialities website, 2015, http://orluems.com/gestor/upload/file/Minutes%20Malta%202010/Minutes%20meeting%20Malta%202010.pdf.

[2] National Institute on Deafness and Other Communication Disorders, Statistics on Voice, Speech and Language, August 2015, http://www.nidcd.nih.gov/health/statistics/pages/vsl.aspx.

[3] US and World Population Clock, United States Census Bureau, 2015, http://www.census.gov/popclock/.

[4] M. Gerek and H. Birkent, "Bölüm 1. Klinik Ses Bozukluklarına Giriş," in *Klinik Ses Bozuklukları*, M. A. Kilic and H. Oguz, Eds., pp. 1–9, Nobel Kitabevi, 2012.

[5] P. Lieberman, "The evolution of human speech," *Current Anthropology*, vol. 48, no. 1, pp. 39–66, 2007.

[6] J. Law, J. Boyle, F. Harris, A. Harkness, and C. Nye, "Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature," *International Journal of Language and Communication Disorders*, vol. 35, no. 2, pp. 165–188, 2000.

[7] J. H. Beitchman, R. Nair, M. Clegg, and P. G. Patel, "Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region," *Journal of Speech and Hearing Disorders*, vol. 51, no. 2, pp. 98–110, 1986.

[8] J. W. Mold, L. E. Reed, A. B. Davis, M. L. Allen, D. L. Decktor, and M. Robinson, "Prevalence of gastroesophageal reflux in elderly patients in a primary care setting," *The American Journal of Gastroenterology*, vol. 86, no. 8, pp. 965–970, 1991.

[9] T. Wilkins, R. A. Gillies, A. M. Thomas, and P. J. Wagner, "The prevalence of dysphagia in primary care patients: a HamesNet research network study," *Journal of the American Board of Family Medicine*, vol. 20, no. 2, pp. 144–150, 2007.

[10] WHO Global estimates on prevalance of hearing loss, Mortality and Burden of Diseases and Prevention of Blindness and Deafness WHO, 2012, http://www.who.int/pbd/deafness/WHO_GE_HL.pdf.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 916356, 6 pages
http://dx.doi.org/10.1155/2015/916356

Clinical Study

Automatic Prosodic Analysis to Identify Mild Dementia

Eduardo Gonzalez-Moreira,[1] Diana Torres-Boza,[1] Héctor Arturo Kairuz,[1] Carlos Ferrer,[1] Marlene Garcia-Zamora,[2] Fernando Espinoza-Cuadros,[3] and Luis Alfonso Hernandez-Gómez[3]

[1]*Center for Studies on Electronics and Information Technologies, Universidad Central "Marta Abreu" de Las Villas, 54830 Santa Clara, Cuba*
[2]*Center for Elderly Adults No. 2, 54830 Santa Clara, Cuba*
[3]*Signal, Systems and Radiocommunications Department, Universidad Politécnica de Madrid, 28040 Madrid, Spain*

Correspondence should be addressed to Eduardo Gonzalez-Moreira; moreira@uclv.edu.cu

Received 22 January 2015; Accepted 10 June 2015

Academic Editor: Markus Hess

This paper describes an exploratory technique to identify mild dementia by assessing the degree of speech deficits. A total of twenty participants were used for this experiment, ten patients with a diagnosis of mild dementia and ten participants like healthy control. The audio session for each subject was recorded following a methodology developed for the present study. Prosodic features in patients with mild dementia and healthy elderly controls were measured using automatic prosodic analysis on a reading task. A novel method was carried out to gather twelve prosodic features over speech samples. The best classification rate achieved was of 85% accuracy using four prosodic features. The results attained show that the proposed computational speech analysis offers a viable alternative for automatic identification of dementia features in elderly adults.

1. Introduction

Dementia is a disorder characterized by an impairment of intellectual and communicative functioning, with high impact among elderly people. Usually this disorder leads to dependency of the patient on their families or caregivers due to the impossibility to carry out their daily tasks. A general agreement of the experts in the field revealed that the number of patients with dementia is increasing around the world due to a progressive aging society [1].

With this fact in mind the early detection of the dementia syndrome becomes an important goal to slow down the development of cognitive deterioration, allowing either the use of alternative nonpharmacological therapies or short periods of pharmacological treatments. Current methods of dementia screening in older adults involve structured interviews. Questionnaire tests such as the Mini-Mental State Examination (MMSE) [2], Clinical Dementia Rating (CDR) [3], or Memory Impairment Screen (MIS) [4] are commonly used. These methods typically rely on prolonged interviews with the patient and a family member. Therefore, an automated method for screening of dementia is highly desirable.

Due to the fact that one of the most significant areas affected by dementia is language, many researches have been oriented towards speech analysis, showing that language impairment is strongly related to cognitive impairment. Even more, the first clues start to appear some years before patient is clinically diagnosed [5, 6].

In this paper we propose a framework that applies speech signal analysis to identify mild dementia (MD). In contrast to previous works in automating the evaluation of cognitive impairment through speech analysis that relied on manual transcripts of audio recording, our system uses a novel method for automatically detecting syllable nuclei in order to measure prosodic features without the need for a transcription. In the next sections we present an overview of our data, followed by the description of the feature extraction procedure we propose and the classification technique used to determine whether the subject has mild dementia or not.

2. Methods

2.1. Experimental Subjects. Within the framework of this exploratory work, speech recordings were conducted at the Center of Elderly Adult #2 in Santa Clara, Cuba. A total of twenty subjects were selected for this pilot experiment from a group of candidates. Our sample comprises participant older than sixty years old with a diagnosis of mild dementia and healthy subjects. Other inclusion criteria were basic reading skills and no significant visual impairments. All the work was performed strictly following the ethical consideration of the center and the participants were notified about the research and their agreement obtained. Table 1 shows demographic data, with no significant differences between groups in terms of gender, age, or years of education.

Table 1: Participant demographic data.

Item	Group	
	MD	Non-MD
Number of patients	10	10
Male	6	9
Female	4	1
Average age (years)	80.3	78.9
Average years of education (years)	4.9	7.8

2.2. Recording Tools and Procedures. Also trying to make the process of speech recording the less annoying as possible for a daily clinical practice, a specific tool was developed. It consisted of the use of a standard laptop equipped with two headworn condenser C520L AKG microphones for capturing both clinician and patient voices. Each microphone was connected to a different channel (left or right) of an M-audio ADC device connected to the laptop through a USB port. This configuration provides some acoustic separation, despite no complete isolation, of patient and clinician voices, thus making it easier for their processing.

A specific software DCGrab v3.0 was created by the authors to allow an easy recording of the audio signals during each one of the parts defined in our recording protocol. The speech sound was recorded in stereo format with 16 bits of resolution and 44.1 KHz of sampling rate. The DCGrab v3.0 software also allows storing clinical data and demographic information for each patient (see Figure 1).

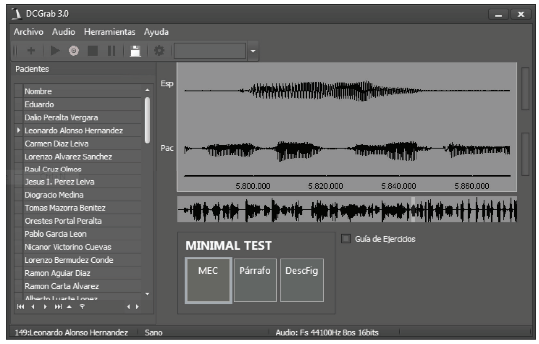


Figure 1: DCGrab v3.0 software.

2.3. Protocol and Speaking Styles. Two major conflicting criteria were considered for the design of the recording protocol in our database. On one side, it should represent the minimum possible burden on the busy schedules of daily clinical practices, while on the other side it should collect the richest variety of speaking styles which can result in a notable increase in testing time.

Consequently we decided to design a protocol consisting of two sequential parts. During the first part we recorded the speech productions from both clinician and patient during structured interviews commonly used in clinical assessment procedures. More specifically we considered the *Mini Examen Cognoscitivo* (MEC) which is the Spanish version of the Mini-Mental State Examination (MMSE) [2]. The MEC evaluation is our gold standard to classify each participant and evaluate the automatic method proposed in this work.

The second part of the interview consists of asking each enrolled subject to read a Spanish version of the paragraph “The Grandfather Passage” [7]: *¿Tú quieres saber todo acerca de mi abuelo? Pues bien, él tiene casi 93 años de edad y aún conserva su mente tan brillante como siempre. Él usa un viejo saco negro al que generalmente le faltan algunos botones. Su barba larga y colgante produce en quienes lo observan un profundo sentimiento de máximo respeto. Cuando el habla, su voz es un poco rajada y tiembla ligeramente. Dos veces al día toca un pequeño órgano con excelencia y virtuosismo. Cada día él hace unas caminatas cortas al aire fresco, excepto en el invierno cuando la nieve o el hielo lo impide. Nosotros siempre lo exhortamos a que camine más y que fume menos, pero él siempre responde “Aceite de plátano”. Al abuelo le gusta ser moderno en su lenguaje.* It is a short reading passage that has evolved into a ubiquitous metric of reading ability and speech intelligibility.

2.4. Prosodic Analysis. To obtain the prosodic features of speech recordings by means of automatic prosodic transcription, a novel algorithm to automatically detect syllable nuclei was used. The proposed algorithm is mainly based on the method described in [8] for speech rate detection.

The overall procedure is illustrated in Figure 2. The input speech signal is processed in parallel to obtain an automatic estimation of both syllable nuclei and fundamental frequency (F0 detection). In order to increase the temporal resolution of the energy envelope, the downsampling process is removed; also smaller windows size (10 ms) and overlap (5 ms) are used in the temporal weighting stage. Then, the syllabic nuclei are detected using the same threshold mechanism described in the peak counting stage in [8].

The vowel nucleus is the place where the pitch estimation reaches a local maximum and this phenomenon is relative to the syllable boundaries because simultaneous changes of intensity, spectral energy distribution, and voicing partially hide the perception of the pitch changes [9]. This feature is more evident for stop and fricative consonants and less significant for liquids and nasals [10]. Consequently the edges of the syllabic nucleus are more suitable for the detection of

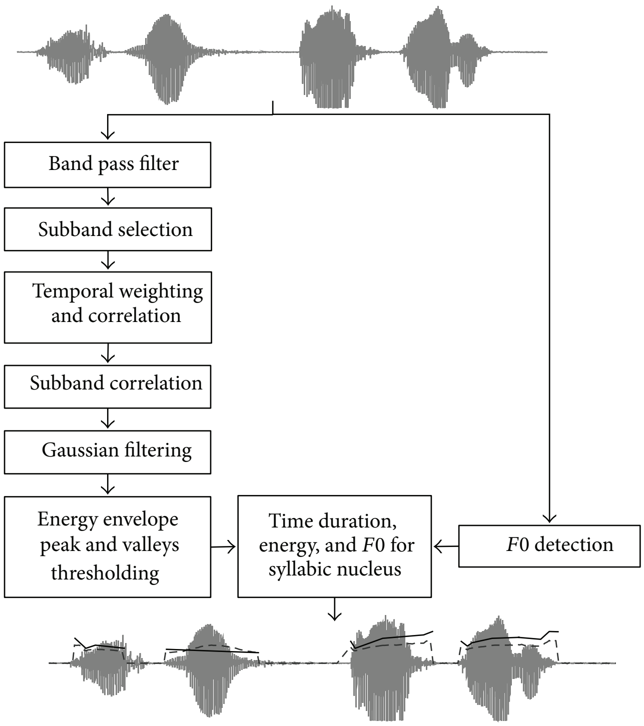


Figure 2: Algorithm block diagram for features set (adapted from [8]).

noticeable changes in the fundamental frequency than the syllable border. The boundaries of the syllable nucleus are estimated by the nearest minimum related to the detected peak on the energy envelope or by vocal activity limits, provided by the robust algorithm for pitch tracking (RAPT) [11].

For each syllable nucleus obtained, a number of features related to measures of intensity, duration, and fundamental frequency are estimated (see central bottom block in Figure 2). Duration and fundamental frequency features are given in milliseconds (ms) and semitones (ST), respectively. Expressing fundamental frequency in semitones diminishes gender differences as suggested in [12, 13].

For our research twelve prosodic features were calculated based on the syllable nucleus position obtained by the novel prosodic method as follows:

(1) speech time (SPT): total speech time from first syllable to last syllable produced,

(2) number of pauses (NPU): total number of silences; a gap between two consecutive syllables over 0.3 s that was considered silence,

(3) proportion of pause (PPU): total number of pauses over 0.3 s divided by the total amount of time spent speaking expressed in seconds,

(4) phonation time (PHT): total time of all syllables produced plus silences lower than 0.3 s,

(5) proportion of phonation (PPH): total time of all syllables produced plus silences lower than 0.3 s divided by the amount of time,

(6) speech rate (SPR): total number of syllables produced in a given speech sample divided by the amount of total time (including pause time),

(7) articulation rate (ARR): total number of syllables produced in a given speech sample divided by the amount of time taken to produce them in seconds,

(8) number of syllables (NSY): total number of syllables produced along the speech sample,

(9) mean of syllables duration (MSD): mean time of all syllables in seconds,

(10) standard deviation of F0 (SDF): standard deviation of fundamental frequency of all syllables produced along speech sample,

(11) maximum variation of F0 (MVF): maximum difference between higher and lower values of fundamental frequency along speech sample,

(12) mean of F0 (MFF): mean of fundamental frequency of all syllables produced.

2.5. Automatic Classification. Many classification techniques have been developed with remarkable performance in the last decades [14]. Since the main goal of this research is to find a set of prosodic parameter with discriminative potential for identifying mild dementia a well-known classifier was selected. In [15] we explored the use of Random Forest for a similar task, but now we have found that better results can be achieved using the Support Vector Machine (SVM) classification technique. Therefore our automatic classification of reading speech is based on SVM technique to evaluate how well the proposed features predicted participant's group membership. The results are evaluated in terms of accuracy (*Accu*), sensitivity (*Sens*), and specificity (*Spec*) measurements [16]. A cross validation technique was used to avoid overfitting, that is, a discriminant function to be created with the same data used later for testing. Specifically the leave-one-out method of cross validation was applied. It involves generating the discriminant function on all but one of the participants ($n-1$) and then testing for group membership on that sample. The process is repeated for each sample (n times) and the percentage of correct classifications is generated through averaging for the n trials. In our case n is equal to the total number of participants ($n = 20$), and, in each iteration of the cross validation method, one fold is used for testing and the other nineteen folds are used for training the SVM classifier.

3. Results

The goal of our experiments was to evaluate the potential of selected features for automatic measurement of the impairment cognition through prosodic analysis. Table 2 contains descriptive statistics for these measures showing the mean, standard deviation and range for every prosodic measure on both groups (MD and non-MD).

Visual inspection reflects that probable differences between subjects with mild dementia and healthy subjects

Table 2: Prosodic features for mild dementia patients (MD) and healthy controls (non-MD).

Features	MD		Non-MD	
	Mean (SD)	Range	Mean (SD)	Range
SPT	156.4 (56.4)	75.0–234.2	127.9 (64.6)	59.2–249.7
NPU	81.3 (38.2)	32–132.0	62.1 (33.7)	27.0–124.0
PPU	54.7 (16.6)	30.1–78.2	52.2 (10.8)	30.7–64.8
PHT	63.2 (13.4)	50.3–90.1	55.3 (16.9)	33.5–107.2
PPH	45.2 (16.6)	21.7–69.8	47.7 (10.8)	35.1–69.2
SPR	2.1 (0.8)	1.2–3.8	2.3 (0.5)	1.6–3.6
ARR	4.8 (0.7)	3.6–5.9	4.8 (0.6)	4.3–6.3
NSY	306.6 (67.6)	225–449	266.8 (75.1)	133–401
MSD	0.1 (0.0)	0.0–0.1	0.1 (0.0)	0.0–0.1
SDF	42.0 (13.5)	24.5–68.6	29.3 (8.7)	17.6–46.6
MVF	279.3 (80.2)	143–377	232.9 (109.2)	99–375
MFF	174.4 (40.6)	108.0–219.9	138.2 (27.7)	105.9–192.3

Table 3: Statistic analysis results.

Features	Kolmogorov-Smirnov		
	h	p values	Ranking
SPT	0	0,675	7
NPU	0	0,675	8
PPU	0	0,675	9
PHT	0	0,312	3
PPH	0	0,675	10
SPR	0	0,312	4
ARR	0	0,675	11
NSY	0	0,312	5
MSD	0	0,974	12
SDF	1	0,030	1
MVF	0	0,312	6
MFF	1	0,032	2

could be found on some measures like SPT, NPU, NSY, SDF, and MFF. Other measures like PPU, PPH, ARR, and MSD show no significant differences suggesting that these measures have no power to discriminate between classes.

Hence, we performed a Kolmogorov-Smirnov test (KS-test) to determine which set of features can significantly discriminate between the two groups (MD and non-MD) [17]. The KS-test has the advantage of making no assumption about the distribution of data. This nonparametric test for the equality of continuous, one-dimensional probability distributions can be used to compare a sample with a reference probability distribution (one-sample KS-test) or to compare two samples (two-sample KS-test). The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case). The two-sample test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples [18]. Nevertheless with a KS-test, we cannot guarantee finding the best set of features to reach the maximum performance of the SVM classifier but we believe it can provide an acceptable first approximation to it. Statistical analyses were conducted using IBM SPSS v21 [19], and the desired significance level of 0.05 was used.

Table 4: Level of significance based on p values for prosodic features.

Significant (SIG)	Possibly significant (PSIG)	Nonsignificant (NSIG)
SDF	PTH	SPT
MFF	SPR	NPU
	NSY	PPU
	MVF	PPH
		ARR
		MSD

Table 5: SVM classification results for significant group combination.

Group	Accu	Sens	Spec
SIG	**75.0**	**72.7**	**77.7**
PSIG	35.0	36.3	33.3
NSIG	45.0	46.1	42.8
SIG-PSIG	60.0	58.3	62.5
SIG-NSIG	65.0	66.7	63.6
PSIG-NSIG	30.0	33.3	25.0
SIG-PSIG-NSIG	65.0	61.5	71.4

Table 6: Summary of dataset size and classification accuracies reported in previous works.

Previous works	Participants (patients)	Classification accuracies
Lehr et al. [20]	72 (35)	75.4%–81.5%
Thomas et al. [21]	85 (50)	58.8%–75.3%
Bucks et al. [22]	24 (8)	87.5%

Summary of results of the hypothesis (h), the p values for KS-test, and a ranking of level of significance for each feature based on the lower p values are shown in Table 3. Despite the small sample used, the ranking attempt illustrates the relative importance of these variables for discriminating between healthy and mild dementia.

Using this information three sets of features (presented in Table 4) were defined according to three different levels of discrimination. The first set included those features with significant differences ($p < 0.05$) between both groups (SDF and MFF). The second set contained features with slight differences ($p < 0.5$) between MD and non-MD participants, for example, PTH, SPR, NSY, or MVF. Features with no significant differences ($p > 0.5$) between groups: SPT, NPU, PPU, PPH, ARR, and MSD were included in the third set. We think that the small sample size may have resulted in this lack of significance and that these temporal measures may yet offer additional explanatory power.

Table 7: SVM classification results for best features combination.

Features	Accu	Sens	Spec
SDF	75.0	69.2	85.7
PHT-MFF	75.0	72.7	77.7
NPU-PHT-MFF	80.0	80.0	80.0
ARR-MSD-SDF-MFF	**85.0**	**81.8**	**88.8**
NPU-NSY-SDF-MVF-MFF	80.0	80.0	80.0
NPU-PPH-NSY-SDF-MVF-MFF	80.0	80.0	80.0
NPU-PPU-PPH-NSY-SDF-MVF-MFF	80.0	80.0	80.0
SPT-NPU-PPU-PPH-NSY-SDF-MVF-MFF	80.0	80.0	80.0
SPT-NPU-PPU-PHT-PPH-SPR-ARR-NSY-MFF	75.0	72.7	77.7
NPU-PPU-PHT-PPH-ARR-NSY-MSD-SDF-MVF-MFF	70.0	66.6	75.0
SPT-NPU-PPU-PHT-PPH-SPR-ARR-NSY-MSD-SDF-MFF	70.0	66.6	75.0
SPT-NPU-PPU-PHT-PPH-SPR-ARR-NSY-MSD-SDF-MVF-MFF	65.0	61.5	71.4

SVM was carried out to determine how well the proposed groups of level of significance predicted participants' group membership. Classification results were obtained using different combinations of the prosodic features included in the three significance groups summarized in Table 4.

As it can be seen in Table 5, using this classification strategy, the best accuracy of 75.0% was achieved using only the set of features in the significant group. While interpreting the results of Table 5, we should note that the classification algorithm did not take advantage from increasing the number of features. Even so these results could be considered relatively good based on the size of the data and published results elsewhere [20–22] as depicted in Table 6.

The former method of feature selection is one of the statistical methods frequently used in similar studies to evaluate the level of significance for measurement under analysis [23, 24]. Due to the nature of this method (KS-test over individual features) it cannot guarantee maximum accuracy for a classifier able to model complementary information between features. The set of features that could represent in a better way both classes (MD and non-MD) cannot be determined by the level of significance of individual features. Even the combination of different features can either improve or worsen the final classification. Consequently one critical question arises and still needs to be answered: how to choose the set of features to able to reach the maximum classification rate for a given classification technique. Trying to answer this question, in the next step, we proceeded with the evaluation of the SVM using all the possible combination of the features obtained from the prosodic analysis. In our work a total of 4094 combinations from 12 features without repetition were tested to find the best feature set. The best classification rates, using leave-one-out cross validation, for each combination of a features amount are shown in Table 7.

Results in Table 6 indicate that the relation between features and the way each affects others in the pattern classification is not well known. The highest accuracy value was obtained for the combination of four features, including the best (SDF) and worse (MSD) ranked in Table 3. We also should note that increasing the number of features does not guarantee an increment in the classification rate.

4. Conclusions

The main goal of this pilot study was to investigate the potential use of automatically extracted prosodic features in the diagnosis of mild dementia in elderly adults. The results demonstrate the existence of significant measureable prosodic differences between the performance of healthy participants and patients with mild dementia in reading speech. Features like ARM, MSD, SDF, and MFF were identified as having a higher discriminative power. Furthermore, due to the relative simple and low cost methodology, the technique for the screening of moderate cognitive impairment is easy to spread out. This study lays the foundation for future research using a larger number of participants and other speech features either in time or in spectral and cepstral domain. In this way, definitive conclusions of prosodic analysis to identify mild dementia could be drawn.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported in part by the Spanish-funded research project "CMC-V2: Caracterización, Modelado y Compensación de Variabilidad en la Señal de Voz" (Grants no. TEC2012-37585-C02-02) and the Cuban-funded research project "Herramienta de Apoyo al Diagnóstico Médico del Deterioro Cognitivo Mediante el Procesamiento Digital de Voz." The authors would like to thank Professors Julian Cardenas and Anja Lowit for their valuable comments.

References

[1] J. J. Llibre, "Aging and dementia: implications for the scientist community, public health and Cuban society," *Revista Anales de la Academia de Ciencias de Cuba*, vol. 2, no. 2, pp. 36–54, 2012.

[2] M. F. Folstein and S. E. Folstein, "'Mini-mental state'. A practical method for grading the cognitive state of patients for the

clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[3] J. C. Morris, "The clinical dementia rating (CDR): current version and scoring rules," *Journal of Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993.

[4] H. Buschke, G. Kuslansky, M. Katz et al., "Screening for dementia with the Memory Impairment Screen," *Neurology*, vol. 52, no. 2, pp. 231–238, 1999.

[5] V. Deramecourt, F. Lebert, B. Debachy et al., "Prediction of pathology in primary progressive language and speech disorders," *Neurology*, vol. 74, no. 1, pp. 42–49, 2010.

[6] M. Mesulam, A. Wicklund, N. Johnson et al., "Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia," *Annals of Neurology*, vol. 63, no. 6, pp. 709–719, 2008.

[7] F. L. Darley, A. E. Aronson, and J. R. Brown, *Motor Speech Disorders*, WB Saunders, Philadelphia, Pa, USA, 3rd edition, 1975.

[8] D. Wang and S. S. Narayanan, "Robust speech rate estimation for spontaneous speech," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2190–2201, 2007.

[9] D. House, "The influence of silence on perceiving the preceding tonal contour," in *Proceedings of the International Congress of Phonetic Sciences*, vol. 13, pp. 122–125, 1995.

[10] P. Mertens, "The prosogram: semi-automatic transcription of prosody based on a tonal perception model," in *Proceedings of the International Conference on Speech Prosody*, Nara, Japan, March 2004.

[11] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds., Elsevier Science, 1995.

[12] J. t'Hart, "Psychoacoustic backgrounds of pitch contour stylization," *IPO Annual Progress Report*, vol. 11, no. 1, pp. 11–19, 1976.

[13] J. t'Hart, "Differential sensitivity to pitch distance, particularly in speech," *The Journal of the Acoustical Society of America*, vol. 69, no. 3, pp. 811–821, 1981.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2nd edition, 2001.

[15] F. Espinoza-Cuadros, M. A. Garcia-Zamora, D. Torres-Boza et al., "A spoken language database for research on moderate cognitive impairment: design and preliminary analysis," in *Advances in Speech and Language Technologies for Iberian Languages*, vol. 8854 of *Lecture Notes in Computer Science*, pp. 219–228, Springer, 2014.

[16] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, 1978.

[17] F. J. Massey Jr., "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[18] L. M. Schultz, "P. Sprent & N.C. Smeeton (2007). Applied nonparametric statistical methods (4th ed.)," *Psychometrika*, vol. 75, no. 3, pp. 579–580, 2010.

[19] M. J. Norušis, *SPSS 13.0 Advanced Statistical Procedures Companion*, Prentice Hall, Englewood Cliffs, NJ, USA, 2004.

[20] M. Lehr, E. Prudhommeaux, I. Shafran, and B. Roark, "Fully automated neuropsychological assessment for detecting mild cognitive impairment," in *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH '12)*, Portland, Ore, USA, September 2012.

[21] C. Thomas, V. Kešelj, N. Cercone, K. Rockwood, and E. Asp, "Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '05)*, vol. 3, pp. 1569–1574, August 2005.

[22] R. S. Bucks, S. Singh, J. M. Cuerden, and G. K. Wilcock, "Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance," *Aphasiology*, vol. 14, no. 1, pp. 71–91, 2000.

[23] S. V. S. Pakhomov, S. E. Marino, and A. K. Birnbaum, "Quantification of speech disfluency as a marker of medication-induced cognitive impairment: an application of computerized speech analysis in neuropharmacology," *Computer Speech & Language*, vol. 27, no. 1, pp. 116–134, 2013.

[24] E. Gonzalez-Moreira, D. Torres-Boza, M. A. Garcia-Zamora, C. A. Ferrer, and L. A. Hernandez-Gomez, "Prosodic speech analysis to identify mild cognitive impairment," in *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014*, vol. 49 of *IFMBE Proceedings*, pp. 580–583, Springer, Berlin, Germany, 2015.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 704121, 9 pages
http://dx.doi.org/10.1155/2015/704121

Clinical Study

Acoustic Correlates of Compensatory Adjustments to the Glottic and Supraglottic Structures in Patients with Unilateral Vocal Fold Paralysis

Luis M. T. Jesus,[1,2] Joana Martinez,[1] Andreia Hall,[1,3] and Aníbal Ferreira[4]

[1] *Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal*
[2] *School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal*
[3] *Department of Mathematics (DMat), University of Aveiro, 3810-193 Aveiro, Portugal*
[4] *Department of Electrical and Computer Engineering, University of Porto, 4200-465 Porto, Portugal*

Correspondence should be addressed to Luis M. T. Jesus; lmtj@ua.pt

Received 22 January 2015; Accepted 24 April 2015

Academic Editor: Haldun Oguz

The goal of this study was to analyse perceptually and acoustically the voices of patients with Unilateral Vocal Fold Paralysis (UVFP) and compare them to the voices of normal subjects. These voices were analysed perceptually with the GRBAS scale and acoustically using the following parameters: mean fundamental frequency ($F0$), standard-deviation of $F0$, jitter (ppq5), shimmer (apq11), mean harmonics-to-noise ratio (HNR), mean first ($F1$) and second ($F2$) formants frequency, and standard-deviation of $F1$ and $F2$ frequencies. Statistically significant differences were found in all of the perceptual parameters. Also the jitter, shimmer, HNR, standard-deviation of $F0$, and standard-deviation of the frequency of $F2$ were statistically different between groups, for both genders. In the male data differences were also found in $F1$ and $F2$ frequencies values and in the standard-deviation of the frequency of $F1$. This study allowed the documentation of the alterations resulting from UVFP and addressed the exploration of parameters with limited information for this pathology.

1. Introduction

A neural dysfunction of the larynx leads to alterations in voice, respiration, and airway protection. Usually, Unilateral Vocal Fold Paralysis (UVFP) is related to a set of well-documented perceptive alterations such as weak voice, breathiness, roughness, diminished voice intensity, vocal effort, low voice efficiency, voice breaks, diplophonia, and air loss [1–5]. Furthermore, vocal strain is a critical component in various vocal pathologies including UVFP. A neuronal dysphonia, such as UVFP, can alter the vibrational patterns of the Vocal Folds (VF) which leads to compensatory adjustments to the glottic and supraglottic structures that increase the vocal effort and vocal strain perception [6, 7]. In addition to the perceptive alterations, UVFP also results in higher values of jitter and shimmer and lower values of the harmonics-to-noise ratio (HNR) [1–4, 8]. Furthermore, values of standard-deviation of fundamental frequency ($F0$) are reported as higher than normal because of the diminished control of the vibrational pattern of the VF, causing greater variability [9–11]. According to Schwarz et al. [6], there is a need to describe and understand the UVFP patient's larynx configuration for a better and more individualised vocal intervention, preventing compensatory adjustments. Formant frequencies provide acoustic cues about the vocal tract configuration [12–14]. According to Lee et al. [15] the formant's values are relevant for discriminating normal from pathologic voices and the configuration of the vocal tract is different during phonation in people with vocal pathologies. The same authors [15] found slightly lower values of the first formant ($F1$) frequency and higher values of the second formant ($F2$) frequency in cases of UVFP. This indicates that UVFP subjects tend to have a more elevated and advanced tongue position during phonation [13, 14]. A breathy voice (common in UVFP) is reported to be associated with the same configuration referred to previously [16]. However, Titze [13] reports an approximation of the values of the frequency of $F1$ and $F2$ in cases of narrower vocal tract. These vocal tract modifications may result from the attempt to compensate the vocal alteration by patients exhibiting UVFP [2]. According to Lee et al. [15] the standard-deviations of the frequency of $F1$ and $F2$ have higher values in cases of UVFP indicating a higher instability of the vocal tract configuration during phonation.

The aim of this study was to compare perceptually and acoustically the voices of subjects with UVFP and the voices of subjects representing normal quality. Measures related to the vocal tract configuration, namely, formant frequencies, were also analysed and correlated with alterations caused by vocal pathology.

2. Materials and Methods

This is a quantitative, descriptive, and cross-sectional study [17–19]. The recordings were made in Hospital de Santo António and Hospital de São João, both in Porto, Portugal, and at the Speech, Language, and Hearing Laboratory (SLH-lab) at the University of Aveiro, Portugal. This took place as part of the data collection process of the first representative European Portuguese pathological voice database [20]. Part of this data was divided into two groups: a group having vocal pathology (UVFP) and a group without vocal pathology. A group of 17 patients, evaluated with videolaryngoscopy and diagnosed with UVFP, formed the pathologic group. The inclusion criteria for this group were having diagnosis of UVFP, not having had speech and language therapy intervention, and being over 18 years old. The exclusion criteria were having other concomitant pathologies to UVFP and/or having been submitted to a surgical intervention to correct the vocal pathology. A group of 85 normal voice volunteers were included in the control group based on two distinct procedures: 43 subjects were evaluated with videolaryngoscopy and diagnosed as normal; 42 subjects were evaluated using a vocal anamnesis and summative evaluation (a similar procedure was used by Roark et al. [21]). The inclusion criteria for the control group were having normal voice quality and being over 18 years old. The exclusion criterion was having vocal or other pathologies that may interfere with normal voice production.

Each pathologic case was individually matched to five subjects of the control group in order to increase the power of statistical tests [17, 22]. The cases were matched according to gender and age. The first variable was gender because after puberty there is a set of different characteristics that differentiate male and female voices [23]. The second variable was age because with aging some functional and structural modifications occur at phonatory level [23, 24]. Taking into account the fact that there are notable voice changes if the subjects' age difference is more than 10 years [25–29] the maximum allowed difference of age between the matched subjects was 5 years, in an attempt to reduce variability.

Four (4) subjects with UVFP were male (23.5%) and 13 subjects were female (76.5%). The youngest patient was 30 years old and the oldest 72. The mean age for the pathologic group was 56.7 years with a standard-deviation of 12.7 years.

Table 1: Values of the autocorrelation method used in *Praat* for the voice analysis.

Parameter	Value
Maximum number of candidates	15
Silence threshold	0.03
Voicing threshold	0.45
Octave cost	0.15
Octave-jump cost	0.35
Voiced/unvoiced cost	0.14

Nine (9) patients had left UVFP (52.9%) and 8 right UVFP (47.1%). In the control group 20 subjects were male (23.5%) and 65 were female (76.5%). The mean age of the control group was 56.1 years and the standard-deviation was 12.7 years.

The voice recordings were made in a clinical setting using *Praat 5.3.56 (32-bit edition)* [30]. A Behringer ECM8000 microphone and a Presonus AudioBox USB (16 bits and 48000 Hz) were used for all of the recordings. The subjects were seated and the microphone was aligned to the mouth at a distance of 30 cm [31, 32]. An informed consent was signed and the vowel [a] was recorded. A parcel of the vowel was then annotated according to criteria defined by Pinho et al. [3]: 200 ms after the onset of phonation and with approximately 100 cycles. This parcel was then manually analysed with *Praat 5.3.56 (64-bit edition)* with an autocorrelation method (used by default by the software) to estimate $F0$. There were some errors in the identification of the period, so a modification of the "octave cost" to a higher value was made (as suggested in *Praat*'s manual). The values of the parameters used to run the autocorrelation method are presented in Table 1.

From the "voice report" *Praat* window the following values were extracted: mean $F0$; standard-deviation of $F0$; jitter (ppq5); shimmer (apq11); mean harmonics-to-noise ratio (HNR). The Burg [33] method (used by default by *Praat*) was used to track the formants. The "formant listing" for the same 100 cycles was obtained and the mean value and standard-deviation were calculated for the frequency of $F1$ and $F2$. The values were double-checked through the spectrogram of each segment.

Each voice was also perceptually assessed using the GRBAS scale [34]. For the pathologic voices a group of five speech and language therapists with expertise in voice assessment made the perceptive evaluation. For the normal voices one speech and language therapist made the perceptive assessment. For these procedures the experts used the following headphones connected to the internal soundcard of a laptop computer: Sennheiser HD 380 Pro; Sennheiser HD201; Sony MDR-CD270; Sony MDRZX100B; Sony MDR-ZX110NA. All of the assessments were made blindly regarding the group (patients or normal subjects).

For the statistical analysis *IBM SPSS Statistics version 20* was used. The interrater consistency was analysed using the Kendall W Coefficient. The Mann-Whitney U test was used to analyse the GRBAS scale parameters. The acoustic parameters that had normal distribution (HNR, $F2$♀, standard-deviation of $F0$♂, $F1$♂) were statistically analysed using the

Table 2: Interrater consistency—Kendall's W test.

Scale parameter	W	p value
G	0.263	0.001
R	0.160	0.033
B	0.381	<0.001
A	0.344	<0.001
S	0.438	<0.001

G: Grade; R: Rough; B: Breathy; A: Asthenic; S: Strained; W: Kendall's W.

t-test and parameters that did not have normal distribution (*Jitter* (ppq5), *Shimmer* (apq11), *F*0♀, standard-deviation of *F*0♀, *F*1♀, standard-deviation of *F*1♀, standard-deviation of *F*2♀, *F*0♂, standard-deviation of *F*1♂, *F*2♂, and standard-deviation of *F*2♂) were analysed with the Mann-Whitney U test. The normality was tested with the Shapiro-Wilk test. A level of significance of 0.05 was used for all statistical analyses.

All of the procedures had the acceptance of the Ethical Commission of the Hospital de Santo António and Hospital de São João. An authorisation from the National Commission for Data Protection was also obtained.

3. Results and Discussion

3.1. Interrater Consistency. The consistency between the five judges that assessed the pathologic voices was analysed using Kendall's W test. Table 2 shows that there is consistency in all of the parameters of the GRBAS scale between judges. The fact that the judges presented consistency between them indicates that they have a similar internal understanding of the used instrument [35]. This consistency is likely to be related to the fact that the GRBAS scale is widely used, understood, and recommended worldwide by clinicians [36]. The W's value, shown in Table 2, can vary between 0 (no general tendency of consistency between judges) and 1 (all judges responded equally) [37]. In Table 2 we can also see that the lowest value of W was found for the R (Rough) parameter. This may be due to the fact that this parameter is a supraclass of perceptive parameters that can lead to various interpretations between different judges [38]. The fact that none of the parameters had a very good consistency was expected because the perceptive assessment is a very complex procedure that includes various subjective elements that are not totally understood [36, 39]. Despite the results varying from *reasonable* to *good*, perceptive evaluation is still a central procedure in the vocal assessment [40].

3.2. Comparison of GRBAS Scale Parameters between Normal and UVFP Voices. The results of the perceptive assessment of the voices of the normal and UVFP subjects were analysed using the Mann-Whitney U test. Table 3 shows that all of the GRBAS parameters were statistically different between groups, being higher in the pathologic group as expected (see Figure 1). The control group had a mean score of zero, which was expected because the control group was intended to have a normal/nonaltered voice quality that would be associated to a 0 value (normal) of all parameters assessed in GRBAS.

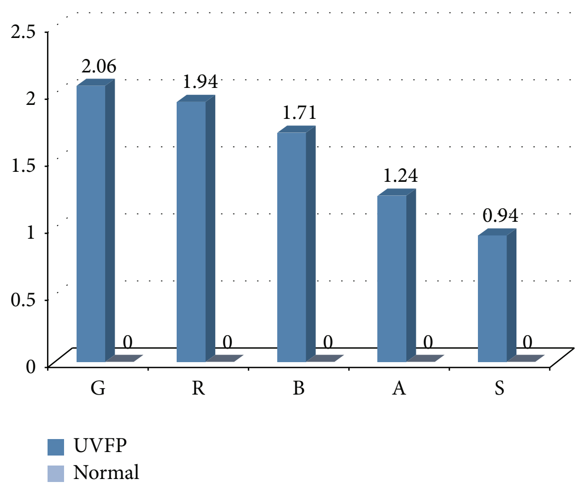


Figure 1: Comparison between mean scores of the GRBAS scale for normal and UVFP subjects.

In the pathologic group we can see that the parameter with the highest values was G (Grade), which has been observed before by other authors [41, 42]. In this group of UVFP there were alterations in all of the GRBAS parameters, varying between a mild and moderate grade of perturbation. Grade (G), Rough (R), and Breathy (B) presented the highest mean scores, as previously observed by various authors [1–4, 43]. Another disturbance that is commonly found in subjects with UVFP is a weak voice [1, 4, 44] which is reflected in parameter A (Asthenic), also found in this sample. In addition to the previous parameters, according to Rosenthal et al. [7], it is usual to find vocal strain (parameter S) in these cases, which could also be observed in this study.

One of the major alterations caused by UVFP is the incomplete glottal closure that originates excess air during phonation that creates a breathy voice (parameter B is altered) [2, 4, 45]. This air leakage leads to a lower voice energy originating a weak voice (parameter A is altered) [2, 4, 45]. The irregularity of the VF cycles (parameter R reflects this) is due to the reduced mobility/immobility of the paralysed VF or to the fact that the unhealthy VF may present a passive vibration [4, 46]. In some cases, in an attempt to overcome the alterations caused by the UVFP, patients create compensations that can lead to strain in the supraglottic region, increasing the vocal effort and giving the voice a strained characteristic (parameter S) [4, 7]. Grade (G) is related with the other parameters and varies according to the severity of the overall voice perturbation [47].

3.3. Comparison of Acoustic Parameters between Normal and UVFP Voices. Although perceptive assessment is the most used technique for vocal assessment, it is a subjective process that leads to some variability issues [8]. Contrary to this, acoustic data allows objective and noninvasive measures about the behaviour of the VF [8, 15, 48–50]. Table 4 shows statistically different values of jitter (ppq5), shimmer (apq11), and HNR between the normal and pathologic voices. Jitter, which is related to the absolute difference between the

Table 3: Comparison of the results of GRBAS scale between UVFP and normal voice subjects.

	UVFP		Normal		U	p value
	N	Mean ± SD	N	Mean ± SD		
G	17	2.06 ± 0.827	85	0	0	<0.001
R	17	1.94 ± 0.899	85	0	0	<0.001
B	17	1.71 ± 0.772	85	0	0	<0.001
A	17	1.24 ± 0.437	85	0	0	<0.001
S	17	0.94 ± 0.556	85	0	25.5	<0.001

UVFP: Unilateral Vocal Fold Paralysis; G: Grade; R: Rough; B: Breathy; A: Asthenic; S: Strained; N: number of cases; SD: standard deviation; U: Mann-Whitney U test.

Table 4: Comparison of jitter, shimmer, and HNR between normal and UVFP subjects.

	UVFP		Normal		t or U	p value
	N	Mean ± SD	N	Mean ± SD		
Jitter ppq5 (%)	17	1.06 ± 1.02	85	0.26 ± 0.18	$U = 249$	<0.001
Shimmer apq11 (%)	17	10.16 ± 3.34	85	7.14 ± 3.22	$U = 376$	0.001
HNR (dB)	17	10.11 ± 4.94	85	14.75 ± 4.46	$t = -3.85$	<0.001

UVFP: Unilateral Vocal Fold Paralysis; SD: standard deviation; t: t-test; U: Mann-Whitney U test.

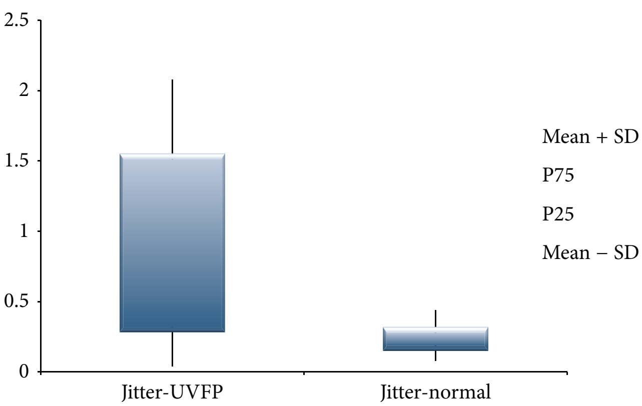


Figure 2: Jitter (%) values for UVFP and normal subjects.

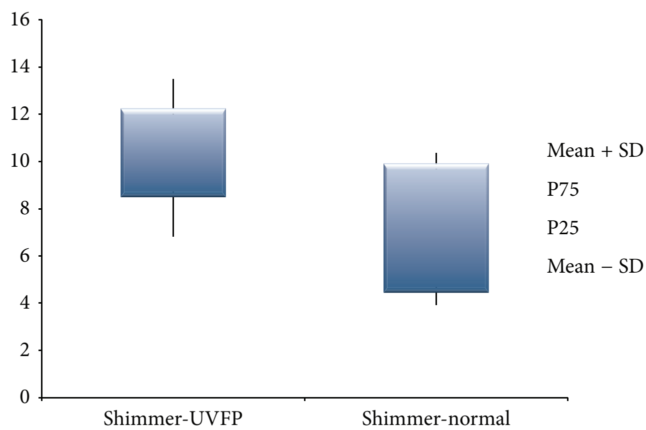


Figure 3: Shimmer (%) values for UVFP and normal subjects.

durations of consecutive cycles [43], is higher in UVFP subjects (see Figure 2). These results were also obtained by other authors [2, 3, 8]. These higher values may be due to the asymmetry at the VF level, caused by the UVFP that leads to vibration irregularities in frequency altering the jitter values [2]. Similarly shimmer, which is related to the absolute difference between the amplitudes of consecutive cycles [43], is also higher in UVFP cases (see Figure 3). These results were also obtained by other authors [2, 3, 8]. The asymmetry caused by UVFP leads to vibration irregularities in amplitude altering shimmer values [2]. This parameter is also increased by a poor and inconsistent contact between VF, which is very common in UVFP [51]. Thus, these UVFP subjects present more cyclic irregularity at frequency and amplitude level compared to the normal voice subjects. It should be noted that we also had higher than normal values of shimmer in the normal sample. This may be due to the fact that the recordings were made in a clinical setting that is not entirely noise-free and this could have interfered with the data calculation of this parameter. Regarding the HNR, which is obtained from the ratio between the harmonic and noise components of the signal [43], the results indicate a lower value in the pathologic group (see Figure 4). These results were consistent with the literature [2, 8]. The alterations in periodicity caused by the UVFP originate a lower ratio between the two components, diminishing the HNR values in the pathologic cases [2]. These results indicate that patients with UFVP have higher relative noise amplitude during phonation (than the normal subjects) lowering the HNR value.

The parameters presented in Tables 5 and 6 were divided by gender because females and males have different inherent vocal tract and VF characteristics, especially in terms of size and mass [52]. For $F0$ (see Figures 5 and 6) we can see that there are no significant statistical differences between pathologic and normal voices in both genders. This fact was also previously described by Oguz et al. [8]. Fundamental frequency is directly related to and dependent of length, tension, mass, rigidity, and the interaction with the subglottic

Table 5: Fundamental frequency ($F0$) and first and second formant frequencies ($F1$ and $F2$) and their standard-deviations, for normal and UFVP female participants.

♀	UVFP		Normal		t or U	p value
	N	Mean ± SD	N	Mean ± SD		
$F0$ (Hz)	13	218.38 ± 72.36	65	195.36 ± 33.02	U = 394	0.335
SD $F0$ (Hz)	13	6.65 ± 12.28	65	2.62 ± 2.15	U = 267	0.018
$F1$ (Hz)	13	826.16 ± 171.73	65	819.03 ± 164.75	U = 421	0.495
SD $F1$ (Hz)	13	117.18 ± 99.91	65	71.28 ± 53.54	U = 327	0.116
$F2$ (Hz)	13	1522.69 ± 96.00	65	1453.51 ± 139.20	t = 1.58	0.059
SD $F2$ (Hz)	13	156.31 ± 146.20	65	62.26 ± 48.67	U = 204	0.002

UVFP: Unilateral Vocal Fold Paralysis; N: number of cases; SD: standard deviation; t: t-test; U: Mann-Whitney U test; $F0$: fundamental frequency; SD $F0$: standard-deviation of the fundamental frequency; $F1$: first formant frequency; SD $F1$: standard-deviation of first formant frequency; $F2$: second formant frequency; SD $F2$: standard-deviation of second formant frequency.

Table 6: Fundamental frequency ($F0$) and first and second formant frequencies ($F1$ and $F2$) and their standard-deviations, for normal and UFVP male participants.

♂	UVFP		Normal		t or U	p value
	N	Mean ± SD	N	Mean ± SD		
$F0$ (Hz)	4	121.43 ± 12.70	20	128.27 ± 23.85	U = 39	0.485
SD $F0$ (Hz)	4	3.41 ± 1.25	20	1.36 ± 0.58	t = 5.27	<0.001
$F1$ (Hz)	4	821.48 ± 331.80	20	677.33 ± 84.95	t = 1.81	0.043
SD $F1$ (Hz)	4	191.91 ± 105.62	20	30.58 ± 18.84	U = 1	<0.001
$F2$ (Hz)	4	1629.09 ± 474.06	20	1282.95 ± 104.35	U = 11	0.011
SD $F2$ (Hz)	4	263.68 ± 144.45	20	36.89 ± 27.53	U = 3	0.001

UVFP: Unilateral Vocal Fold Paralysis; N: number of cases; SD: standard deviation; t: t-test; U: Mann-Whitney U test; $F0$: fundamental frequency; SD $F0$: standard-deviation of the fundamental frequency; $F1$: first formant frequency; SD $F1$: standard-deviation of first formant frequency; $F2$: second formant frequency; SD $F2$: standard-deviation of second formant frequency.

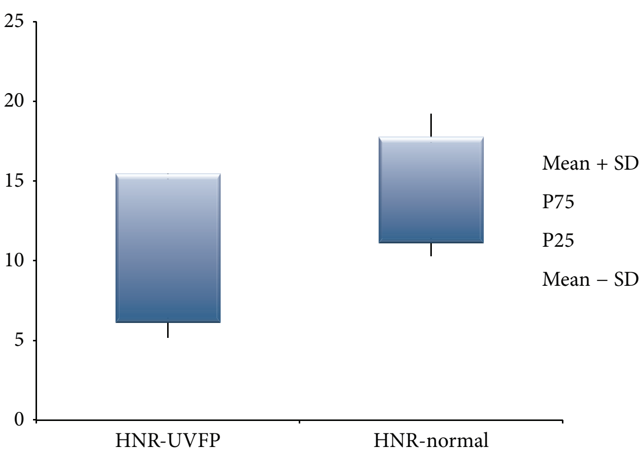


Figure 4: HNR (dB) values for UVFP and normal subjects.

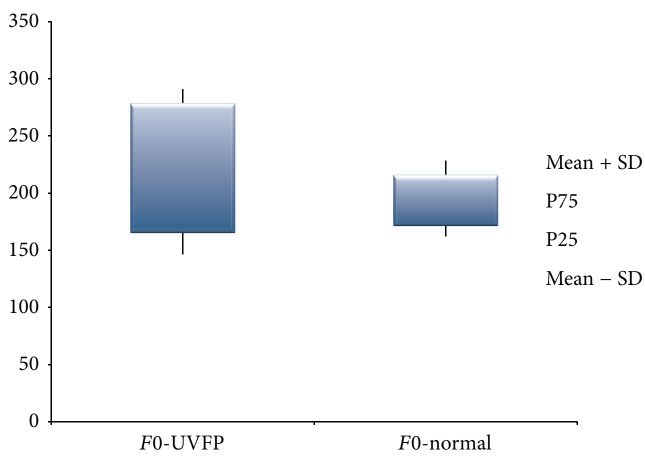


Figure 5: $F0$ (Hz) values for female UVFP and normal subjects.

pressure [53]. The fact that there are no differences between the two groups indicates that, in this sample, the modifications at VF level caused by UVFP are not sufficient to create real alterations in $F0$. Also according to Woo et al. [54] the majority of UVFP subjects present $F0$ values close to normal.

The standard-deviation of $F0$ (see Figures 7 and 8), which is related to the variations in vibration and muscular control of the VF, is higher in the pathologic group indicating important alterations in the described aspects [53]. Thus, subjects with UVFP present more $F0$ variability indicating a poorer muscular control and lower vibrational stability of VF. These results are supported by other authors [10, 11, 46, 53].

The vocal tract configuration interacts with VF oscillation; that is, vocal tract configuration constrains VF functioning during phonation [15, 55]. After the onset of UVFP patients usually develop some compensatory adjustments at glottic and supraglottic level altering voice and vocal tract configuration [6]. The description of vocal tract

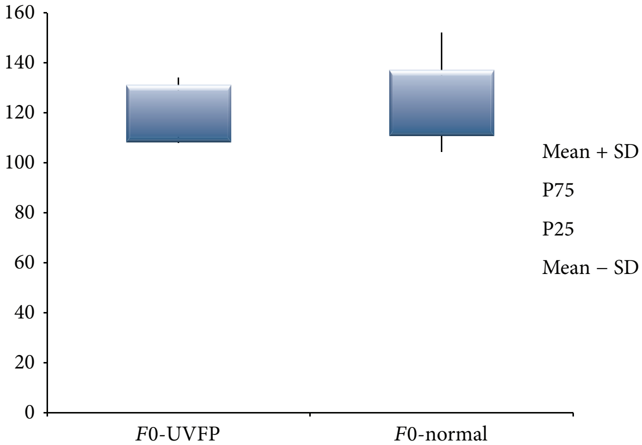


Figure 6: $F0$ (Hz) values for male UVFP and normal subjects.

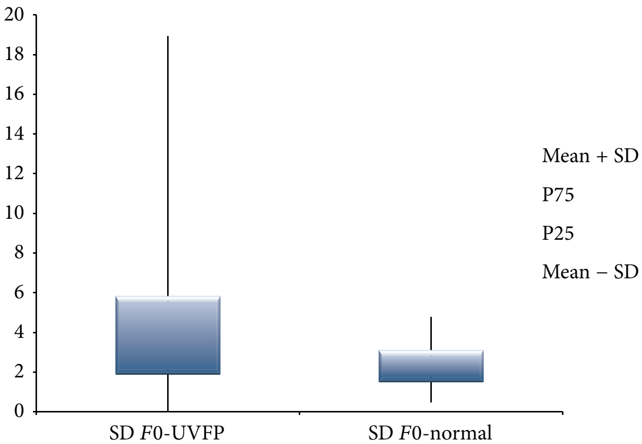


Figure 7: SD of $F0$ (Hz) values for female UVFP and normal subjects.

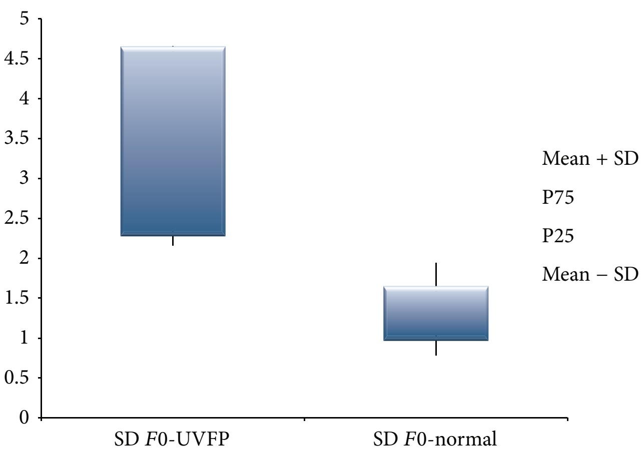


Figure 8: SD of $F0$ (Hz) values for male UVFP and normal subjects.

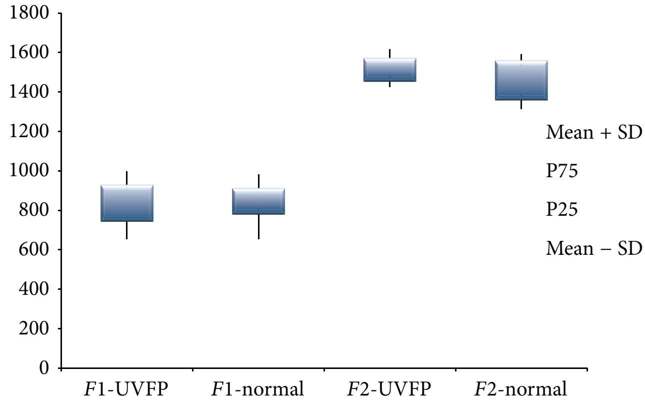


Figure 9: $F1$ and $F2$ frequency (Hz) values for female UVFP and normal subjects.

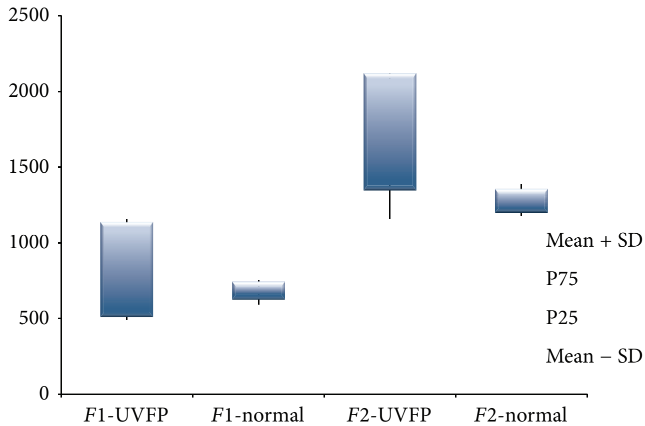


Figure 10: $F1$ and $F2$ frequency (Hz) values for male UVFP and normal subjects.

configurations in subjects with UVFP could guide treatments and help prevent negative compensations [6, 7].

Regarding $F1$ frequency, Table 5 shows that for females there are no statistically significant differences between pathologic and normal subjects (see Figures 9 and 10). A similar result was obtained by Lee et al. [15]. Formant frequency values shown in Table 6 reveal that, for males, differences between groups are statistically significant. Lower values of $F1$ frequencies in UVFP cases were expected (based on data reported previously [15]); however, Table 6 clearly shows that the $F1$ frequency values were higher in the pathologic group. However, authors such as Hartl et al. [2] and D. H. Klatt and L. C. Klatt [56] have also reported higher $F1$ frequency values for voices with similar characteristics to UVFP patients. Since the frequency of $F1$ is inversely related to the vertical movement of the tongue, higher values of this formant (in UVFP subjects) indicate a lower tongue position during phonation for the pathologic subjects. This result is in line with what was found by Higashikawa et al. [57] for whispered voices.

The second formant ($F2$) frequency, which is related to the horizontal tongue movement, is higher in UVFP male subjects (see Figures 9 and 10). This result was also obtained

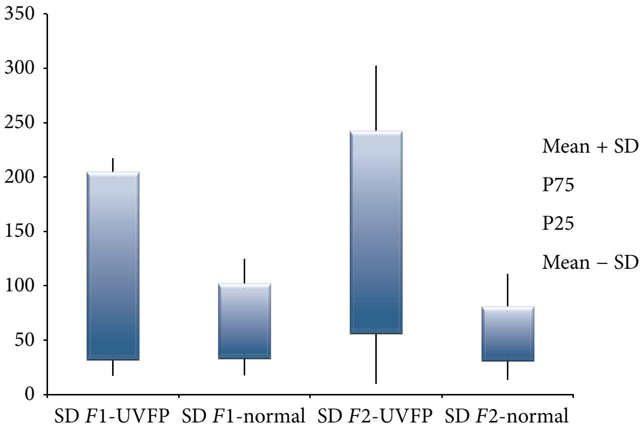


Figure 11: SD of $F1$ and $F2$ frequency (Hz) for female UVFP and normal subjects.

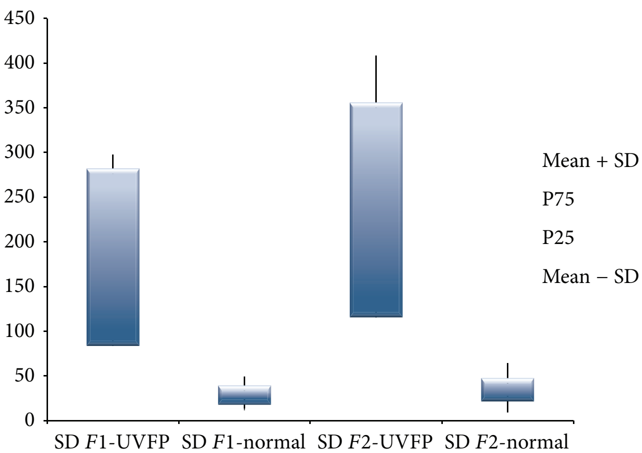


Figure 12: SD of $F1$ and $F2$ frequency (Hz) for male UVFP and normal subjects.

in other studies [2, 15]. For females, although the p value is very close to the significance level, there are no statistically significant differences between normal and UVFP subjects. However, we can see a slightly higher value of $F2$ frequency in the female pathologic group compared to normal females. Therefore results indicate that there could be a tendency to a more advanced tongue position during phonation in cases of UVFP. This is consistent with the results presented by Lotto et al. [16] who studied breathy voices (typical of UVFP).

As for the SD of the frequency of $F1$ shown in Table 5, there were significant differences between the two groups being SD of $F1$ frequency higher in the patients, for male participants. There were no significant differences between groups for females (see Table 6). As for the SD of the frequency of $F2$ there were statistically significant higher values in the UVFP group for both genders (see Figures 11 and 12). Therefore these parameters, especially the SD of the frequency of $F2$, may have an important role in discriminating normal and UVFP voices. Pathologic voices showed higher values of formant frequency SD. These results were also obtained by Lee et al. [15]. This indicates a greater instability of the vocal tract configuration in UVFP during phonation.

Overall results related to the vocal tract configuration ($F1$ and $F2$) show great potential to discriminate between normal and UVFP voices (especially for males) in spite of the localisation of the lesion being at the VF level. This is in agreement with the literature which clearly indicates that the behaviour of the VF is not entirely independent of the vocal tract [55, 58, 59]. Thus, these parameters can add useful information to the assessment procedure and may be used as a complement to the more traditional VF behavioural assessment.

It should be noted that the overall results obtained for females distance themselves from what was initially expected. These differences between genders may be due to a greater technical difficulty in analysing female voices [56, 60]. To a large extent, these difficulties are associated with the identification of formants, due to the fact that $F0$ is higher, and this increases the difficulty in $F1$ estimation [56].

4. Conclusions

In this study various ways of assessing the UVFP voice were combined. Since vocal therapy is one of the first noninvasive treatment options with potential to help the client to reacquire a functional voice, it is fundamental to know in detail the alterations created by the pathology at VF and vocal tract level to better guide the treatment. Perceptual differences between normal and UVFP voices were found. The perceptual parameters that better characterised this data of UVFP subjects were Rough (R) and Breathy (B), but altered values of Asthenic (A) and Strained (S) were also found. As far as acoustic parameters are concerned there were no differences in $F0$ values between normal and UVFP voices in this sample. Jitter (ppq5), shimmer (apq11), HNR, and SD of $F0$ had an important role in discriminating normal and UVFP voices. Measures related to the vocal tract configuration were also indicative of alterations at VF level; therefore the analysis of formant frequencies values and their SD may have an important role in a clinical setting contributing to a better knowledge of the alterations caused by the vocal pathology. Future work should continue to explore formants and their relation to vocal pathology.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the Otorhinolaryngology Team from Hospital de Santo António and Hospital de São João. This work was partially funded by National Funds through FCT (Foundation for Science and Technology), in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014. The Advanced Voice Function

Assessment Databases (AVFAD) project is supported by the School of Health Sciences (ESSUA), University of Aveiro, Portugal.

References

[1] A. Blitzen, M. Brin, and L. Ramig, *Neurologic Disorders of the Larynx*, vol. 101, Thieme, New York, NY, USA, 2nd edition, 2009.

[2] D. M. Hartl, S. Hans, J. Vaissière, M. Riquet, and D. F. Brasnu, "Objective voice quality analysis before and after onset of unilateral vocal fold paralysis," *Journal of Voice*, vol. 15, no. 3, pp. 351–361, 2001.

[3] C. M. R. Pinho, L. M. T. Jesus, and A. Barney, "Aerodynamic measures of speech in unilateral vocal fold paralysis (UVFP) patients," *Logopedics Phoniatrics Vocology*, vol. 38, no. 1, pp. 19–34, 2013.

[4] L. Sulica and A. Blitzer, *Vocal Fold Paralysis*, Springer, New York, NY, USA, 2006.

[5] K. Verdolini, C. Rosen, and R. Branski, *Classification Manual for Voice Disorders-I*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2006.

[6] K. Schwarz, C. A. Cielo, N. Steffen, J. Becker, and G. P. Jotz, "Voice and laryngeal configuration of men with unilateral vocal fold paralysis before and after medialization," *Journal of Voice*, vol. 25, no. 5, pp. 611–618, 2011.

[7] A. L. Rosenthal, S. Y. Lowell, and R. H. Colton, "Aerodynamic and acoustic features of vocal effort," *Journal of Voice*, vol. 28, no. 2, pp. 144–153, 2014.

[8] H. Oguz, M. Demirci, M. A. Safak, N. Arslan, A. Islam, and S. Kargin, "Effects of unilateral vocal cord paralysis on objective voice measures obtained by Praat," *European Archives of Oto-Rhino-Laryngology*, vol. 264, no. 3, pp. 257–261, 2007.

[9] D. K. Chhetri, J. Neubauer, J. L. Bergeron, E. Sofer, K. A. Peng, and N. Jamal, "Effects of asymmetric superior laryngeal nerve stimulation on glottic posture, acoustics, vibration," *Laryngoscope*, vol. 123, no. 12, pp. 3110–3116, 2013.

[10] C. Madill and P. McCabe, "Acoustic analysis using freeware: praat," in *Handbook of Voice Assessments*, Singular, San Diego, Calif, USA, 2011.

[11] A. Vogel, "Multidimensional analysis of voice: computerized speech lab," in *Handbook of Voice Assessments*, Singular, San Diego, Calif, USA, 2011.

[12] P. Alku, "Glottal inverse filtering analysis of human voice production—a review of estimation and parameterization methods of the glottal excitation and their applications," *Sadhana*, vol. 36, no. 5, pp. 623–650, 2011.

[13] I. Titze, *Principles of Voice Production*, National Center for Voice and Speech, Iowa City, Iowa, USA, 2nd edition, 2000.

[14] G. Fant, *Acoustic Theory of Speech Production—With Calculations Based on X-Ray Studies of Russian Articulations*, Mouton, The Hague, The Netherlands, 2nd edition, 1970.

[15] J.-W. Lee, H.-G. Kang, J.-Y. Choi, and Y.-I. Son, "An investigation of vocal tract characteristics for acoustic discrimination of pathological voices," *BioMed Research International*, vol. 2013, Article ID 758731, 11 pages, 2013.

[16] A. J. Lotto, L. L. Holt, and K. R. Kluender, "Effect of voice quality on perceived height of English vowels," *Phonetica*, vol. 54, no. 2, pp. 76–93, 1997.

[17] L. Gordis, *Epidemiology*, Elsevier, Philadelphia, Pa, USA, 3rd edition, 2004.

[18] G. Breakwell, J. Smith, and D. Wright, Eds., *Research Methods in Psychology*, SAGE Publications, London, UK, 4th edition, 2012.

[19] F. Kerlinger and H. Lee, *Foundations of Behavioral Research*, Wadsworth and Thomson, Orlando, Fla, USA, 4th edition, 2000.

[20] L. Jesus, "University of Aveiro's advanced voice function assessment databases (AVFAD)," *Revista de Saúde Pública*, vol. 48, p. 291, 2014.

[21] R. M. Roark, B. C. Watson, R. J. Baken, D. J. Brown, and J. M. Thomas, "Measures of vocal attack time for healthy young adults," *Journal of Voice*, vol. 26, no. 1, pp. 12–17, 2012.

[22] N. Breslow and N. Day, *Statistical Methods in Cancer Research: Vol.1—The Analysis of Case-Control Studies*, IARC, Lyon, France, 1980.

[23] J. Beck, "Organic variation of the vocal apparatus," in *The Handbook of Phonetic Sciences*, W. Hardcastle, J. Laver, and F. Gibbon, Eds., pp. 155–201, Blackwell Publishing, Oxford, UK, 2nd edition, 2010.

[24] S. Linville, "The aging voice," in *Voice Quality Measurements*, R. Kent and M. Ball, Eds., pp. 359–376, Singular, San Diego, Calif, USA, 2000.

[25] I. Chatterjee, H. Halder, S. Bari, S. Kumar, A. Roychoudhury, and P. Murthy, "An analytical study of age and gender effects on voice range profile in bengali adult speakers using phonetogram," *International Journal of Phonosurgery and Laryngology*, vol. 1, no. 2, pp. 65–70, 2011.

[26] H. B. Fisher and S. E. Linville, "Acoustic characteristics of women's voices with advancing age," *Journals of Gerontology*, vol. 40, no. 3, pp. 324–330, 1985.

[27] E. Perrin, C. Berger-Vachon, and L. Collet, "Acoustical recognition of laryngeal pathology: a comparison of two strategies based on sets of features," *Medical and Biological Engineering and Computing*, vol. 37, no. 5, pp. 652–658, 1999.

[28] L. A. Ramig and R. L. Ringel, "Effects of physiological aging on selected acoustic characteristics of voice," *Journal of Speech and Hearing Research*, vol. 26, no. 1, pp. 22–30, 1983.

[29] T. Shipp and H. Hollien, "Perception of the aging male voice," *Journal of Speech and Hearing Research*, vol. 12, no. 4, pp. 703–710, 1969.

[30] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, pp. 341–345, 2001.

[31] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices: a voice analysis system for the screening and laryngeal diseases," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 4, pp. 74–82, 1997.

[32] J. G. Švec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," *American Journal of Speech-Language Pathology*, vol. 19, no. 4, pp. 356–368, 2010.

[33] J. Burg, "Maximum entropy spectral analysis," in *Proceedings of the 37th Meeting of the Society of Exploration Geophysicists*, 1967.

[34] M. Hirano, *Clinical Examination of Voice*, Springer, New York, NY, USA, 1981.

[35] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics," *Computational Linguistics*, vol. 34, no. 4, pp. 555–596, 2008.

[36] F. Jalalinajafabadi, C. Gadepalli, F. Ascott, J. Homer, M. Lujan, and B. Cheetham, "Perceptual evaluation of voice quality and its correlation with acoustic measurement," in *Proceedings of the European Modelling Symposium (EMS '13)*, pp. 283–286, IEEE, Manchester, UK, November 2013.

[37] M. G. Kendall and B. B. Smith, "The problem of *m* rankings," *Annals of Mathematical Statistics*, vol. 10, no. 3, pp. 275–287, 1939.

[38] C. Moers, B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdt, and T. Haderlein, "Vowel- and text-based cepstral analysis of chronic hoarseness," *Journal of Voice*, vol. 26, no. 4, pp. 416–424, 2012.

[39] C. Sellarsa, A. Stantona, A. McConnachiea et al., "Reliability of perceptions of voice quality: evidence from a problem asthma clinic population," *The Journal of Laryngology & Otology*, vol. 123, no. 7, pp. 755–763, 2009.

[40] M. P. Karnell, S. D. Melton, J. M. Childes, T. C. Coleman, S. A. Dailey, and H. T. Hoffman, "Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders," *Journal of Voice*, vol. 21, no. 5, pp. 576–590, 2007.

[41] M. Hirano, S. Hibi, R. Terasawa, and M. Fujiu, "Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia," *Journal of Phonetics*, vol. 14, pp. 445–456, 1986.

[42] P. Yu, R. Garrel, R. Nicollas, M. Ouaknine, and A. Giovanni, "Objective voice analysis in dysphonic patients: new data including nonlinear measurements," *Folia Phoniatrica et Logopaedica*, vol. 59, no. 1, pp. 20–30, 2007.

[43] M. A. Little, D. A. E. Costello, and M. L. Harries, "Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures," *Journal of Voice*, vol. 25, no. 1, pp. 21–31, 2011.

[44] S. Bielamowicz and S. V. Stager, "Diagnosis of unilateral recurrent laryngeal nerve paralysis: laryngeal electromyography, subjective rating scales, acoustic and aerodynamic measures," *Laryngoscope*, vol. 116, no. 3, pp. 359–364, 2006.

[45] C. Baylor, K. Yorkston, E. Strand, T. Eadie, and J. Duffy, "Measurment of treatment outcome in unilateral vocal fold paralysis: a systematic review," UVFP Technical Report 5, Academy of Neurologic Communication Disorders and Sciences, Washington, DC, USA, 2005.

[46] S. Pinho, D. Tsuji, and S. Bohadana, *Fundamentos em Laringologia e Voz*, Guanabara Koogan, Rio de Janeiro, Brasil, 2006.

[47] H. Takahashi, "Assessment of auditory impression of dysphonia," in *Voice Examination Methods*, Japan Society of Logopedics and Phoniatrics, Ed., Interna, Tokyo, Japan, 1979.

[48] P. H. Dejonckere, P. Bradley, P. Clemente et al., "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS)," *European Archives of Oto-Rhino-Laryngology*, vol. 258, no. 2, pp. 77–82, 2001.

[49] N. Yan, L. Wang, and M. L. Ng, "Acoustical analysis of voices produced by Cantonese patients of unilateral vocal fold paralysis acoustical analysis of voices by Cantonese UVFP," in *Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC '13)*, pp. 1–5, IEEE, Kunming, China, August 2013.

[50] V. Teles and A. Rosinha, "Acoustic analysis of formants and measures of the sonorous signal disturbance in non-smoker and non-alcoholic women without vocal complaints," *International Archives of Otorhinolaryngology*, vol. 12, no. 4, pp. 523–530, 2008.

[51] P. Reijonen, S. Lehikoinen-Söderlund, and H. Rihkanen, "Results of fascial augmentation in unilateral vocal fold paralysis," *Annals of Otology, Rhinology and Laryngology*, vol. 111, no. 6, pp. 523–529, 2002.

[52] D. G. Childers and K. Wu, "Gender recognition from speech. Part II: fine analysis," *Journal of the Acoustical Society of America*, vol. 90, no. 4, pp. 1841–1856, 1991.

[53] M. Behlau, *Voz: O Livro do Especialista*, Revinte, Rio de Janeiro, Brazil, 2001.

[54] P. Woo, R. Colton, D. Brewer, and J. Casper, "Functional staging for vocal cord paralysis," *Otolaryngology—Head and Neck Surgery*, vol. 105, no. 3, pp. 440–448, 1991.

[55] R. Kent and C. Read, *The Acoustic Analysis of Speech*, Singular, San Diego, Calif, USA, 2nd edition, 2002.

[56] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–857, 1990.

[57] M. Higashikawa, K. Nakai, A. Sakakura, and H. Takahashi, "Perceived pitch of whispered vowels-relationship with formant frequencies: a preliminary study," *Journal of Voice*, vol. 10, no. 2, pp. 155–158, 1996.

[58] M. Rothenberg, "Source-tract acoustic interaction in breathy voice," in *Vocal Fold Fisiology—Biomechanics, Acoustic and Phonatory Control*, pp. 465–481, The Denver Center for the Performing Arts, Denver, Colo, USA, 1980.

[59] I. R. Titze and B. H. Story, "Acoustic interactions of the voice source with the lower vocal tract," *Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2234–2243, 1997.

[60] H. M. Hanson, "Glottal characteristics of female speakers: acoustic correlates," *Journal of the Acoustical Society of America*, vol. 101, no. 1, p. 466, 1997.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 932519, 8 pages
http://dx.doi.org/10.1155/2015/932519

Research Article

Spoken Word Recognition Errors in Speech Audiometry: A Measure of Hearing Performance?

Martine Coene,[1,2] Anneke van der Lee,[1] and Paul J. Govaerts[1,2]

[1] *Department of Language, Literature and Communication, Language and Hearing Center Amsterdam, VU Free University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands*
[2] *The Eargroup, Herentalsebaan 75, 2100 Antwerpen, Belgium*

Correspondence should be addressed to Martine Coene; m.m.r.coene@vu.nl

Received 19 January 2015; Revised 1 June 2015; Accepted 8 June 2015

Academic Editor: Markus Hess

This report provides a detailed analysis of incorrect responses from an open-set spoken word-repetition task which is part of a Dutch speech audiometric test battery. Single-consonant confusions were analyzed from 230 normal hearing participants in terms of the probability of choice of a particular response on the basis of acoustic-phonetic, lexical, and frequency variables. The results indicate that consonant confusions are better predicted by lexical knowledge than by acoustic properties of the stimulus word. A detailed analysis of the transmission of phonetic features indicates that "voicing" is best preserved whereas "manner of articulation" yields most perception errors. As consonant confusion matrices are often used to determine the degree and type of a patient's hearing impairment, to predict a patient's gain in hearing performance with hearing devices and to optimize the device settings in view of maximum output, the observed findings are highly relevant for the audiological practice. Based on our findings, speech audiometric outcomes provide a combined auditory-linguistic profile of the patient. The use of confusion matrices might therefore not be the method best suited to measure hearing performance. Ideally, they should be complemented by other listening task types that are known to have less linguistic bias, such as phonemic discrimination.

1. Introduction

Speech perception refers to the mapping of acoustic and sometimes visual or haptic signals onto language forms [1]. It is part of the *speech chain* which includes the processes of speech production, transmission, and perception [2]. In this paper, we will be concerned with the latter process. Whenever a speaker utters a sentence, speech sound waves travel to the outer ear, middle ear, and cochlea and are transformed into neural activity which is received and decoded by the brain. This involves two types of hearing processes: the first one is based on sensory information obtained at the level of the ear itself, providing the necessary information in a *bottom-up* way; the second refers to the cognitive part of the perception process and brings in *top-down* information.

A particular way of capturing the proper functioning of the speech chain is by measuring input and output [3], that is, by comparing the speech stimulus as it was originally uttered by the speaker to the way it is understood by the listener in the so-called stimulus-repetition tasks. An implementation of this principle can be found in speech audiometry, a clinical examination that is commonly used to assess the impact of a potential hearing deficit on speech understanding.

In this study, we will investigate the performance of participants with normal hearing on a speech audiometric test battery. Conventional speech audiometric test settings involve the repetition of short words that are presented acoustically to the tested individual. The stimuli typically consist of monosyllabic CVC words. The use of this type of stimuli has the advantage of providing little linguistic context from which to derive information that has been missed. Hence, the results of the test are generally considered to be highly informative with respect to the bottom-up processing of speech information in hearing, that is, at the level of the inner ear.

Our main aim is to gain more insight into phoneme substitutions which are found in speech audiometric data obtained from listeners with Dutch as a native language. In the clinical audiological practice, error patterns in phoneme perception are often analyzed by means of a phoneme confusion matrix. The outcomes are interpreted in terms of the particular configuration of hearing loss of the patient [4]. Apparently, the underlying assumption to do so is that deficits in the auditory periphery are the most apparent source of speech perception difficulties.

The central question that is raised in this paper addresses this particular assumption: Is it indeed the case that the erroneous replacement of a missed consonant is best explained by peripheral influences? In other words, do listeners make maximal use of the auditory information that is accessible to them? Or do they rather fill in the gap based on linguistic information? This question is operationalized by means of three hypotheses which relate to the previously mentioned bottom-up and top-down processes that underlie successful speech understanding.

The remaining part of the paper is structured as follows. In Section 2, we will first discuss the auditory component of the phonemes by which the speech stimuli are built up by means of articulatory features and how they may influence word identification. In Section 3, we will discuss the influence of linguistic factors on the replacement of a missed phoneme by a particular alternative. The aim of the study, its methods, and materials are given in Sections 4 and 5, respectively. In Section 6, the results of the analyses are described and are further discussed in Section 7. Finally, the conclusions of our research are found in Section 8.

2. Auditory Component

As was previously mentioned, the acoustic stimuli that are typically used in speech audiometric test batteries consist of short words. These are combinations of phonemes, that is, small units of speech sound (consonants or vowels) that are capable of conveying a distinction in meaning. A particular way to describe the phonetic content of the phonemes by which the stimuli are made up is by means of articulatory features. These speech features commonly involve voicing, nasality, affrication, duration, and place of articulation. Particular models of speech perception take listeners to be sensitive to these different phonetic features.

In information transfer (IT) analysis [5, 6], for instance, speech stimuli are compared to their repetitions in view of determining the fraction of the original information that has been transmitted to the listener for each feature independently. Information transfer is said to be complete if the listener does not make any confusions between phonemes belonging to a different articulatory feature category; for example, if a voiced bilabial /b/ is replaced by a voiced labiodental /d/, voicing is said to be transmitted whereas place of articulation is not. In case of random guessing and biased responses, input and output will be independent and the IT metric for a speech feature will yield a score of 0%. As such, the model measures the information encoded in the input and in the output and the covariance between them.

Under optimal listening conditions, listeners with normal hearing will rather easily obtain information transfer scores which are close to 100%; that is, they hardly make any errors when repeating the short CVC words that are presented to them. However, in difficult listening conditions, for example, when the stimulus word is presented in the presence of masking noise, the relative information transfer is significantly reduced. As expected, the portion of information which is transmitted drops as a function of the decreasing signal-to-noise ratio (SNR). Importantly, not all features are equally affected in their transmission: for example, whereas voicing and nasality are still discriminable at a SNR of −12 dB, the phoneme's place of articulation is hardly distinguishable even at a SNR of 6 dB [5]. By comparing the transmission rates of the articulatory features under similar listening conditions, it thus becomes possible to determine their relative prominence for a given population of listeners. For hearing impaired listeners, it has been shown that places of articulation errors are more prevalent followed by errors in the manner of articulation [7, 8].

3. Linguistic Component

There is an overwhelming body of literature reporting on nonauditory factors influencing speech understanding. It thus seems reasonable to expect that the potential variation in speech repetition errors cannot be fully explained in terms of the abovementioned articulatory features. Adequate processing of the perceived stimulus crucially relies on postcochlear processes as well, including key features of the central auditory system and the cortex, individual cognitive skills, and, most importantly, also information which comes from the linguistic system itself. In recent studies on speech perception, there is an important focus on nonauditory parameters, giving rise to a new interdisciplinary field of research ("cognitive hearing science"; see, e.g., [9]).

In healthy hearing adults, both auditory and linguistic factors may be taken to contribute to word identification accuracy. This is especially the case for stimuli presented in noise conditions. In individuals in which lower-level sensory processing is failing, top-down influences become proportionally more important: to fill in the missing gaps in the incoming speech signal, listeners can receive feedback from different levels of linguistic knowledge (see, e.g., [10–15] amongst many others).

Fortunately, language is an intrinsically redundant system in which a high amount of information that is relevant for speech understanding is represented more than once in the signal. This holds for different components of language, ranging from speech over morphology to complex syntax. Linguistic redundancy becomes particularly relevant when part of the acoustic information is missing. The *phonemic restoration effect* [16], for instance, is one of several phenomena proving that listeners are able to fill in missed phonemes based on nonauditory information.

For the past few decades, many scholars have investigated how listeners accomplish such a complex task. It is commonly believed that the appropriate comprehension of incoming speech is partially guided by the preceding linguistic context; that is, it enables listeners to make predictions with respect to how a sentence is likely to be continued. Evidence from (computational) psycholinguistic experiments indicates that words are identified more rapidly and more accurately when occurring in sentence contexts in which they have high probability to occur based on semantic and/or syntactic grounds [17–19].

In the literature, there is an ongoing debate with respect to the particular mechanisms underlying the potential serving role of linguistic context in speech comprehension. Within particular models of speech processing, it has been claimed that auditory processing, even at the level of early sensory analysis of the speech signal, is affected by top-down constraints from lexical and discourse processes [20–24]. Yet other, that is, modular, accounts rather follow the perspective of a *feed-forward* architecture, in which the output of early stage auditory processing module is passed on to the next interpretational level without feedback mechanisms that would allow the output of the first module to be corrected [25, 26].

In the context of the present study on speech understanding at the word level, two particular linguistic factors are of interest: (i) the *phonological neighborhood size* and (ii) the *frequency* of the word itself. With respect to the first factor, it is taken that the number of existing alternatives a given word has based on just one different phoneme can affect the listener's understanding of that word. *In concreto*, it takes more time to identify a word when several phonological neighbors are potential candidates. Part of the task of the listener is thus to eliminate the alternatives from his/her accessible lexical memory [27, 28]. Importantly, the impact of phonological neighborhood size on speech perception has also been attested in listeners with a hearing impairment [29].

Secondly, the distribution of a word in a given neighborhood may also be described in terms of frequency of usage [30, 31]. Most of the studies investigating the effect of word frequency on speech perception have found that there is a significant bias favoring the identification of words with a higher frequency of occurrence as compared to low-frequent words [32]. Again, this is the case both for listeners with normal hearing [27] and for listeners with a hearing impairment [29].

4. Aim, Research Questions, and Hypotheses

It has been well documented that there is no one-to-one relation between hearing performance based on pure-tone thresholds and speech understanding (see amongst others [33–35]). Sometimes listeners with relatively high pure-tone thresholds may perform unexpectedly well on word recognition tasks and vice versa. As oral communication is a social skill that heavily relies on the ability to hear and understand speech, in current clinical audiological practice, speech audiometry has become a fundamental tool in the assessment of hearing performance. In tandem with pure-tone audiometry, speech audiometric outcomes are taken to help the audiologist in determining the degree and type of hearing loss, to provide information regarding discomfort or tolerance to speech stimuli, and to measure an individual's functional hearing ability.

The information obtained from this complementary test procedure may also be used to predict a patient's gain in hearing performance with hearing devices and may help to optimize the device settings in view of maximum output.

Yet, an important drawback in the use of speech audiometric test batteries for hearing performance assessment is that current test batteries do not typically use stimuli that are controlled for both auditory and linguistic features. This implies that obtained word identification accuracy scores might not be unequivocally attributed to hearing performance. Against the background of the above sketched state of the art, the main aim of the present study is therefore to gain more insight into the proportional contribution of auditory versus linguistic factors in current speech audiometric word identification tasks.

In this paper, the central question that is raised is at the heart of the ongoing debate with respect to the interaction versus autonomous processing models in speech perception research: How do auditory cues, phonological neighborhood, and word frequency contribute to phoneme identification errors in word recognition tasks?

To answer this question, three hypotheses are raised in which particular predictions are made with respect to word repetition based on the three cues under investigation:

(i) The *auditory* hypothesis takes phoneme identification errors to be mainly driven by bottom-up (sensory) information alone: it predicts that the variance in erroneous responses to a given stimulus is best explained by the random choice out of the entire set of phonemes which is available in a given language, regardless of whether this yields an existing word or not.

(ii) The *lexical* hypothesis predicts that erroneous responses will contain significantly more "existing" words than "nonsense" words.

(iii) The *frequency* hypothesis takes the frequency usage of a word to be the most important factor explaining the variance in erroneous responses given by the listener.

5. Materials and Methods

5.1. Materials. The complete database on which the present analysis is built consists of 184 435 stimulus-response pairs of CVC words. These are drawn either from prerecorded wordlists that are commonly used in speech audiometric test batteries in the Dutch-speaking area (see, e.g., [36]) or from phonemically balanced lists obtained from daily readings of the participants themselves [38]. For an example of such CVC wordlist, see Table 4.

In agreement with a classical speech audiometric test procedure, the CVC words were presented acoustically to the participants, and their repetitions were recorded and subsequently scored on a phonemic level by an experienced audiologist. In line with clinical audiological standards, the presentation level of the stimuli ranged from 40 to 70 dB [39].

A total of 230 participants (146 men, 84 women) with normal hearing abilities participated in this study. They were listeners recruited from the Netherlands or Belgium (Flanders) having Dutch as a native language. Prior to

Table 1: Database of stimulus-response pairs of Dutch CVC words.

	N
Number of participants	230
Total number of word tokens analyzed	21285
Total number of single phoneme errors	1957
Word-initial position	1252
Word-medial position	113
Word-final position	592

participation, their speech production was judged according to the Speech Intelligibility Rate (SIR [40]). Participants whose SIR did not reach the level of *complete intelligibility* were excluded from the study. As can be read from Table 1, the obtained data from these participants contained over 21 000 uttered word repetitions of which 1957 stimulus-response pairs contained single phoneme errors.

5.2. Method. From the abovementioned database, all word types were selected which had at least 1% of consonant confusions. Based on this criterion, the database was narrowed down to word-initial and word-final consonant confusions; these were further analyzed in view of testing the abovementioned *auditory, lexical*, and *frequency* hypotheses.

With respect to the auditory hypothesis, each (erroneous) consonant response was analyzed in relation to its potential alternatives, the latter consisting of the set of consonants occurring in either word-initial or word-final position in the Dutch language. Specifically, within this hypothesis, is it predicted that in a stimulus such as Dutch /bEl/ "bell" the odds that the word-final consonant /l/ is replaced by (erroneous) /m/ (resulting in a "nonsense" word /bEm/) are the same as for any other phonemic alternative whether yielding "nonsense" words (/bEr/) or words that are part of the Dutch lexicon such as /bEn/ "am," /bEs/ "berry," and /bEk/ "beak." Importantly, alternatives are defined based on the consonant inventory of the target language, regardless of whether a particular replacement will yield an existing word.

As for the *lexical hypothesis*, a similar comparison is made between the given response and its potential alternatives, but at the exclusion of nonexisting words in Dutch. Within this hypothesis, based on the above given stimulus /bEl/ "bell," /bEr/ would not be considered as a potential alternative whereas /bEk/ "beak" would. The decision of whether a word is part of the Dutch lexicon has been defined based on *Van Dale's* explanatory dictionary of the Dutch language [41].

Finally, under the *frequency hypothesis*, it is expected that the frequency of usage of the potential response determines the choice of a particular consonant replacing the original one. In this study, word frequencies were calculated based on the Corpus of Spoken Dutch (Corpus Gesproken Nederlands [42]), a 9 000 000-word corpus of contemporary Dutch as spoken in the Netherlands and Flanders by adult-speakers. From this reference corpus, mispronunciations, foreign words, or uncertain pronunciations were excluded, but dialectal or regional pronunciations were not. With respect to the stimulus-response pair /bEl/-/bEm/, the frequency hypothesis would predict the listener to prefer /bEn/ "am" over other less frequent alternatives such as /bEK/ "beak."

At a more detailed level, within each hypothesis, the obtained responses are compared to their potential alternatives in terms of categories of phonetic features such as "voicing," "place," and "manner of articulation" (see Table 5 for an overview, based on [37]). For each articulatory feature, the observed and expected values are calculated and standardized following

$$F = \sqrt{\frac{\left(Fa_{\mathrm{OBS+HE}} - Fa_{\mathrm{EXP}}\left(Fa_{\mathrm{OBS+HE}} + \cdots + Fn_{\mathrm{OBS}}\right)\right)^2 + \cdots + \left(Fn_{\mathrm{OBS}} - Fn_{\mathrm{EXP}}\left(Fa_{\mathrm{OBS+HE}} + \cdots + Fn_{\mathrm{OBS}}\right)\right)^2}{\left(Fa_{\mathrm{OBS+HE}} + \cdots + Fn_{\mathrm{OBS}}\right)^2}}, \quad (1)$$

where Fa = articulatory feature (voicing, place, or manner of articulation), $Fa, \ldots, n$ = feature value (e.g., voiced/voiceless), Fa_{OBS} = the observed value, Fa_{EXP} = the expected value, and HE = hypothetical error.

6. Results

Statistical analysis was done by means of an analysis of variance with the word-internal position of the target consonant (initial versus final) and its phonetic make-up in terms of articulatory features (voicing versus place versus manner) as independent factors and the results for the different consonant replacement strategies (auditory versus lexical versus frequency of usage) as dependent variables. The significance of the main and interaction effects was determined based on Greenhouse-Geisser or Huynh-Feldt corrected values [43]. Post hoc testing was done using the Bonferroni method. Effect-sizes have been determined based on the partial η^2.

(1) The main effect of *replacement strategy* (i.e., the auditory, lexical, or frequency route) yielded an F ratio of $F(1.288, 171.34) = 11.258$, $p < 0.001$, and partial η^2 0.078. The contrast analysis between the different strategies revealed that the lexical replacement "route" ($M = 1.07$, SD = 0.42) is significantly better at predicting erroneous replacements of missed consonants than the purely auditory driven strategy ($M = 1.30$, SD = 0.47), $F(1, 133) = 50.813$, $p < 0.001$, and partial η^2 0.276. The comparison between the "lexical" ($M = 1.07$, SD = 0.42) and "frequency" ($M = 1.03$, SD = 0.68) replacement strategies did not yield any significant results ($F(1, 74) = 0.408$, $p = 0.524$, and partial η^2 0.003). This indicates that listeners prefer to replace missed consonants by alternatives that will yield existing words in the target Dutch language instead of phonologically licit but

Table 2: Mean outcomes with standard deviations of the standardized distances between observed and expected values for the auditory, lexical, and frequency routes for each population separately. Post hoc test statistics represent the obtained p values after Bonferroni correction for multiple testing.

	Auditory	Lexical	Frequency	Test statistics	
NH	1.30 (SD 0.47)	1.07 (SD 0.42)	1.03 (SD 0.68)	Auditory-lexical Lexical-frequency Auditory-frequency	**$p < 0.001$** $p > 0.05$ **$p = 0.001$**

Table 3: Mean outcomes with standard deviations of the standardized distances between observed and expected values for the speech features "voice," "place," and "manner" of articulation for each population separately. Post hoc test statistics represent the obtained p values after Bonferroni correction for multiple testing.

	Voice	Place	Manner	Test statistics	
NH	0.28 (SD 0.29)	0.33 (SD 0.28)	0.42 (SD 0.31)	Voice-place Voice-manner Place-manner	$p = 0.192$ **$p < 0.001$** **$p = 0.001$**

Table 4: Two wordlists from [36].

List 17	List 19
Loop	Goud
Fout	Doek
Maai	Hooi
Rood	Sap
Hoek	Lag
Zich	Jong
Dam	Door
Tien	Pijl
Geeuw	Zin
Kok	Bes
Bel	Kieuw
Huis	Neef

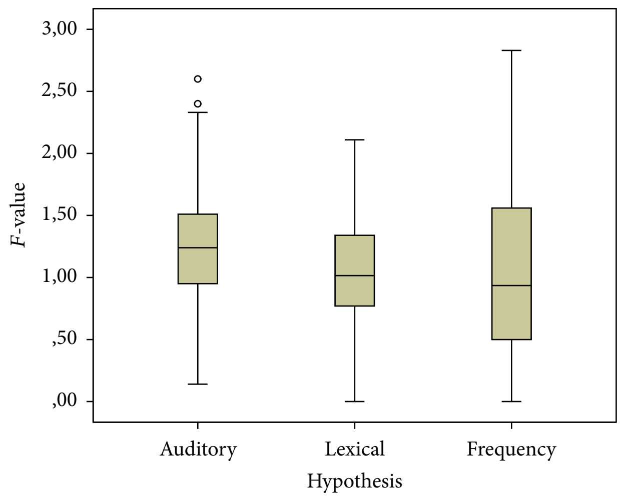


Figure 1: Standardized distances between observed and expected values for erroneous consonants replacements for auditory, lexical, and frequency routes in hearing listeners. Boxes: range between 25th and 75th percentile, whiskers: 1.5 ∗ IQR, and central point: median. Circles: outliers (>1.5 ∗ IQR).

nonsensical alternatives. Yet, at the same time, within the set of possible words, listeners do not opt for the more frequent ones. The mean standardized distances between observed and expected values for erroneous consonant replacement for the auditory, lexical and frequency routes are given in Table 2 and their variation is depicted in Figure 1.

(2) In order to analyze the potential *transfer of phonetic features* of the consonant between stimulus and response, a repeated measures ANOVA was performed comparing the outcomes for voicing, place, and manner of articulation within the lexical replacement strategy only (i.e., the best "model" for the stimulus-response variation resulting from the statistical analysis in (1)). The word-internal position (initial, final) was used as between-subject factor. Test results show that the three articulatory features yield significantly different outcomes ($F(1.805, 240.02) = 12.506$, $p < 0.001$, and partial η^2 0.086). With respect to the between-subject factor, the outcomes for the two possible positions of the consonant within the word (initial versus final) were not significantly different. The mean standardized distances between observed and expected values for erroneous consonant replacement for the speech features "voicing", "place" and "manner" of articulation are given in Table 3 and their variation is depicted in Figure 2.

Post hoc testing with respect to the three articulatory features indicated that "voicing" is best preserved and that "place" of articulation yields significantly lower, that is, better, scores than "manner" of articulation.

7. Discussion

The present study examined to what extent auditory, lexical, and frequency factors may influence erroneous phonemic replacements in an open-set spoken word recognition paradigm. Although the intent of the present study was to verify the presumed auditory nature of consonant replacement strategies in speech audiometric testing, the results of the analysis also address several aspects of human spoken word recognition models.

Table 5: Dutch consonant inventory with articulatory features, according to [37].

Voicing		Place			Manner			
Voiced	Voiceless	Front	Mid	Back	Plosive	Nasal	Fricative	Approximant
b	p	b	d	k	b	m	v	ʋ
d	t	p	t	ŋ	p	n	f	r
v	k	m	n	h	d	ŋ	s	l
z	f	v	z	x	t		z	j
m	s	f	s		k		h	
n	h	ʋ	r				x	
ŋ	x		l					
ʋ			j					
ɣ								
r								
l								
j								

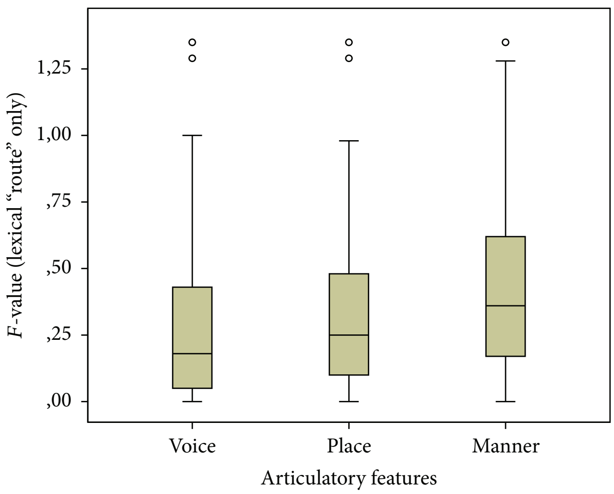


Figure 2: Standardized distances between observed and expected values for erroneous consonants replacements for the speech features "voicing," "place," and "manner" of articulation in hearing listeners. Boxes: range between 25th and 75th percentile, whiskers: 1.5 ∗ IQR, and central point: median. Circles: outliers (values > 1.5 ∗ IQR).

According to the literature, processing an incoming speech signal may occur in two distinct modes: the *autonomous* mode of processing is characterized by an apparent lack of influence from sources of linguistic knowledge or bias (e.g., the lexicon) whereas the *interactive* mode is actually characterized by the influence of such linguistic knowledge (e.g., a wordhood bias) [44]. Previous research has shown that the processing mode may be determined by the type of speech perception task that is used: whereas recognition tasks (e.g., phoneme or word identification) are expected to show a larger effect of lexical influence, discrimination tasks will rather produce an autonomous mode of processing [45].

Building on these insights the methodology followed in this study may be taken to represent a typical case of the first, that is, the interactive, mode due to the fact that listeners have received the instruction to repeat as much of the words that they have heard. This is essentially a "listen-and-repeat" task that calls for speech recognition at the phonemic and word level. Lexical information is therefore expected to be an important predictor of the speech perception errors that are found in such a listening task. The fact that listeners are encouraged to report anything they hear, even if this yields a nonsense word, may well mitigate this linguistic interaction effect, but it will not be able to eliminate it altogether.

A first important finding of this study is therefore that the erroneous replacement of consonants obtained through speech audiometric testing procedures is the result of combined bottom-up and top-down processes in which linguistic factors take up an important portion. Our results indicate that in the case of word identification lexical information is essentially a better predictor of errors than acoustic similarity.

From the point of view of the audiological practice, spoken word recognition measures are commonly used to describe the extent of a patient's hearing impairment, to make a differential diagnosis of auditory disorders, to determine the needs and type of audiological rehabilitation, to verify benefits of hearing aid use, and to monitor the patient's performance over time for diagnostic or rehabilitation purposes [46]. Within this context, it is therefore important to observe that word recognition errors are highly influenced by nonauditory factors. An identical interpretation of correct responses and erroneous consonant replacements in terms of hearing performance would place disproportionate emphasis on top-down processing strategies and would thus seriously overestimate the proper functioning of the inner ear hearing mechanisms.

Secondly, a more systematic analysis of the reception of the different articulatory features of consonants such as voicing, place, and manner of articulation has shown that in erroneous replacement of consonants "voicing" is best preserved whereas "manner" of articulation is most prone to perception errors. This finding is in line with the literature [7, 8] and with our own expectations: as stated in Section 2, voicing is known to be a perceptually very robust feature. According to the *Auditory Enhancement* hypothesis, its redundant acoustic specifications (presence of low-frequency energy, ratio of consonant duration to preceding vowel duration, and presence/absence of aspiration; see [46]) all work together to make this feature acoustically distinctive even under adverse listening conditions.

8. Conclusions

In this study, we have analyzed erroneous phonemic replacements from an open-set spoken word recognition task which is part of a speech audiometric test battery for the Dutch-speaking area. The task was administered to 230 hearing listeners with a native Dutch-speaking background. Erroneous consonant replacements in CVC words were analyzed in terms of the probability of choosing a particular consonant based on acoustic-phonetic, lexical, and frequency variables.

The results indicate that erroneous replacements of consonants are best predicted by lexical information; that is, when part of the incoming speech signal is missing, listeners tend to fill in the gap by picking out one of the stimulus word's phonological neighbors which are part of their mental lexicon. In doing so, listeners do not, however, prefer words that are more frequently heard in the ambient language over alternatives that are less frequent.

Taken together, these results are thought to be of importance for current models of speech perception, pointing in the direction of an interaction between bottom-up and top-down factors even at the lowest levels of spoken word recognition. At the same time, the results are highly relevant for the audiological practice. They draw attention to the fact that erroneous consonant replacements are highly affected by linguistic knowledge. Word repetitions tasks provide a combined auditory-linguistic profile of the patient and might thus not be best suited to measure hearing performance or to guide the rehabilitation of hearing impaired patients. To factor out possible lexical influences on hearing performance measures, they should be complemented by other listening task types that are known to have less linguistic bias, such as phonemic discrimination.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This project has received funding from the European Union's Seventh Framework Program for research, technological development, and demonstration under FP7-PEOPLE-2012-IAPP project "Hearing Minds," Grant Agreement no. 324401.

References

[1] C. A. Fowler and J. S. Magnuson, "Speech perception," in *The Cambridge Handbook of Psycholinguistics*, M. J. Spivey, K. McRae, and M. F. Joanisse, Eds., pp. 3–26, Cambridge University Press, New York, NY, USA, 2012.

[2] P. B. Denes and E. N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language*, W.H. Freeman, Oxford, UK, 2nd edition, 1993.

[3] A. C. M. Rietveld and V. J. van Heuven, *Algemene Fonetiek*, Coutinho, Bussum, The Netherlands, 3rd edition, 2009.

[4] H. H. Dunn, R. J. Roeser, and M. Valente, *Audiological Practice Management*, Thieme Medical Publishers, New York, NY, USA, 2007.

[5] G. A. Miller and P. A. Nicely, "An analysis of perceptual confusions among some English consonants," *The Journal of the Acoustical Society of America*, vol. 27, pp. 338–352, 1955.

[6] E. Sagi and M. A. Svirsky, "Information transfer analysis: a first look at estimation bias," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2848–2857, 2008.

[7] T. Baer, B. C. J. Moore, and K. Kluk, "Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies," *Journal of the Acoustical Society of America*, vol. 112, no. 3 I, pp. 1133–1144, 2002.

[8] D. A. Vickers, B. C. J. Moore, and T. Baer, "Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," *Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1164–1175, 2001.

[9] S. Arlinger, T. Lunner, B. Lyxell, and M. Kathleen Pichora-Fuller, "The emergence of cognitive hearing science," *Scandinavian Journal of Psychology*, vol. 50, no. 5, pp. 371–384, 2009.

[10] E. L. J. George, A. A. Zekveld, S. E. Kramer, S. T. Goverts, J. M. Festen, and T. Houtgast, "Auditory and nonauditory factors affecting speech reception in noise by older listeners," *Journal of the Acoustical Society of America*, vol. 121, no. 4, pp. 2362–2375, 2007.

[11] S. E. Kramer, A. A. Zekveld, and T. Houtgast, "Measuring cognitive factors in speech comprehension: the value of using the Text Reception Threshold test as a visual equivalent of the SRT test," *Scandinavian Journal of Psychology*, vol. 50, no. 5, pp. 507–515, 2009.

[12] A. A. Zekveld, D. J. Heslenfeld, J. M. Festen, and R. Schoonhoven, "Top-down and bottom-up processes in speech comprehension," *NeuroImage*, vol. 32, no. 4, pp. 1826–1836, 2006.

[13] M. K. Pichora-Fuller, "Auditory and cognitive processing in audiologic rehabilitation," in *Adult Audiologic Rehabilitation: Advanced Practices*, J. Spitzer and J. Montano, Eds., pp. 519–536, Plural Publishing, San Diego, Calif, USA, 2013.

[14] M. Rudner, C. Foo, J. Rönnberg, and T. Lunner, "Cognition and aided speech recognition in noise: specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids," *Scandinavian Journal of Psychology*, vol. 50, no. 5, pp. 405–418, 2009.

[15] J. Rönnberg, M. Rudner, T. Lunner, and A. A. Zekveld, "When cognition kicks in: working memory and speech understanding in noise," *Noise and Health*, vol. 12, no. 49, pp. 263–269, 2010.

[16] R. M. Warren, "Perceptual restoration of missing speech sounds," *Science*, vol. 167, no. 3917, pp. 392–393, 1970.

[17] F. Grosjean, "Spoken word recognition processes and the gating paradigm," *Perception & Psychophysics*, vol. 28, no. 4, pp. 267–283, 1980.

[18] V. Steinbiss, H. Ney, X. Aubert et al., "The Philips Research system for continuous-speech recognition," *Philips Journal of Research*, vol. 49, no. 4, pp. 317–352, 1995.

[19] A. J. Lowe, *The relative contribution of top-down and bottom-up information during lexical access [Ph.D. thesis]*, University of Edinburgh, Edinburgh, Scotland, 1990.

[20] J. Morton, "Word recognition," in *Psycholinguistics 2: Structures and Processes*, J. Morton and J. D. Marshall, Eds., pp. 107–156, MIT Press, Cambridge, UK, 1979.

[21] W. D. Marslen-Wilson and A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech," *Cognitive Psychology*, vol. 10, no. 1, pp. 29–63, 1978.

[22] D. J. Foss and M. A. Blank, "Identifying the speech codes," *Cognitive Psychology*, vol. 12, no. 1, pp. 1–31, 1980.

[23] J. L. McClelland, D. Mirman, and L. L. Holt, "Are there interactive processes in speech perception?" *Trends in Cognitive Sciences*, vol. 10, no. 8, pp. 363–369, 2006.

[24] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," *Cognitive Psychology*, vol. 18, no. 1, pp. 1–86, 1986.

[25] D. Norris and J. M. McQueen, "Shortlist B: a Bayesian model of continuous speech recognition," *Psychological Review*, vol. 115, no. 2, pp. 357–395, 2008.

[26] D. Norris, J. M. McQueen, and A. Cutler, "Merging information in speech recognition: feedback is never necessary," *Behavioral and Brain Sciences*, vol. 23, no. 3, pp. 299–370, 2000.

[27] J. L. Metsala, "An examination of word frequency and neighborhood density in the development of spoken-word recognition," *Memory & Cognition*, vol. 25, no. 1, pp. 47–56, 1997.

[28] E. T. Auer Jr., "The influence of the lexicon on speech read word recognition: contrasting segmental and lexical distinctiveness," *Psychonomic Bulletin & Review*, vol. 9, no. 2, pp. 341–347, 2002.

[29] C. Richard, L. Tordella, A. Bernard, C. H. Martin, S. Roy, and A. Moulin, "Ageing and linguistic factors influence on speech audiometry," *Otolaryngology—Head and Neck Surgery*, vol. 147, p. 198, 2012.

[30] P. Assman and Q. Summerfield, "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System, Springer Handbook of Auditory Research*, S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay, Eds., pp. 231–308, Springer, New-York, NY, USA, 2004.

[31] K. I. Forster, "Accessing the mental lexicon," in *New Approaches to Language Mechanisms*, R. J. Wales and E. Walker, Eds., pp. 257–287, North-Holland, Amsterdam, The Netherlands, 1976.

[32] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: the neighborhood activation model," *Ear and Hearing*, vol. 19, no. 1, pp. 1–36, 1998.

[33] P. Era, J. Jokela, Y. Qvarnberg, and E. Heikkinen, "Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages," *Audiology*, vol. 25, no. 6, pp. 338–352, 1986.

[34] T. Letowski, P. Hergenreder, and H. Tang, "Relationships between speech recognition threshold, average hearing level, and speech importance noise detection threshold," *Journal of Speech and Hearing Research*, vol. 35, no. 5, pp. 1131–1136, 1992.

[35] S. Coren, L. M. Ward, and J. T. Enns, *Sensation and Perception*, Wiley, New York, NY, USA, 2004.

[36] A. J. Bosman, J. Wouters, and W. Damman, "Realisatie van een cd voor spraakaudiometrie in Vlaanderen," *Logopedie en Audiologie*, vol. 9, pp. 218–225, 1995.

[37] G. Booij, *The Phonology of Dutch*, Clarendon Press, Oxford, UK, 1995.

[38] A. Hammer, B. Vaerenberg, W. Kowalcyk, L. F. M. ten Bosch, M. Coene, and P. Govaerts, "Balancing word lists in speech audiometry through large spoken language corpora," in *Proceedings of the Interspeech*, pp. 3613–3616, Lyon, France, 2013.

[39] D. Omar Robinson and M. J. Koenigs, "A comparison of procedures and materials for speech reception thresholds," *Journal of the American Auditory Society*, vol. 4, no. 6, pp. 227–230, 1979.

[40] R. M. Cox and D. M. McDaniel, "Development of the speech intelligibility rating (Sir) test for hearing aid comparisons," *Journal of Speech and Hearing Research*, vol. 32, no. 2, pp. 347–352, 1989.

[41] *Van Dale Groot Woordenboek van de Nederlandse Taal*, Van Dale Uitgevers, Antwerp, Belgium, 14th edition, 2005.

[42] N. Oostdijk, "Het Corpus Gesproken Nederlands: veelzijdig onderzoeksinstrument voor o.a. taalkundig en taal- en spraaktechnologisch onderzoek," *Link*, vol. 14, no. 1, pp. 3–6, 2003.

[43] A. Field, *Discovering Statistics Using SPSS*, SAGE Publications, Thousand Oaks, Calif, USA, 2007.

[44] M. Key, "Interactive and autonomous modes of speech perception: phonological knowledge and discrimination in English and French listeners," *Laboratory Phonology*, vol. 11, pp. 73–74, 2011.

[45] J. Kingston, S. Kawahara, D. Chambless, M. Key, D. Mash, and S. Watsky, "Context effects as auditory contrast," *Attention, Perception, & Psychophysics*, vol. 76, no. 5, pp. 1437–1464, 2014.

[46] R. L. Diehl, "On the robustness of speech perception," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS '11)*, Hong Kong, August 2011.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 754876, 7 pages
http://dx.doi.org/10.1155/2015/754876

Research Article

Stem Cell Therapy in Injured Vocal Folds: A Three-Month Xenograft Analysis of Human Embryonic Stem Cells

Bengt Svensson,[1,2] Srinivasa R. Nagubothu,[3] Christoffer Nord,[4] Jessica Cedervall,[5] Isabell Hultman,[6] Lars Ährlund-Richter,[6] Anna Tolf,[7] and Stellan Hertegård[8,9]

[1] *Department of Public Health and Clinical Medicine, Unit of Clinical Research Center, Umeå University, 90187 Umeå, Sweden*
[2] *Department of Otorhinolaryngology, 83183 Östersund, Sweden*
[3] *Center for Hematology and Regenerative Medicine, Department of Laboratory Medicine, Karolinska Institutet, 14186 Stockholm, Sweden*
[4] *Centre for Molecular Medicine, Umeå University, 90187 Umeå, Sweden*
[5] *Department of Medical Biochemistry and Microbiology, Uppsala University, 75185 Uppsala, Sweden*
[6] *Department of Women's and Children's Health, Karolinska Institutet, 17177 Stockholm, Sweden*
[7] *Department of Immunology, Genetics, and Pathology, University Hospital, 75185 Uppsala, Sweden*
[8] *Department of Clinical Sciences and Intervention, Karolinska Institutet, 14186 Stockholm, Sweden*
[9] *Department of Otorhinolaryngology, Karolinska University Hospital, Huddinge, 14186 Stockholm, Sweden*

Correspondence should be addressed to Stellan Hertegård; stellan.hertegard@ki.se

Received 21 January 2015; Accepted 8 April 2015

Academic Editor: Haldun Oguz

We have previously shown that human embryonic stem cell (hESC) therapy to injured rabbit vocal folds (VFs) induces human tissue generation with regained VF vibratory capacity. The aims of this study were to test the sustainability of such effect and to what extent derivatives of the transplanted hESCs are propagated in the VFs. The VFs of 14 New Zealand rabbits were injured by a localized resection. HESCs were transplanted to 22 VFs which were analyzed for persistence of hESCs after six weeks and after three months. At three months, the VFs were also analyzed for viscoelasticity, measured as dynamic viscosity and elastic modulus, for the lamina propria (Lp) thickness and relative content of collagen type I. Three months after hESC cell therapy, the dynamic viscosity and elastic modulus of the hESC treated VFs were similar to normal controls and lower than untreated VFs ($p \leq 0.011$). A normalized VF architecture, reduction in collagen type I, and Lp thickness were found compared with untreated VFs ($p \leq 0.031$). At three months, no derivatives of hESCs were detected. HESCs transplanted to injured rabbit VFs restored the vibratory characteristics of the VFs, with maintained restored function for three months without remaining hESCs or derivatives.

1. Introduction

Tissue defects in a vocal fold (VF) heal with scar formation. The scar tissue causes stiffness in the lamina propria (Lp) which renders disturbed viscoelastic properties to the VF [1]. A scarred VF causes severe voice problems [2]. Treatment is difficult. Different injectable substances have been used to augment scarred VFs [3–12]. Several of these substances have rendered improved vibratory characteristics to the scarred VF, but presently there is no effective method to prevent VF scarring or to heal VF scars.

In xenograft studies, human mesenchymal stem cells (hMSCs) in different preparations have lately shown promising results in healing VF scars [13–18].

However, hMSCs have only been found to perform single divisions and have not been shown to have the capacity to differentiate into new tissue in the VFs [15, 16].

Since the first description of successful *in vitro* culture of human embryonic stem cells (hESCs), such cells have been recognized as to provide a potential resource for cell transplantations. HESCs are derived from the blastocyst of an early embryo and are unique in the sense that they have

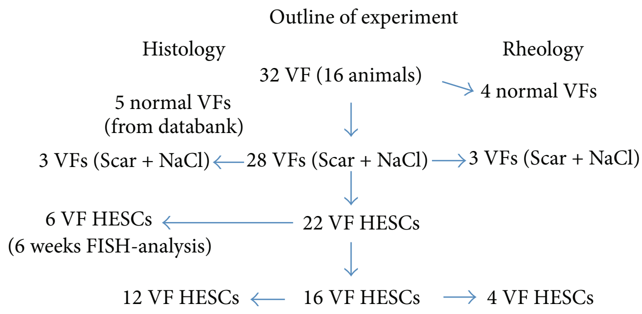


Figure 1

the capacity to self-renew but also that of pluripotency, that is, to differentiate into all cell types of the human body. In a previous study with human embryonic stem cells (hESCs) transplanted to injured VFs of rabbits, we found that hESCs can survive one month in a xenograft rabbit model and that they during this time can differentiate into fully developed epithelium, muscle, and cartilage tissue, adequately placed close to or interpositioned with the rabbit corresponding tissue, thus replacing lost rabbit VF tissue [19]. This study also showed that the VFs treated with hESCs gained a significantly improved viscoelastic function, measured as dynamic viscosity and elastic modulus, compared with untreated scarred VFs.

Using the same xenograft model, the aim of this study was to analyze in longer term, that is, three months, the sustainability of the improved healing in the hESC treated scarred rabbit VFs. Another aim was to examine the destiny of the hESCs and to explore if malignancies or teratomas would develop in the hESC transplanted VFs.

2. Material and Methods

The principal study design has been used by several investigators [7, 10, 13]. American and Swedish principles of laboratory animal care were followed. The experiment was approved by the Local Ethics Committee at the Karolinska Institutet and by the Regional Committee for Animal Experimentation, Stockholm, Sweden, S29-06, S115-08.

Sixteen female New Zealand white rabbits (bw 3.0 kg–4.0 kg) were used in the study. Twenty-eight VFs were operated on and the remaining four VFs were left as normal controls and used in the viscoelastic measurements. HESCs were transplanted to 22 of the 28 operated VFs and the remaining 6 VFs were left untreated, that is, scarred-untreated. Data for another five normal VFs were collected from a data bank from earlier experiments and were added in the histologic analyses, $n = 32 + 5$ [15, 16]. Figure 1 shows outline of experiment.

2.1. Vocal Fold Scarring. After premedication with glycopyrrolate (0.1 mg/kg s.c.) and Hypnorm (fentanyl citrate 0.3 mg/mL mixed with fluanizonum 10 mg/mL, 0.3 mL/kg i.m., Janssen Pharmaceutica, Beerse, Belgium) the animals were anaesthetized with diazepam (2 mg/kg i.v.). The laryngeal structures and the mobility of the cricoarytenoid joints were found normal at examination by means of a modified 4.0 mm paediatric laryngoscope (model 8576E, Karl Storz Endoskope, Tuttlingen, Germany) and a Storz-Hopkins 0° 2.7 mm rigid endoscope (model 7218A). The scarring procedure was performed with a 1.5 mm microcup forceps (MicroFrance Medtronic, Düsseldorf, Germany) excising the mucosa and the superficial layer of the thyroarytenoid muscle. A digital video recorded on a computer was made of the VFs before and after the operation (Richard Wolf video camera Number 5512 and a Canopus ADVC100 digital video converter, Reading, UK).

2.2. Human Embryonic Stem Cell Preparation and Characterization. The hESC line HS181 (46; XX) derived by the hESC network at Karolinska Institutet, Stockholm, Sweden, was kindly provided by Professor Hovatta et al. [20]. The HS 181 cells were maintained as previously described on mitotically inactivated, by 35 Gy irradiation, human foreskin fibroblasts [21]. HS181 cells corresponding to passage 33 were used in the study.

2.3. Vocal Fold hESC Transplantation. Three to four undifferentiated hESC colonies were dissected directly from the culture plates and aspirated by a 27-gauge needle of a laryngeal injector with a syringe of 1 mL saline (Medtronic Xomed, Inc., Jacksonville, FL). The 1 mL saline then contained approximately 10^4 cells per 0.1 mL. Under video monitoring, using the 27-gauge needle Xomed laryngeal injector, the hESCs were transplanted by an injection of 0.1 mL of the solution into the lamina propria and/or the superficial part of the thyroarytenoid muscle of the scarred VF. The injection was carried out directly after the scar excision procedure. The correct injection site was stated by observed bulging of the VF corresponding to the injected volume.

To reduce rejection, the animals that received hESCs were treated with immunosuppressant Tacrolimus (TC), (0.05 mg/kg bw s.c.) every second day. The dose was based on the recommended dose/kg from the manufacturer and our previous experiments in rabbits [15, 16].

2.4. Sample Procurement. The animals were sacrificed with an overdose of pentobarbital sodium. Each larynx was dissected out and divided in the posterior midline. Three animals were sacrificed after 6 weeks and the VFs were analyzed for persistence of hESCs. The remaining thirteen animals were sacrificed after 3 months. Eleven hemilarynges were fresh frozen and kept at −70°C until viscoelastic analyses. The remaining hemilarynges were placed in 4% formaldehyde for histologic and antibody analyses.

2.5. Histologic Measurements. After fixation in 4% formaldehyde and 70% ethanol, the VFs removed from the larynges were further processed, dehydrated, and finally embedded in paraffin wax and cut into 5 μm thick horizontal sections covering the whole thickness of each VF. Staining was made with hematoxylin-eosin (HE) for histologic analyses. Image analyses were made at 10x or 20x magnification after digitization of the microscopic images. The slides were blindly analyzed at the Department of Pathology, University Hospital, Uppsala, Sweden. Inter and intrareliability were assessed by blind reexamination of 10% of the slides, randomly chosen. The results were identical.

Twelve out of the 22 hESC treated VFs were prepared for histologic measurements at time point of three months. Comparisons with three scarred untreated (Scar + NaCl) VFs and with data for 5 normal VFs from the databank were performed [15, 16].

2.5.1. Immunohistochemistry for Collagen Type I Staining. Staining was performed as previously described [13, 19]. Briefly, slides were deparaffinized in xylene, rehydrated in alcohol, and blocked in PBS containing 3% BSA. Slides were incubated with a primary antibody (antibody 6308, Abcam, Cambridge, UK), followed by incubation with a secondary antibody (nr.A21127 Jackson Immuno Research labs Inc., West Grove, PA). Sections were rehydrated in ethanol and xylene and mounted with Vectashield containing DAPI (Vector labs Inc., Burlingame, CA). The relative contents of collagen type I in the VFs were measured from the digitized stains after a color filtering and normalization process with Photoshop (version 8.0) and a custom made software that automatically summarizes color change for the collagen type I linked fluorescent antibody (Software by Hans Larsson, Karolinska Institutet, Department of Phoniatrics).

2.5.2. Lamina Propria (Lp) Thickness. After being embedded in paraffin wax, each VF was cut in horizonal 5 μm thick sections covering the whole VF. The right angle of the microtome toward the specimen was meticulously adjusted for each sample. The measurements of the Lp thickness were carried out on the digitized HE image representing the optimal level of each VF (custom made software by Hans Larsson, Karolinska Institutet). The Lp was measured at three spots representing each third of the VF. If a tendency of polyp formation was seen, the polyp was included in that section's measure point. Each single value was then used in the statistic evaluation.

The Lp of two of the hESC treated VFs were partly damaged in the cutting preparation processes and were left out in the Lp thickness measurements, resulting in $n = 10$ for the hESC treated VFs in the Lp thickness calculations.

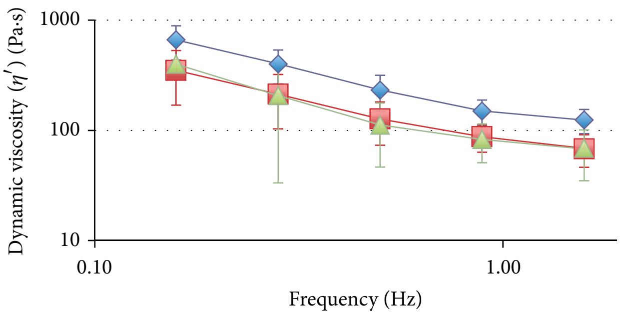
2.5.3. Hematoxylin-Eosin Staining for Analysis of General Fibrosis. The VFs were characterized into four categories depending on grade of scarring, that is, fibrosis. Grade a showed no or minimal signs of fibrosis. Grade b showed a focal or noncompact fibrosis in the Lp or superficial vocal muscle. Grade c showed a more compact fibrosis in the Lp and superficial muscle and Grade d showed a compact fibrosis in Lp and superficial muscle as well as fibrosis in the deeper part of the vocal muscle [16].

2.6. Fluorescence In Situ Hybridization (FISH-Analysis) for Persistence of Cell Derivatives from the Transplanted hESCs. Detection of human cells in the VFs was performed with a human DNA specific reference probe linked to a fluorescent molecule, that is, FISH-analysis.

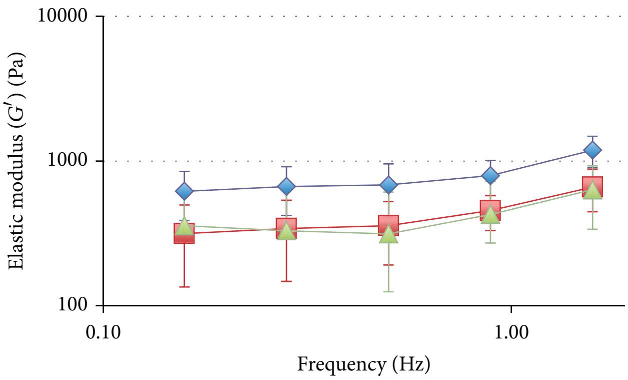
The FISH-analysis was accomplished as previously described [19]. Briefly, slides were deparaffinized in xylene and rehydrated in alcohol, followed by pretreatment with pepsin and hybridization over night at 38°C with a human specific fluorescent probe (CEP X (DXZ1) Spectrum Green SRY Probe, human genomic DNA, Vysis Inc., Burlingame, CA). Six hESC treated VFs were analyzed after six weeks and 12 after three months.

2.7. Viscoelastic Measurements. The viscoelastic shear properties of VF tissue have been studied by several researchers [22, 23]. The parallel-plate rheometer in this experiment produces sinusoidal shear small amplitude oscillations at increasing frequency (within 0.01–15 Hz). We used an AR 2000 Rheometer (TA Instrument) with a stationary lower plate (8 mm diameter) separated by about 0.5 mm from a rotating upper plate. Tissue samples from the eleven fresh frozen VFs (4 scarred VFs injected with hESCs (HESC), 3 scarred injected with saline (Scar + NaCl), and 4 untreated, i.e., normal VFs (Normal)) were thawed in room temperature, dissected, and analyzed at 37°C in the parallel-plate rheometer. The samples included Lp and the superficial part of the thyroarytenoid muscle. The tissue was kept moist with saline during the measurements. All rheometric measurements were performed with a constant strain level transferred from the sample to the upper plate where it was measured with a linear variable displacement transducer. In this experiment, the response and reproducibility were stable up to 2-3 Hz. For higher frequencies, the results were not stable probably due to inertia of the measurement system as the tissue samples were not geometrically perfectly flat and did not completely fill out the 8 mm plate space. The dynamic viscosity (η' in Pa·s) and elastic modulus (G' in Pa) were derived as a function of frequency. Dynamic viscosity is a measure of material's resistance to shear flow. The elastic (storage) modulus (G') represents a measure of material's stiffness in shear. As mentioned in this experiment, the gap between the plates was not completely filled with tissue. Thus the absolute levels of η' and G' may not be accurate. However, the same dissection procedure and amount of tissue were used for all samples which allows for comparison between the different groups.

Measurements of G' (Pa) and η' (Pa·s) as functions of frequency, f, (in Hz) were plotted in log-log scale as shown in Figures 2(a) and 2(b). Curve-fitting regression was then performed for each curve to examine the relationships between G' and f and between η' and f. The obtained data were properly described using the quadric model rather than



(a)



(b)

Figure 2: Rheological data showing (a) dynamic viscosity and (b) elastic modulus (as means ± 2 SD) versus frequency. Both were significantly reduced in the vocal folds (VFs) treated with human embryonic stem cells (hESCs) compared with untreated VFs (Scar + NaCl) ($p = 0.011$ and $p < 0.001$, resp.). Dynamic viscosity and elastic modulus of the hESC treated VFs did not significantly differ from normal VFs.

the linear one. The quadratic model was used for both G' and η', that is, $\log(G' \text{ or } \eta') = B_0 + B_1 \cdot \log(f)$, where B_0, and B_1 are coefficients of parameterization. The curve-fitting estimations, based on least-squares regression analysis, resulted in highly significant findings using the ANOVA F test in all cases (G' $p = 0.001$ and η' $p = 0.006$). The significant values of the F test suggested that the variation explained by the model was not due to chance. Goodness of fit was also estimated by the coefficient of determination, R^2. The R^2 statistic is a measure of the strength of association between the observed and model-predicted values for both $\log(G')$ and $\log(\eta')$. The values of R^2 were high for each regression model indicating goodness of fit ($R^2 > 0.99$ for both $\log(G')$ and $\log(\eta')$).

2.8. Statistics. Differences between groups were assessed using Mann-Whitney U test for independent data. For the histologic measurements, each single value was included when differences between the various groups were estimated. Calculations, whether or not the dynamic viscosity and the elastic modulus, respectively, differed between normal, hESC treated VFs, and untreated scarred controls, were performed with the binomial test. In the regression analyses, the ANOVA F test was used. To calculate differences between groups in the analyses of general fibrosis, Fisher's exact test was used. Statistical significance was considered when $p < 0.05$.

3. Results

3.1. Viscoelastic Analyses

Dynamic Viscosity, η' (Pa·s). Scarring significantly increased the dynamic viscosity, indicating stiffer folds, compared with the normal VFs ($p = 0.006$). Treatment with hESCs significantly decreased the dynamic viscosity compared with the untreated scarred controls (Scar + NaCl) ($p = 0.011$) and was not significantly different from the unscarred controls, that is, normal VFs ($p = 0.4$) (Figure 2(a)).

Elastic Modulus, G' (Pa). Scarring also significantly increased the elastic modulus compared to normal VFs ($p = 0.001$). Treatment with hESCs significantly decreased the elastic modulus in comparison with the untreated scarred controls (Scar + NaCl) ($p < 0.001$). No significant difference was shown for the hESC treated VFs compared with the unscarred controls, that is, normal VFs ($p = 0.4$) (Figure 2(b)).

3.2. FISH-Analysis for Persistence of Transplanted hESC Derivatives. Six weeks after the hESC injections, four out of six treated VFs showed presence of human cells as detected by FISH-analysis. Mitotic activity was rare and detected in 2 out of 6 VFs (Figure 3).

Three months after the hESC injections none of the 12 hESC treated VFs showed persistence of human cells or derivatives, as indicated by the lack of cells positive for the FISH analysis.

3.3. Histologic Analyses

3.3.1. Lamina Propria Thickness. The hESC treated VFs showed a significantly reduced Lp thickness compared with the scarred untreated VFs ($p < 0.001$). No significant difference was shown between hESC treated VFs and normal VFs ($p > 0.05$). The difference between untreated VFs and normal VFs was significant ($p < 0.001$) (Figure 4). The Lp thickness was performed blindly. Mean Inter and intrareliability were examined on 20% of the samples and the difference between the measurements was found to be less than 5% with a correlation coefficient of 0.99 at repeated measurements.

3.3.2. Collagen Type I Staining. The hESC treated VFs showed significantly reduced collagen type I in comparison with the scarred untreated VFs ($p = 0.031$). The difference between normal and hESC treated VFs was not significant ($p > 0.05$). The difference between normal and untreated VFs was significant ($p = 0.037$) (Figure 5).

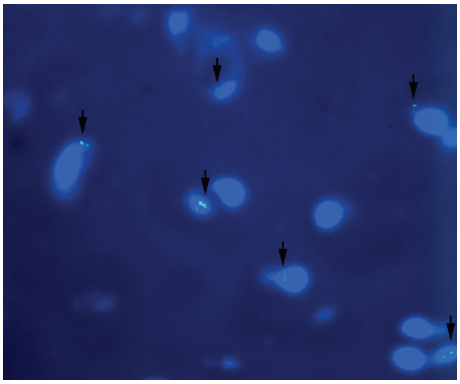


FIGURE 3: Fish staining (fluorescence in situ hybridization linked to a green fluorescent molecule; see text) showing hESCs in division at six weeks. Green enlightening represents human cells. Blue cells represent DAPI (4′,6-diamidino-2-phenylindole) fluorescent stain colored nuclei, in both rabbit and human cells. Black arrow marks human cells. 40x magnification.

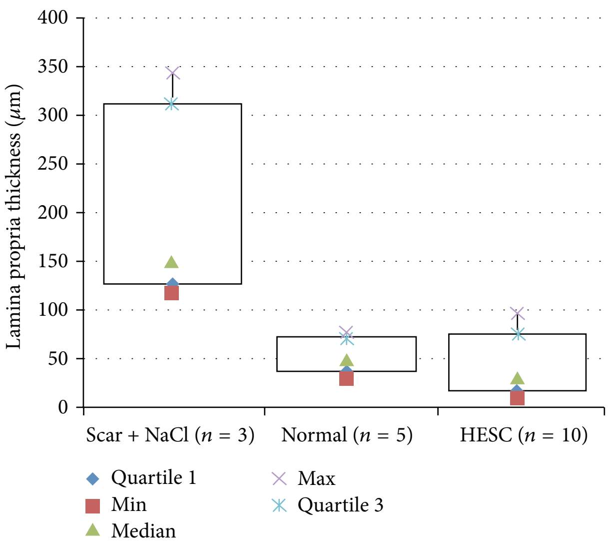


FIGURE 4: Lamina propria (Lp) thickness (μm) is reduced in the vocal folds (VFs) treated with human embryonic stem cells (hESCs) compared with untreated scarred VFs (Scar + NaCl) ($p < 0.001$). The difference between hESC treated and normal VFs is nonsignificant.

3.3.3. Hematoxylin-Eosin Staining for Analysis of General Fibrosis. Ten of the hESC treated VFs were placed in group b, one in each of groups a and c, and none in d ($n = 12$). Of the three scarred untreated VFs, two were placed in the c group and one in the d group ($n = 3$). When the a and b groups were compared with the c and d groups, the hESC treated VFs were placed in the a-b group and the untreated VFs in the c-d group ($p < 0.011$). Inter and intrareliability were identical in the blinded analyses (Figure 6).

3.3.4. Hematoxylin-Eosin Staining for Malignancies or Teratomas. From the HE staining, there were no malignancies or teratomas found in the hESC transplanted VFs.

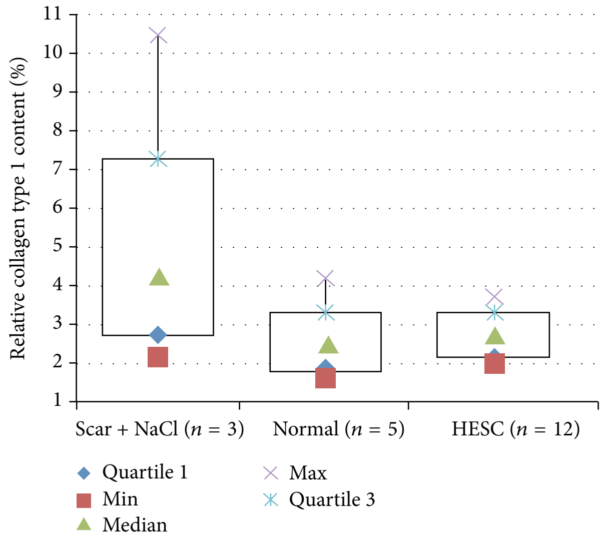


FIGURE 5: Collagen type I content (%). HESC treated vocal folds (VFs) show a reduction in collagen type I content compared with untreated scarred VFs (Scar + NaCl) ($p = 0.031$) and no significant difference to normal VFs.

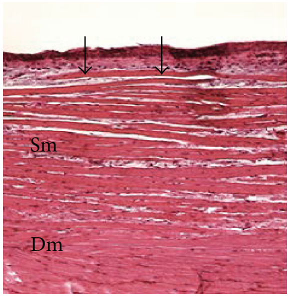
4. Discussion

The study is the second one published on hESC treatment of VFs. It shows that the hESC improved healing of scarred rabbit VFs seen after one month is sustainable over a longer period of time. Studies by our group and others have shown that VF scarring is established affecting both tissue viscoelasticity and histology within 3 months in a rabbit model. The specific maturation process of the scar may then take 3–6 months [15, 16, 24].

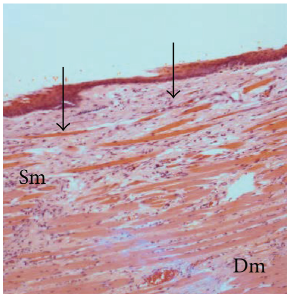
The derivatives of the transplanted hESCs were able to survive for six weeks but were not detected after three months in this rabbit model. Although a survival of hESC derivatives under the present level of detection cannot be ruled out, this indicates that the regenerated human tissue (epithelium, muscle, and cartilage) found after one month [19] can be expected to undergo apoptosis or rejection in a three-month period.

However, the present study shows that the vibratory characteristics, measured as dynamic viscosity and elastic modulus, were improved compared with untreated scarred VFs and showed statistically no difference to normal VFs after the three months. Moreover, in the histologic analyses no extensive scarring was found in the hESC treated VFs. In fact the collagen type I content and the Lp thickness values for the hESC treated VFs were not found different from the normal VFs. This indicates that lost regenerated human tissue is not replaced by scar tissue but by regeneration of compatible native rabbit tissue. By such factors, the apoptotic hESCs signal to the surrounding rabbit cells, in order to make them proliferate and not to create scar tissue, is unknown and is a challenge for further research.

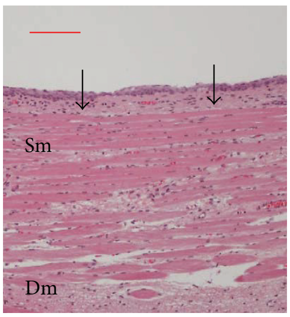
Two hESC treated VFs did not show any human cells already after six weeks. These two VFs were classified to group b in the classification system of general fibrosis, that is,



(1) Normal vocal fold (group a)



(2) Scarred untreated VF (Scar + NaCl) (group d)



(3) HESC treated VF at 3 months (group b)

Figure 6: Longitudinal hematoxylin-eosin staining of the midmembranous part of (1) normal vocal fold (VF) with minimal loose connective tissue superficially under the lamina propria (Lp) with some inflammatory cells in the space of Lp, seen as black dots (group a, in the classification of general fibrosis; see Section 2.5.3). (2) Scarred untreated VF at three months showing compact connective tissue /fibrosis/ in deep Lp expanding into superficial muscle (Sm) and down into deep muscle (Dm). Plenteous inflammatory cells are spread in Sm and far into Dm (group d). (3) A hESC treated VF at three months showing minimal connective tissue in the Lp, slight loose connective tissue with limited inflammatory cells in Sm (group b). Arrows mark boarder between Lp and Sm. Scale bar is 100 μm. The images slightly decolorized to visualize the greyish fibrosis.

showing no more fibrosis than the average hESC treated VF. This suggests in this case an initiation of apoptosis already before six weeks, rather than a failed stem cell engraftment.

In all of the 12 VFs, the injected hESCs stayed in the injection site of the Lp and superficial vocal muscle. In none of the VFs, the hESCs were seen to migrate into the deeper part of the vocal muscle. This may be specific for VFs representing a relatively closed compartment [25].

None of the hESC treated VFs developed signs of malignancies or teratomas in this heterologous model. Notably, a human environment may behave differently. The general safety of hESCs however remains for future solutions.

In the study the immunosuppressant (Tacrolimus) was used to reduce the host versus graft reaction. The immunosuppressant was only administrated to the hESC treated VF individuals. This may have influenced the inflammatory reaction and also affected the hESCs. In a study with human mesenchymal stem cells comparing the healing process with and without immunosuppressant (Tacrolimus) we did not find any improved healing with immunosuppressant alone [26]. If the immunosuppressant has had any effect on the hESCs, it is reasonable to believe it has been negative. If so the hESCs have a potential to enhance their effect in an autologous environment.

Currently there is no ideal animal model representing the human VF. However, higher mammals as rabbits seem to have a rather similar VF healing capacity and pattern of scar formation as humans [24, 27, 28]. Therefore it seems reasonable to assume that the histologic and viscoelastic findings are transferable to the human VF.

5. Conclusion

We have previously shown that human embryonic stem cell (hESC) therapy to injured rabbit vocal folds (VFs) induces human tissue generation in the VFs with regained vibratory capacity of the VFs [19]. The present study shows that the hESC transplanted VFs maintained their restored function with improved vocal fold architecture as well as reduction in collagen and Lp thickness for three months although no residual hESCs or derivatives could be detected at this time point. At the three months, no malignancies or teratomas were revealed in the hESC treated VFs.

Conflict of Interests

The authors declare no competing interests regarding this paper.

Acknowledgments

The authors acknowledge Professor Frans Maurer, Department of Polymer Science & Engineering, Lund University, Sweden, for assistance with the viscoelastic measurements and Viktor Kempi, Ph.D., Department of Physiology, Östersund Hospital, Sweden, for statistic advice. The study was supported by The Swedish Larynx Foundation (Laryngfonden), the County Council of Jamtland Sweden Foundation (FoU Jamtlands Lans Landsting), and Karolinska Institutet.

References

[1] M. S. Benninger, D. Alessi, S. Archer et al., "Vocal fold scarring: current concepts and management," *Otolaryngology—Head and Neck Surgery*, vol. 115, no. 5, pp. 474–482, 1996.

[2] I. R. Titze, “Phonation threshold pressure: a missing link in glottal aerodynamics,” *Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2926–2935, 1992.

[3] C. N. Ford, D. M. Bless, and J. M. Loftus, “Role of injectable collagen in the treatment of glottic insufficiency: a study of 119 patients,” *Annals of Otology, Rhinology & Laryngology*, vol. 101, no. 3, pp. 237–247, 1992.

[4] L. S. Zaretsky, M. L. Shindo, and D. H. Rice, “Autologous fat injection for unilateral vocal fold paralysis,” *Annals of Otology, Rhinology and Laryngology*, vol. 105, no. 8, pp. 602–606, 1996.

[5] M. Remacle, M. Delos, G. Lawson, and J. Jamart, “Correcting vocal fold immobility by autologous collagen injection for voice rehabilitation: a short-term study,” *Annals of Otology, Rhinology and Laryngology*, vol. 108, no. 8, pp. 788–793, 1999.

[6] M. C. Neuenschwander, R. T. Sataloff, M. M. Abaza, M. J. Hawkshaw, D. Reiter, and J. R. Spiegel, “Management of vocal fold scar with autologous fat implantation: perceptual results,” *Journal of Voice*, vol. 15, no. 2, pp. 295–304, 2001.

[7] S. Hirano, D. M. Bless, B. Rousseau et al., “Prevention of vocal fold scarring by topical injection of hepatocyte growth factor in a rabbit model,” *Laryngoscope*, vol. 114, no. 3, pp. 548–556, 2004.

[8] D. K. Chhetri, C. Head, E. Revazova, S. Hart, S. Bhuta, and G. S. Berke, “Lamina propria replacement therapy with cultured autologous fibroblasts for vocal fold scars,” *Otolaryngology—Head and Neck Surgery*, vol. 131, no. 6, pp. 864–870, 2004.

[9] K. Tsunoda, K. Kondou, K. Kimitaka et al., “Autologous transplantation of fascia into the vocal fold: long-term result of type-1 transplantation and the future,” *The Laryngoscope*, vol. 115, supplement 108, pp. 1–10, 2005.

[10] J. K. Hansen, S. L. Thibeault, J. F. Walsh, X. Z. Shu, and G. D. Prestwich, “In vivo engineering of the vocal fold extracellular matrix with injectable hyaluronic acid hydrogels: early effects on tissue repair and biomechanics in a rabbit model,” *Annals of Otology, Rhinology and Laryngology*, vol. 114, no. 9, pp. 662–670, 2005.

[11] S. Hertegård, Å. Dahlqvist, and E. Goodyer, “Viscoelastic measurements after vocal fold scarring in rabbits—short-term results after hyaluronan injection,” *Acta Oto-Laryngologica*, vol. 126, no. 7, pp. 758–763, 2006.

[12] J. Gaston and S. L. Thibeault, “Hyaluronic acid hydrogels for vocal fold wound healing,” *Biomatter*, vol. 3, no. 1, Article ID e23799, 2013.

[13] S. Hertegård, J. Cedervall, B. Svensson et al., “Viscoelastic and histologic properties in scarred rabbit vocal folds after mesenchymal stem cell injection,” *Laryngoscope*, vol. 116, no. 7, pp. 1248–1254, 2006.

[14] B. J. Johnson, R. Fox, X. Chen, and S. Thibeault, “Tissue regeneration of the vocal fold using bone marrow mesenchymal stem cells and synthetic extracellular matrix injections in rats,” *Laryngoscope*, vol. 120, no. 3, pp. 537–545, 2010.

[15] B. Svensson, R. S. Nagubothu, J. Cedervall et al., “Injection of human mesenchymal stem cells improves healing of scarred vocal folds—analysis using a xenograft model,” *Laryngoscope*, vol. 120, no. 7, pp. 1370–1375, 2010.

[16] B. Svensson, S. R. Nagubothu, J. Cedervall et al., “Injection of human mesenchymal stem cells improves healing of vocal folds after scar excision—a xenograft analysis,” *Laryngoscope*, vol. 121, no. 10, pp. 2185–2190, 2011.

[17] W. Xu, R. Hu, E. Fan, and D. Han, “Adipose-derived mesenchymal stem cells in collagen—hyaluronic acid gel composite scaffolds for vocal fold regeneration,” *Annals of Otology, Rhinology and Laryngology*, vol. 120, no. 2, pp. 123–130, 2011.

[18] S. Ohno, S. Hirano, S.-I. Kanemaru et al., “Implantation of an atelocollagen sponge with autologous bone marrow-derived mesenchymal stromal cells for treatment of vocal fold scarring in a canine model,” *Annals of Otology, Rhinology and Laryngology*, vol. 120, no. 6, pp. 401–408, 2011.

[19] J. Cedervall, L. Åhrlund-Richter, B. Svensson et al., “Injection of embryonic stem cells into scarred rabbit vocal folds enhances healing and improves viscoelasticity: short-term results,” *Laryngoscope*, vol. 117, no. 11, pp. 2075–2081, 2007.

[20] O. Hovatta, M. Mikkola, K. Gertow et al., “A culture system using human foreskin fibroblasts as feeder cells allows production of human embryonic stem cells,” *Human Reproduction*, vol. 18, no. 7, pp. 1404–1409, 2003.

[21] M. P. Imreh, S. Wolbank, C. Unger et al., “Culture and expansion of the human embryonic stem cell line HS181, evaluated in a double-color system,” *Stem Cells and Development*, vol. 13, no. 4, pp. 337–343, 2004.

[22] S. L. Thibeault, S. D. Gray, D. M. Bless, R. W. Chan, and C. N. Ford, “Histologic and rheologic characterization of vocal fold scarring,” *Journal of Voice*, vol. 16, no. 1, pp. 96–104, 2002.

[23] R. W. Chan and M. L. Rodriguez, “A simple-shear rheometer for linear viscoelastic characterization of vocal fold tissues at phonatory frequencies,” *Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1207–1219, 2008.

[24] B. Rousseau, S. Hirano, R. W. Chan et al., “Characterization of chronic vocal fold scarring in a rabbit model,” *Journal of Voice*, vol. 18, no. 1, pp. 116–124, 2004.

[25] E. C. Wang, E. J. Damrose, A. H. Mendelsohn et al., “Distribution of class I and II human leukocyte antigens in the larynx,” *Otolaryngology—Head and Neck Surgery*, vol. 134, no. 2, pp. 280–287, 2006.

[26] B. Svensson, C. Svensson, S. R. Nagubothu et al., “Effects of immunosuppression on human mesenchymal stem cells in healing scarred rabbit vocal folds,” in *Restoration of Scarred Vocal Folds with Stem Cell Implantation*, B. Svensson, Ed., pp. 978–991, Karolonska Institutet, Stockholm, Sweden, 2011.

[27] B. Rousseau, S. Hirano, T. D. Scheidt et al., “Characterization of vocal fold scarring in a canine model,” *Laryngoscope*, vol. 113, no. 4, pp. 620–627, 2003.

[28] S. Hirano, S. Minamiguchi, M. Yamashita, T. Ohno, S.-I. Kanemaru, and M. Kitamura, “Histologic characterization of human scarred vocal folds,” *Journal of Voice*, vol. 23, no. 4, pp. 399–407, 2009.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 906356, 7 pages
http://dx.doi.org/10.1155/2015/906356

Research Article

A Fast Semiautomatic Algorithm for Centerline-Based Vocal Tract Segmentation

Anton A. Poznyakovskiy,[1] Alexander Mainka,[2,3] Ivan Platzek,[4] and Dirk Mürbe[2,3]

[1] *Department of Otorhinolaryngology, University Hospital Carl Gustav Carus, Technische Universität Dresden, 01062 Dresden, Germany*
[2] *Division of Phoniatrics and Audiology, Department of Otorhinolaryngology, University Hospital Carl Gustav Carus, Technische Universität Dresden, 01062 Dresden, Germany*
[3] *Voice Research Laboratory, Hochschule für Musik Carl Maria von Weber, 01067 Dresden, Germany*
[4] *Department of Radiology, University Hospital Carl Gustav Carus, Technische Universität Dresden, 01062 Dresden, Germany*

Correspondence should be addressed to Anton A. Poznyakovskiy; anton.poznyakovskiy@uniklinikum-dresden.de

Received 23 January 2015; Accepted 8 April 2015

Academic Editor: Haldun Oguz

Vocal tract morphology is an important factor in voice production. Its analysis has potential implications for educational matters as well as medical issues like voice therapy. The knowledge of the complex adjustments in the spatial geometry of the vocal tract during phonation is still limited. For a major part, this is due to difficulties in acquiring geometry data of the vocal tract in the process of voice production. In this study, a centerline-based segmentation method using active contours was introduced to extract the geometry data of the vocal tract obtained with MRI during sustained vowel phonation. The applied semiautomatic algorithm was found to be time- and interaction-efficient and allowed performing various three-dimensional measurements on the resulting model. The method is suitable for an improved detailed analysis of the vocal tract morphology during speech or singing which might give some insights into the underlying mechanical processes.

1. Introduction

The process of human voice production involves a complex interaction of different components and mechanisms. It involves the generation of a pulsating transglottal airflow which is filtered by the vocal tract (VT) resonator. The shape of the VT, the aeroacoustic cavity between the vocal folds and the lips, defines the formant frequencies and the frequency response of the filter which, in turn, defines vowels, consonants, and essential parts of voice timbre [1].

Magnetic resonance imaging (MRI) has become a promising technique for investigating the VT at a functional stage. MRI delivers images of high spatial resolution which can be analyzed in two dimensions within a single sagittal plane allowing for detailed analysis of dynamic VT adjustments during speech or even singing. For a detailed overview, see [2].

The elongated structure of the VT along with its curved shape makes segmentation feasible with the use of centerline-based methods. In medical imaging, these approaches have been successfully used in vessel segmentation [3]. The image stack is transformed into a coordinate system which is aligned with the centerline, and the cross sections are segmented. This makes a reduction of the 3D segmentation problem to a set of two-dimensional problems possible.

In many cases, the estimation of the centerline is a problem in itself, as in the aforementioned vessel segmentation task. There exist a broad variety of segmentation methods which compute the centerline on runtime, based on segmentation results of previous cross sections. Some of them postulate a circular [4] or else analytically defined [5] cross-section shape. Other authors make use of deformable models, subsequently skeletonizing the shape and extrapolating the resulting centerline, such as Li and Ourselin [6]. In our recent study on the cochlea [7], we also used deformable models in form of active contours [8] but predicted the centerline via mass centers of cross sections with the help of a Kalman filter [9].

With regard to segmentation, the geometrical structure of the VT poses similar challenges as vessel or cochlear segmentation. The potential of tomographic imaging techniques, especially MRI, to deliver three-dimensional (3D) image stacks of the VT has been exploited, for instance, to analyze area functions during sustained phonation [10, 11]. So far though, applications to the VT mostly included manual segmentation or relatively simple segmentation algorithms. Although publications on 3D modeling of the VT exist for quite some time [12, 13], they rely on manual segmentation.

A few centerline-based methods have been applied to the VT. However, most studies focus solely on the extraction of the area function by employing 2D methods such as threshold segmentation [14, 15]. A 3D centerline-based approach was presented by Vampola et al. [16], yet it still suggests segmenting individual cross sections manually. While manual segmentation requires little implementation error and is thereby a good tool for exploratory analysis, it has the drawback of being time-consuming, a factor that precludes its application to large sets of individual data.

There are several studies published which use region growing for the segmentation of VT [17, 18]. This method does not allow for the immediate studying of VT cross sections as it lacks a centerline. In order to measure cross-sectional areas, the centerline has to be constructed a posteriori, and the resulting segmented body cut along this centerline. This process might be nontrivial if there are bifurcations along the path. Indeed, VT has minor bifurcations: the piriform sinuses (*sinus piriformes*) and *vallecula*. On the other hand, vocal tract cross-sectional area has attracted considerable interest within the research community [19–23], since it plays an important role in the acoustics of speech and singing. Functional VT adjustments during phonation seem to be of importance not only for educational purposes but also for the medical field, where voice problems, for example, among professional voice users continue to bring about considerable socioeconomic burdens for the health care systems [24, 25].

Thus, in this study, we attempt to develop a VT segmentation algorithm which satisfies the following: (1) reduced operator interaction and time efforts and (2) direct data output on both VT cross sections and 3D geometry.

2. Materials and Methods

2.1. Image Data Acquisition and Sound Recording. A 43-year-old male test subject (height: 1.90 m, weight: 108 kg) was asked to produce a sustained vowel in a 3.0-T MR system (Verio; Siemens Medical Solutions, Erlangen, Germany) and to keep articulation constant during the recording. The task was specified regarding vowel quality (closed midback rounded vowel /o/ as in German "Boot"), pitch (220 Hz/ A3), and phonatory condition (speaking voice). The MRI recording was initiated as soon as the subject had started phonation. The MRI was performed with a 12-element head-neck coil. The applied MRI sequence was a volumetric interpolated breath-hold examination sequence with an acquisition time of about 12 s. A set of 52 sagittal slices of the whole VT was obtained. The parameter setting was the following: slice thickness 1.8 mm, repetition time 4.01 ms, echo time 1.22 ms, matrix 288 × 288, field of view 300 × 300 mm, and flip angle 9°. The obtained resolution of the images was 1.04 mm. Due to the known limitations of the MRI to visualize structures with low water content, the teeth were not detected in the MRI scan. For the segmentation of the oral cavity the segments were forced manually to remain between the tongue and the maxillary bone leaving out the space of the teeth.

An optical microphone unit (MO 2000 from Sennheiser) and a laptop PC running Audacity software (Dominic Mazzoni et al., http://audacity.sourceforge.net/, retrieved on January 20, 2015) were used for sound recording within the MRI facility. The acoustical recording was used to ensure vowel quality and pitch correctness.

2.2. Processing of Images and Coordinate Transform. The 52 sagittal images were stacked and scaled by a factor of 3.0 with ImageJ (National Institutes of Health, Bethesda, MD, USA) resulting in 156 images with a pixel size of 0.35 mm. This scaling was necessary to facilitate the later segmentation. Then, the images were resliced to the coronal view in order to fit the distance between slices to 0.35 mm and to obtain uniformly sized voxels. The reslice was repeated a second time with default settings to obtain sagittally oriented images.

For further image processing, the used algorithms were implemented in our software IPTools (freeware: http://www.uniklinikum-dresden.de/das-klinikum/kliniken-poliklini-ken-institute/hno/forschung/forschungslabor-gehor/links, last inspected January 20, 2015). In order to increase the grayscale gradient at the air-tissue border of the VT, the image stacks were filtered using anisotropic diffusion [26].

On the midsagittal image of the stack, the centerline was drawn (see Figure 1). This was done by defining node points such that the centerline intersects the tip of the uvula and the crossing of ventricular folds and the arytenoids. This procedure was established to ensure repeatability for application to other subjects while keeping the orientation of transformed slices near-orthogonal to the pharyngeal axis. Moreover, this provides a near-parallel slicing of the ventricular folds which is essential for calculating the area function in this region.

The centerline was piecewise interpolated between these nodes using cubic splines:

$$\mathbf{c}_i(t_c) = \mathbf{a}_{i,3}t^3 + \mathbf{a}_{i,2}t^2 + \mathbf{a}_{i,1}t^1 + \mathbf{a}_{i,0}, \quad (1)$$

with $\mathbf{c}_i = [c_{i,x}, c_{i,y}, c_{i,z}]^{\mathrm{T}}$ and $t_c \in \{0, \ldots, 1\}$. The four coefficients were defined by setting the positions at $t = 0$ and $t = 1$ to the coordinates of the nodes and equalizing the first derivatives $d\mathbf{v}/dt$ at these positions with adjacent spline segments.

Subsequently, the image stack was transformed along the curve with a fixed spacing of 1 pixel between new images resulting in a distance of 1.04 millimeters between centers of images. The center of each image of the new stack was set at the respective position on $\mathbf{c}(t_c)$. The coordinate axes were set

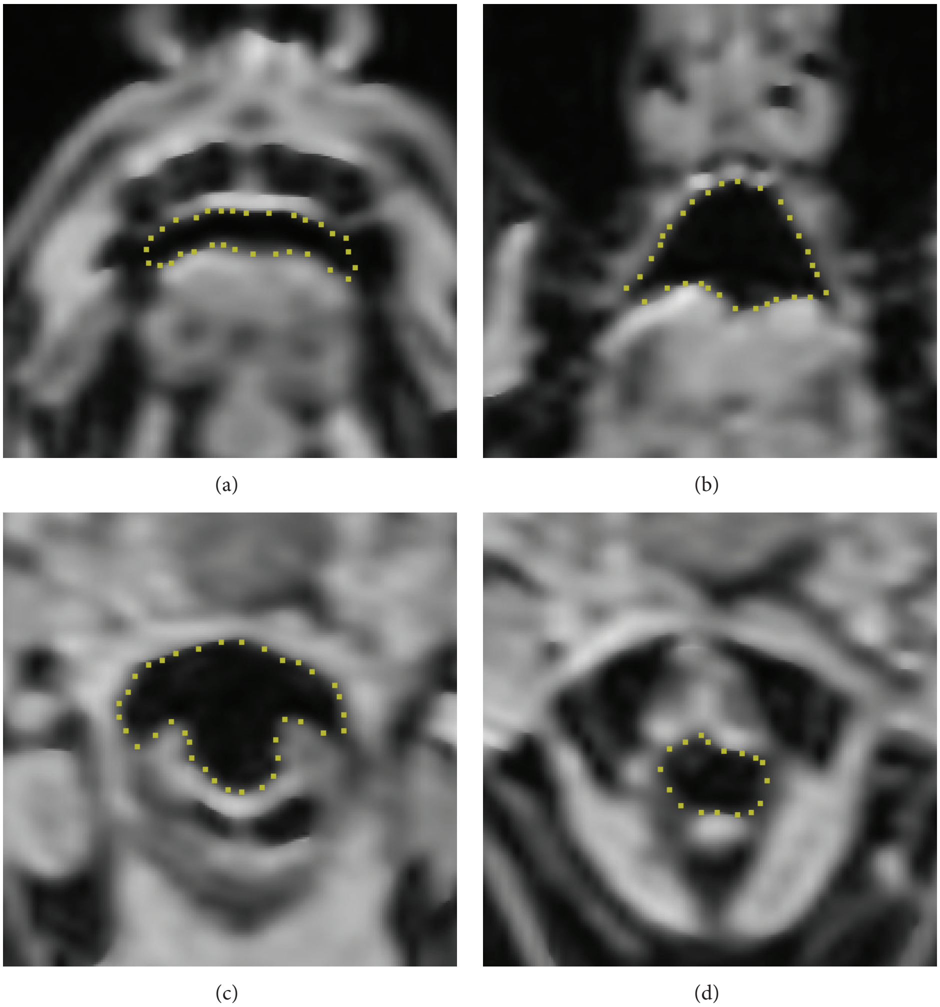


Figure 1: Four examples of vocal tract cross sections at different levels: (a) anterior oral cavity, (b) central oral cavity, (c) hypopharynx at inferior vallecula, and (d) larynx.

to the Frenet vectors of the curve at this position; specifically, the x-axis was set to the normalized binormal vector $\mathbf{b}$, the y-axis to the negated normalized normal vector $-\mathbf{n}$, and the z-axis to the normalized tangent vector $\mathbf{t}$. The negation of the normal vector was necessary since the coordinate system of an image stack was defined as left-hand.

From these definitions, we could derive the following affine transform matrix for each t_c:

$$\mathbf{T} = \begin{bmatrix} b_x & b_y & b_z & c_x \\ -\ddot{c}_x & -\ddot{c}_y & -\ddot{c}_y & c_y \\ \dot{c}_x & \dot{c}_y & \dot{c}_z & c_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

Note that the binormal vector $\mathbf{b} = [b_x, b_y, b_z]^{\mathrm{T}}$ is defined as

$$\mathbf{b} = \dot{\mathbf{c}} \times \ddot{\mathbf{c}}. \quad (3)$$

The transformation with a set of such matrices delivered a stack of several hundred images, where each image displayed the cross section in a manner feasible for 2D segmentation.

2.3. Segmentation. The segmentation was performed with a greedy variant of active contours [8]. A circular discrete starting contour was initialized with a center at user-defined position. Each node of the contour had the energy balance

$$E = E_{\mathrm{cont}} + E_{\mathrm{curv}} + E_{\mathrm{ext}} + E_{\mathrm{dev}}. \quad (4)$$

Here, E_{cont} is the contour energy which controls the expansive behavior of the contour. It is defined by a first-order derivative of the active contour curve function $\mathbf{v}(t_v)$:

$$E_{\mathrm{cont}} = \alpha \cdot \left|\nabla \mathbf{v}(t_v)\right|^2. \quad (5)$$

E_{curv} is the curvature energy which models the bending stiffness of the contour via the second-order derivative of the curve function:

$$E_{\mathrm{curv}} = \beta \cdot \left|\Delta \mathbf{v}(t_v)\right|^2. \quad (6)$$

The external energy E_{ext} provides the contour with edge detection and is proportional to the negated square norm of the grayscale gradient:

$$E_{\mathrm{ext}} = -\gamma \cdot \left|\nabla I(x, y)\right|^2. \quad (7)$$

Finally, the deviation energy E_{dev} provides cross links between contours on adjacent tomogram images:

$$E_{dev} = \delta \cdot d^4. \quad (8)$$

Here, d is the distance to the nearest node of the contour on previous image. In the first contour, this energy is set to zero. The Greek letters in (5)–(8) indicate user-defined parameters.

A search for the local minimum of the energy sum over all nodes was performed. This caused the contour to expand and adapt to the cross section of the vocal tract iteratively. When the contour stopped moving, the finding of local energy minimum was stated and the algorithm moved to the next image. Alternatively, the processing of the contour on an image stopped when the number of iterations exceeded a predefined threshold ($n = 30$). On the next image, a new contour was initialized with the end result of the previous image.

As the contour expanded, new points were added between any two neighboring points, whose spacing exceeded a predefined value s_{max}. Similarly, one of two points was deleted if the spacing became lower than s_{min} after any iteration. For this purpose, the values were defined as $s_{max} = 4s_{min}$ and $s_{min} = 3$ px.

The algorithm progress through the image stack was terminated at user's command. The accuracy of the segments was checked by an experienced laryngologist and corrected manually if needed.

The resulting segment stack was realigned with the information of the centerline curvature and visualized using Amira (FEI Visualization Sciences Group, Burlington, MA, USA).

3. Results and Discussion

Using the abovementioned methods, we were able to segment cross sections along the entire vocal tract (Figure 1). Total time used for segmentation (excluding filtering with anisotropic diffusion) was about 90 minutes.

To estimate algorithm objectivity, we performed the segmentation including centerline positioning twice. Out of the segmented data, we computed the area functions which are plotted against each other on Figure 2. The graphs appear to be highly correspondent down to the ventricular folds, scattering only in the region of the laryngeal ventricle. We assume that this is due to phonatory vibrations which cause blurring artifacts on VT borders and decrease the precision of segmentation.

We calculated reference volumes by superposition of segment areas. The total volume of the VT was estimated at 50528 mm^3. The volume of the oral cavity was 25965 mm^3, the combined volume of oropharynx and hypopharynx (segments from *uvula* to arytenoids including the *sinus piriformes*) was 21900 mm^3, and the volume of larynx (segments from arytenoids to glottis) was 2663 mm^3.

As a method of validation, we calculated the acoustic transfer function of the VT using PRAAT (Paul Boersma and David Weenink, http://www.fon.hum.uva.nl/praat/). It is displayed in Figure 3 and shows distinct peaks which are

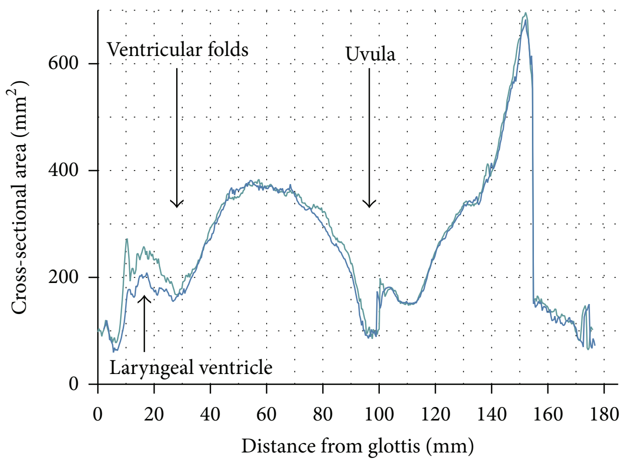


Figure 2: Cross-sectional area function obtained from two segmentations of the same data set.

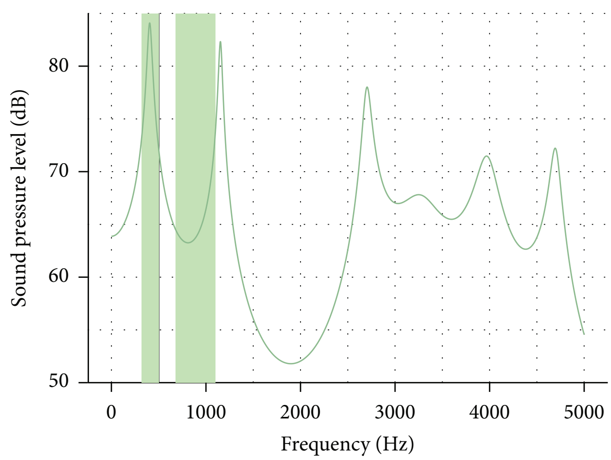


Figure 3: Transfer function computed with PRAAT based on the calculated area functions (Figure 2, green line). Light green stripes denote frequencies of the first two formants according to [27].

fairly coincident with the 1st and the 2nd formants for the utilized vowel [27].

With the obtained VT cross-section model, the geometrical analysis within all three spatial dimensions becomes feasible. The complete set of segmented cross sections, transformed from centerline-based coordinate system back to the global coordinate system (Figure 4(a)), is shown in Figure 4(b). Triangulation of this set yielded a surface mesh of the vocal tract shown on Figure 4(c). A close-up on the lower VT showing the high-detailed resolution of the larynx segmentation is displayed in Figure 4(d). This resulting mesh can serve as direct input for further numerical simulations using, for example, finite element modeling.

The accuracy of the model is dependent on the used MRI tomography device and the stability of the test subject over time. A natural challenge to the stability is the requirement to the subject to keep a constant articulatory setting and to maintain phonation during the entire recording procedure. This is necessary to produce enough images to

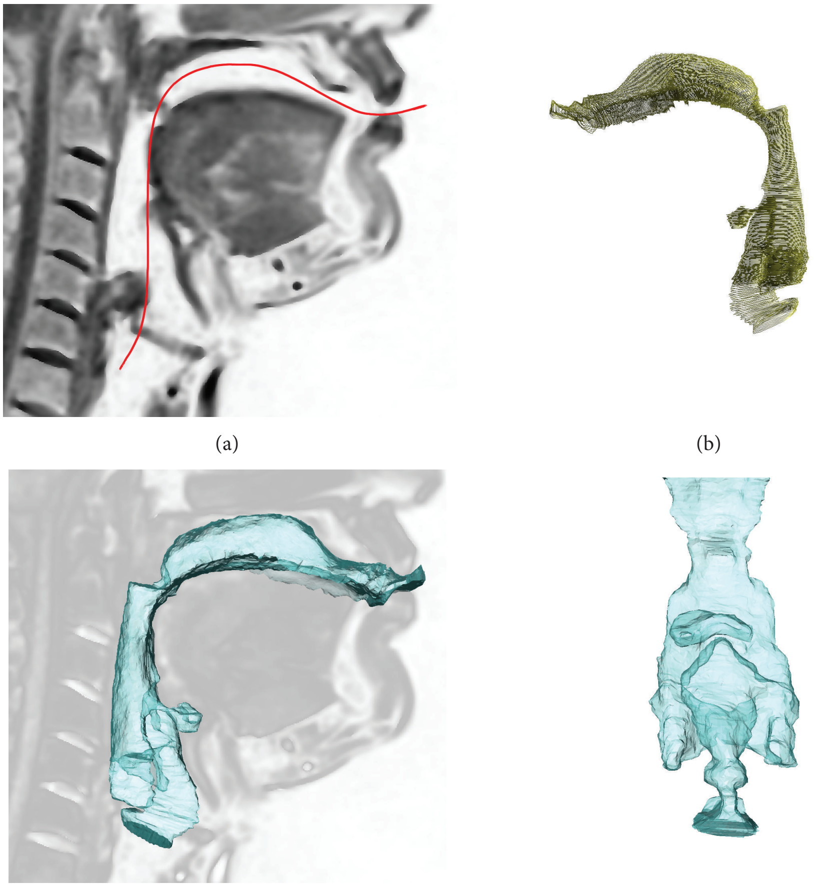


Figure 4: Vocal tract geometry modeling results: (a) centerline position within a midsagittal slice, (b) spatial alignment of the resulting cross-sectional segments, (c) generated surface mesh and its location within the image stack, and (d) detailed coronal view of the lower vocal tract.

cover the whole VT. There are hints that the movement artifacts of the jaw during sustained phonation are in the submillimeter order [28]. Yet this data represents only a single subject. A detailed discussion of the accuracy of MRI investigations prior to image processing is beyond the scope of this methodological study but ought to require further scientific attention.

The technical accuracy of the segmentation algorithm is constrained by the obtained resolution of MRI images. Since active contours cannot perform segmentation to a higher precision than 1 pixel, the uncertainty in border estimation corresponds to the resolution value, that is, 1.04 mm, which is well in range of state-of-the-art publications [15, 18]. Hence, the error in cross-section area estimation is between ca. 20 and 50 mm^2, depending on the area value.

The amount of input image data calls for a time-efficient and at least semiautomatic algorithm in order to reduce man-hours spent on segmentation. Unlike vessels whose cross-section shape does not vary much along the centerline, the VT has highly variable cross-section geometry (cf. Figure 1). A heuristic algorithm which is capable of handling arbitrary shapes is active contours which are widely used in biological imaging [29, 30]. It is based on the search of a steady-state shape of discrete deformable contour under influence of an equilibrium of internal and external forces. Internal forces govern the intrinsic properties of the contour, expansion, and stiffness. The external force creates a link to the image data, attracting the contour to regions with the greatest gradient of grayscale intensity.

4. Conclusion

By the presented approach for a complete MRI based 3D segmentation of the VT at a functional state, we were able to obtain a high-detailed model. The method could be used for answering questions regarding the physics and mechanical properties of the VT.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors acknowledge support by the German Research Foundation and the Open Access Publication Funds of the TU Dresden. The authors kindly acknowledge the help of Mario Fleischer who calculated the VT transfer function values. The authors also thank the test subject who participated in this study.

References

[1] G. Fant, *Acoustic Theory of Speech Production*, Mouton & Co, The Hague, The Netherlands, 1960.

[2] M. Echternach, M. Markl, and B. Richter, "Dynamic real-time magnetic resonance imaging for the analysis of voice physiology," *Current Opinion in Otolaryngology & Head & Neck Surgery*, vol. 20, no. 6, pp. 450–457, 2012.

[3] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes," *Medical Image Analysis*, vol. 13, no. 6, pp. 819–845, 2009.

[4] T. Behrens, K. Rohr, and H. S. Stiehl, "Robust segmentation of tubular structures in 3-D medical images by parametric object detection and tracking," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 33, no. 4, pp. 554–561, 2003.

[5] P. J. Yim, J. J. Cebral, R. Mullick, H. B. Marcos, and P. L. Choyke, "Vessel surface reconstruction with a tubular deformable model," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1411–1421, 2001.

[6] R. Li and S. Ourselin, "Accurate curvilinear modelling for precise measurements of tubular strcutures," in *Proceedings of the 7th International Conference on Digital Image Computing: Techniques and Applications*, pp. 243–252, Sydney, Australia, 2003.

[7] A. A. Poznyakovskiy, T. Zahnert, Y. Kalaidzidis et al., "A segmentation method to obtain a complete geometry model of the hearing organ," *Hearing Research*, vol. 282, no. 1-2, pp. 25–34, 2011.

[8] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[10] H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, and K. Honda, "Acoustic roles of the laryngeal cavity in vocal tract resonance," *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2228–2238, 2006.

[11] P. Clément, S. Hans, D. M. Hartl, S. Maeda, J. Vaissière, and D. Brasnu, "Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. A preliminary study," *Journal of Voice*, vol. 21, no. 5, pp. 522–530, 2007.

[12] T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels," *Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 799–828, 1991.

[13] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 537–554, 1996.

[14] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," *Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1037–1049, 2006.

[15] X. Zhou, J. Woo, M. Stone, J. L. Prince, and C. Y. Espy-Wilson, "Improved vocal tract reconstruction and modeling using an image super-resolution technique," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. EL439–EL445, 2013.

[16] T. Vampola, J. Horáček, and J. G. Švec, "FE modeling of human vocal tract acoustics. Part I: production of Czech vowels," *Acta Acustica united with Acustica*, vol. 94, no. 3, pp. 433–447, 2008.

[17] A. Wismueller, J. Behrends, P. Hoole, G. L. Leinsinger, M. F. Reiser, and P.-L. Westesson, "Human vocal tract analysis by in vivo 3D MRI during phonation: a complete system for imaging, quantitative modeling, and speech synthesis," *Medical Image Computing and Computer-Assisted Intervention*, vol. 11, no. 2, pp. 306–312, 2008.

[18] B. Delvaux, D. Howard, and D. A. Robin, "A new method to explore the spectral impact of the piriform fossae on the singing voice: benchmarking using MRI-based 3D-printed vocal tracts," *PLoS ONE*, vol. 9, no. 7, Article ID e102680, 2014.

[19] P. Perrier, L.-J. Boe, and R. Sock, "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients," *Journal of Speech and Hearing Research*, vol. 35, no. 1, pp. 53–67, 1992.

[20] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions for an adult female speaker based on volumetric imaging," *Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 471–487, 1998.

[21] A. F. Brito, J. A. Redinz, and J. A. Plascak, "Dynamics of rough surfaces generated by two-dimensional lattice spin models," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 75, no. 4, Article ID 046106, 2007.

[22] J. Brito, "Genetic learning of vocal tract area functions for articulatory synthesis of Spanish vowels," *Applied Soft Computing Journal*, vol. 7, no. 3, pp. 1035–1043, 2007.

[23] Y. C. Kim, J. M. Kim, M. I. Proctor et al., "Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging," in *Proceedings of the ISCA Workshop on Speech Production in Automatic Speech Recognition (SPASR '13)*, Lyon, France, August 2013.

[24] K. Verdolini and L. O. Ramig, "Review: occupational risks for voice problems," *Logopedics Phoniatrics Vocology*, vol. 26, no. 1, pp. 37–46, 2001.

[25] R. J. Ruben, "Redefining the survival of the fittest: communication disorders in the 21st century," *Laryngoscope*, vol. 110, pp. 241–245, 2000.

[26] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[27] W. Sendlmeier and J. Seebode, *Formantkarten des Deutschen Vokalsystems*, Technical University of Berlin, Berlin , Germany, 2006, http://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf.

[28] D. Aalto, J. Malinen, M. Vainio, J. Saunavaara, and P. Palo, "Estimates for the measurement and articulatory error in MRI data from sustained vowel production," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS '11)*, Hong Kong, August 2011.

[29] D. J. Kang and I. S. Kweon, "A fast and stable snake algorithm for medical images," *Pattern Recognition Letters*, vol. 20, no. 5, pp. 507–512, 1999.

[30] F. L. Valverde, N. Guil, and J. Muñoz, "Segmentation of vessels from mammograms using a deformable model," *Computer Methods and Programs in Biomedicine*, vol. 73, no. 3, pp. 233–247, 2004.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 259239, 13 pages
http://dx.doi.org/10.1155/2015/259239

Research Article

Modulation Spectra Morphological Parameters: A New Method to Assess Voice Pathologies according to the GRBAS Scale

Laureano Moro-Velázquez, Jorge Andrés Gómez-García, Juan Ignacio Godino-Llorente, and Gustavo Andrade-Miranda

ETSIST, Universidad Politécnica de Madrid, Campus Sur, Carretera de Valencia km 7, 28031 Madrid, Spain

Correspondence should be addressed to Laureano Moro-Velázquez; laureano.moro@upm.es

Received 23 January 2015; Revised 4 May 2015; Accepted 4 May 2015

Academic Editor: Adam Klein

Disordered voices are frequently assessed by speech pathologists using perceptual evaluations. This might lead to problems caused by the subjective nature of the process and due to the influence of external factors which compromise the quality of the assessment. In order to increase the reliability of the evaluations, the design of automatic evaluation systems is desirable. With that in mind, this paper presents an automatic system which assesses the Grade and Roughness level of the speech according to the GRBAS perceptual scale. Two parameterization methods are used: one based on the classic Mel-Frequency Cepstral Coefficients, which has already been used successfully in previous works, and other derived from modulation spectra. For the latter, a new group of parameters has been proposed, named *Modulation Spectra Morphological Parameters*: MSC, DRB, LMR, MSH, MSW, CIL, PALA, and RALA. In methodology, PCA and LDA are employed to reduce the dimensionality of feature space, and GMM classifiers to evaluate the ability of the proposed features on distinguishing the different levels. Efficiencies of 81.6% and 84.7% are obtained for Grade and Roughness, respectively, using modulation spectra parameters, while MFCCs performed 80.5% and 77.7%. The obtained results suggest the usefulness of the proposed *Modulation Spectra Morphological Parameters* for automatic evaluation of Grade and Roughness in the speech.

1. Introduction

With the aim of diagnosing and evaluating the presence of a voice disorder clinicians and specialists have developed different assessment procedures [1] such as exploration using laryngoscopic techniques, acoustic analysis, or perceptual evaluations. The latter is widely used by clinicians to quantify the extent of a dysphony. Some well-known perceptual evaluation procedures are the Buffalo Rating Voice Profile [2], Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) [3], and GRBAS [4]. The main problem with perceptual analysis is the high intra/interrater variability [5, 6] due to the subjectivity of the assessment in which the experience of the evaluator, his/her physical fatigue, mental condition, and some other factors are involved. Hence, means such as acoustic analysis based on signal processing might be valuable in clinical scenarios, providing objective tools and indices which can directly represent the level of affection or at least help clinicians to make a more reliable and less subjective perceptual assessment. This noninvasive technique can complement and even replace other invasive methods of evaluation.

Besides, the large amount of improvements in the field of speech signal processing is addressed mostly to areas such as speech or speaker recognition. Many of these advances are being transferred to biomedical applications for clinical purposes; some recent examples are related to different uses such as telemonitoring of patients [7], telerehabilitation [8], or clinical-support systems [9]. However, there is a substantial quantity of research to be done for further enhancements. Roughly speaking, most of the studies in this field can be divided into three main categories: the first one is focused on developing automatic detectors of pathological voices

[10–14] capable of categorizing voices between normal and pathological; the second group works with classifiers of pathologies [11, 15, 16] which consists in determining the speech disorder of the speaker using the acoustic material; and the third and last group aims to evaluate and assess the voice quality [8, 17–23]. The present study can be framed in the third mentioned category, highlighting the fact that the main goal is the development of new parameterization methods.

The essential common characteristic of all the automatic systems found in the literature is the need to extract a set of parameters from the acoustic signal to accomplish a further classification task. Regarding these parameters, some works use *amplitude perturbations* such as Shimmer or Amplitude Tremor Intensity Index (ATRI) [24–26] as input features while others are centered on *frequency perturbations* using Jitter [8, 25, 26], frequency and cepstral-based analysis [8, 13, 14, 16, 17, 27, 28], F_0 Tremor Intensity Index (F_0TRI) [24, 26], or Linear Predictive Coding (LPC) [29]. *Noise-based parameters* [30, 31] and *nonlinear analysis features* [12, 18, 25, 32] are likewise widely used in this kind of automatic detectors. Moreover, other varieties of feature-extraction techniques such as biomechanical attributes or signatures can be applied for the same purposes [10].

Focusing on the third kind of the aforementioned categories of detectors, those assessing the quality of voice, some are employed for simulating a perceptual evaluation such as GRBAS. For instance, several classification methods were used in [19, 20] to study the influence of the voice signal bandwidth in perceptual ratings and automatic evaluation of GRBAS Grade (G) trait using cepstral parameters (Linear Frequency Spectrum Coefficients and Mel-Frequency Spectrum Coefficients). Efficiencies up to 80% were obtained using Gaussian Mixture Models (GMM) classifiers and leave-x-out [33] cross-validation techniques. Similar parameterization methods were used in [9] to automatically evaluate G with a Back-and-Forth Methodology in which there is feedback between the human experts that rated the database and the automatic detector, and vice versa. On [22] a group of 92 features comprising different types of measurements such as noise, cepstral and frequency parameters among others were used to detect GRBAS Breathiness (B). After a reduction to a four-dimensional space, a 77% of efficiency was achieved using a 10-fold cross-validation scheme. Authors in [34] fulfilled a statistical study of acoustic measures provided by two commonly used analysis systems, *Multidimensional Voice Analysis Program* by Kay Elemetrics and *Praat* [35] obtaining good correlations for G and B traits. On [21] Mel-Frequency Spectrum Coefficients (MFCCs) were utilized obtaining 68% and 63% of efficiency for Grade and Roughness (R) traits, respectively, using Learning Vector Quantization (LVQ) methods for the pattern recognition stage but without any type of cross-validation techniques. The review of the state of the art reports that only [36] has used the same database and perceptual assessment used in the present study. The mentioned work proposed a set of complexity measurements and GMM to emulate a perceptual evaluation of all GRBAS traits, but its performance does not surpass 56% for G or R.

In general, results seldom exceed 75% of efficiency; hence, there is still room for enhancement in the field of voice quality automatic evaluation. Thus, new parameterization approaches are needed and the use of Modulation spectrum (MS) emerges as a promising technique. MS provides a visual representation of sound energy spread in acoustic and modulation axes [37, 38] supplying information about perturbations related to amplitude and frequency modulation of the voice signal. Numerous acoustic applications use these spectra to extract features from audio signals from which some examples can be found in [39–42]. Although there are few publications centered in the characterization of dysphonic voices using this technique [11, 12, 23, 43], it can be stated that MS has not been studied deeply in the field of the detection of voice disorders and specially as a source of information to determine patient's degree of pathology. Some of the referred works have used MS to simulate an automatic perceptual analysis but, to the best of our knowledge, none of them offer well-defined parameters with a clear physical interpretation but transformations of MS which are not easily interpretable, limiting their application in the clinical practice.

The purpose of this work is to provide new parameters obtained from MS in a more reasoned manner, making them more comprehensible. The use of this spectrum and associated parameters as support indices is expected to be useful in medical applications since they provide easy-to-understand information compared to others such as MFCC or complexity parameters, for instance. The new parameterization proposed in this work has been used as the input to a classification system that emulates a perceptual assessment of voice following the GRBAS scale in G and R traits. These two traits have been selected over the other three (Aesthenia (A), Breathiness, and Strain (S)) since its assessment seems to be more reliable. De Bodt et al. [5] point that G is the less unambiguously interpreted and R has an intermediate reliability on its interpretation. These conclusions are coherent with those exposed in [44, 45]. Similar findings are revealed in [6] which considers R as one of the most reliable traits when using sustained vowel $/ah:/$ as source of evaluation. It is convenient to specify that each feature of the GRBAS scale ranges from 0 to 3, where 0 indicates no affection, 1 slightly affected, 2 moderately affected, and 3 severely affected voice regarding the corresponding trait. Thus evaluating according to this perceptual scale means developing different 4-class classifiers, one for each trait.

In this work, the results obtained with the proposed MS-based parameters are compared with a classic parameterization used to characterize voice in a wide range of applications: Mel-Frequency Cepstral Coefficients [46]. MFCCs have been traditionally used for speech and speaker recognition purposes since the last two decades and many works use these coefficients to detect voice pathologies with a good outcome.

The paper is organized as follows: Section 2 develops the theoretical background of modulation spectra features. Section 3 introduces the experimental setup and describes the database used in this study. Section 4 presents the obtained results. Lastly, Section 5 presents the discussions, conclusions, and future work.

2. Theoretical Background

2.1. Modulation Spectra. This study proposes a new set of parameters based on MS to characterize the voice signal. MS provides information about the energy at modulation frequencies that can be found in the carriers of a signal. It is a three-dimensional representation where abscissa usually represents modulation frequency, ordinate axis depicts acoustic frequency, and applicate, acoustic energy. This kind of representation allows observing different voice features simultaneously such as the harmonic nature of the signal and the modulations present at fundamental frequency and its harmonics. For instance, the presence of tremor, understood as low frequency perturbations of the fundamental frequency, can be easily noticeable since it implies a modulation of pitch as an usual effect of laryngeal muscles improper activity. Other modulations associated with fundamental or harmonic frequencies could indicate the presence of a dysfunction of the phonatory system. Some examples can be found in [11].

To obtain MS, the signal is filtered using a short-time Fourier transform (sTFT) filter bank whose output is used to detect amplitude and envelope. This outcome is finally analyzed using FFT [47] producing a matrix E where MS values at any point can be represented as $E(f_a, f_m)$. The columns at E (fixed f_m) are modulation frequency bands, and rows (fixed f_a) are acoustic frequency bands. Therefore, a can be interpreted as the index of acoustic bands and m, the index of modulation bands while f_a and f_m are the central frequencies of the respective bands. Due to the fact that values $E(f_a, f_m)$ have real and imaginary parts, the original matrix can be represented using the modulus $|E|$ and the phase $\arg(E)$ of the spectrum. Throughout this work, the MS has been calculated using the Modulation Toolbox library version 2.1 [48]. Some different configurations can be used to obtain E, where the most significant degrees of freedom are the use of coherent or noncoherent (Hilbert envelope) [49] modulation, the number of acoustic bands, and acoustic and modulation frequency ranges. The three-dimensional phase unwrapping techniques detailed in [50] are used to solve the phase ambiguity problems which appear when calculating $\arg(E(f_a, f_m))$.

Figure 1 shows an example of MS extracted from two different voices on which the voice of a patient with gastric reflux, edema of larynx, and hyperfunction exhibits a more spread modulation energy in comparison to a normal voice.

However, one of the principal drawbacks of MS is that it provides a large amount of information that can not be easily processed automatically due to limitations of the existing pattern recognition techniques and voice disorders databases available. In this sense, MS matrices have to be processed to obtain a more compact but precise enough representation of the represented speech segments. Thus, after obtaining the MS, some representative parameters are extracted to feed a further classification stage. With this in mind, a new group of *Morphological Parameters* based on MS is proposed in this work: centroids [51] (MSC), dynamic range per band (DRB), Low Modulation Ratio (LMR), Dispersion Parameters (CIL, PALA, and RALA), Contrast (MSW), and Homogeneity (MSH). All these parameters use the MS modulus as input source, except the last two which also use the phase.

2.1.1. Centroids (MSC) and Dynamic Range per Band (DRB). Centroids provide cues about the acoustic frequency that represents the central energy or the energy center at each modulation band. To obtain MSC, MS modulus is reduced to an absolute number of modulation bands usually ranging from 4 to 26, each containing information about the modulation energy in that band along the acoustic frequency axis. Once the reduced MS is computed, centroids are calculated following the expression

$$\mathrm{MSC}(f_m) = \frac{\sum_a f_a \cdot |E(f_a, f_m)|}{f_{\mathrm{pitch}} \cdot \sum_a |E(f_a, f_m)|}, \tag{1}$$

where f_a and f_m represent the central frequency of the acoustic and modulation bands, respectively, and f_{pitch} is the pitch frequency.

As a matter of example, Figure 2 depicts a representation of MSC extracted from a MS.

Once MS is reduced to a small number of modulation bands, the dynamic range is calculated for every band (DRB) as the difference between the highest and the lowest levels in the band. These parameters provide information about the flatness of the MS depending on the modulation frequency.

2.1.2. Low Modulation Ratio (LMR). LMR, expressed in dB, is the ratio between energy in the first modulation band $\varepsilon(f_{a(f_{\mathrm{pitch}})}, f_1)$ at acoustic frequency f_{pitch} and the global energy in all modulation bands covering at least from 0 to 25 Hz at acoustic frequency f_{pitch}, $\varepsilon(f_{a(f_{\mathrm{pitch}})}, f_{m(25\,\mathrm{Hz})})$. Its calculation is carried out according to the following expressions. These bands are represented in Figure 3:

$$\mathrm{LMR} = 10 \cdot \log\left(\frac{\varepsilon\left(f_{a(f_{\mathrm{pitch}})}, f_1\right)}{\varepsilon\left(f_{a(f_{\mathrm{pitch}})}, f_{m(25\,\mathrm{Hz})}\right)}\right) \tag{2}$$

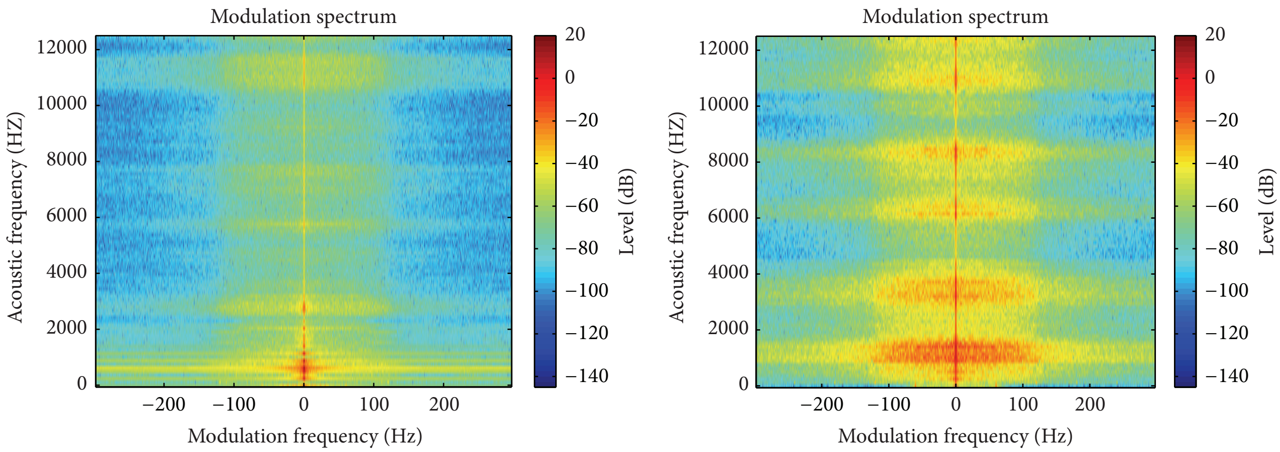
being

$$\varepsilon(f_a, f_k) = \sum_{m=1}^{k} |E(f_a, f_m)|^2, \tag{3}$$

where $a(f_{\mathrm{pitch}})$ is the index of the acoustic band including pitch frequency and $m(25\,\mathrm{Hz})$, the index of the modulation band including 25 Hz.

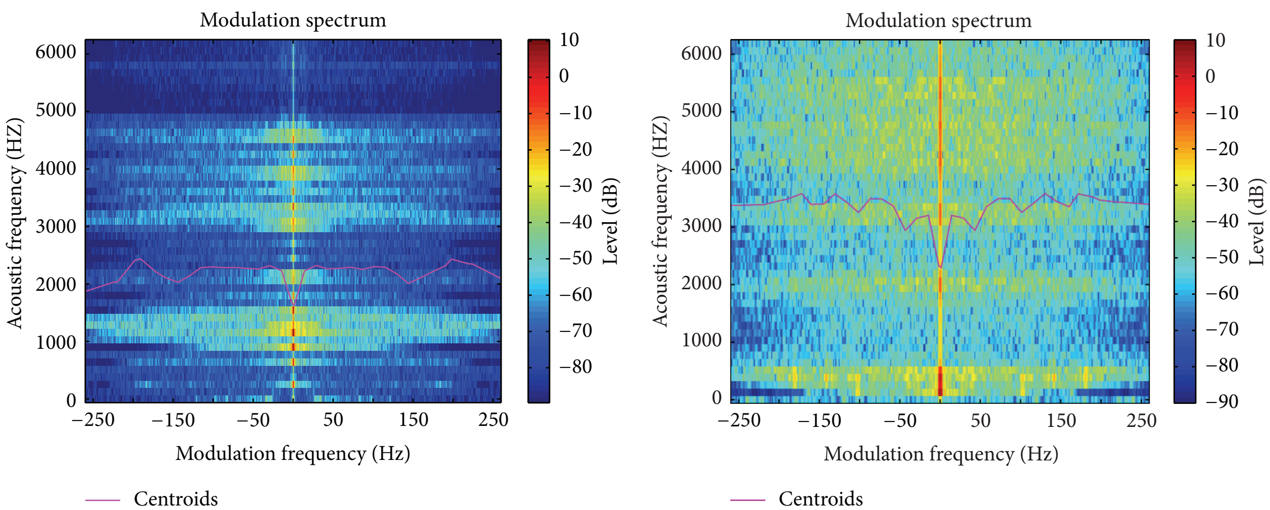
The 0–25 Hz band has been selected to represent all possible cases of tremor and low frequency modulations around pitch frequency [52, 53].

2.1.3. Contrast and Homogeneity. Representing MS (modulus or phase) as two-dimensional images let observe that pathological voices usually have more complex distributions. Images related to normal voices are frequently more homogeneous and present less contrast, as can be seen in Figure 1. Accordingly, Homogeneity and Contrast are used as two MS features since they provide information about the existence of voice perturbations.



(a) (b)

Figure 1: MS modulus of a normal voice (a) and pathological voice of a patient with gastric reflux, edema of larynx, and hyperfunction (b).



(a) (b)

Figure 2: MS centroids of a normal voice (a) and a pathological voice (b) of a patient with gastric reflux, keratosis, and laryngocele.

Homogeneity is computed using the Bhanu method described by the following expression, as stated in [54]:

$$\mathrm{MSH} = \sum_{a}\sum_{m}\left[E(f_a, f_m) - \overline{E(f_a, f_m)_{3\times3}}\right]^2, \quad (4)$$

with MSH being the MS Homogeneity value; $E(f_a, f_m)$ the modulation spectra computation (modulus or phase) at point (f_a, f_m); and $\overline{E(f_a, f_m)_{3\times3}}$ the average value in a 3×3 window centered at the same point.

Contrast is computed using a variation of the Weber-Fechner contrast relationship method described by the following expression as stated in [54]:

$$\mathrm{MSW}(f_a, f_m) = \sum_{a'}\sum_{m'} C_{f_a, f_m}, \quad (5)$$

where

$$C_{f_a, f_m} = \frac{|E(f_a, f_m) - E(f_{a'}, f_{m'})|}{|E(f_a, f_m) + E(f_{a'}, f_{m'})|} \quad (6)$$

representing $(f_{a'}, f_{m'})$ the vertical and horizontal adjacent points to (f_a, f_m). The global MSW is considered the sum of all points in $\mathrm{MSW}(f_m, f_a)$ divided by the total number of points to normalize.

The MS used to calculate MSH and MSW at each point of the matrix is represented in Figure 3.

2.1.4. Dispersion Parameters. As MS differs from normal to pathological voices, changes in the histograms of MS modulus reflect the effects of a dysfunction in a patient's voice. A short view to the MS permits to observe that voices with high G and R traits usually have a larger number of points with levels above the average value of $|E(f_m, f_a)|$. The level of

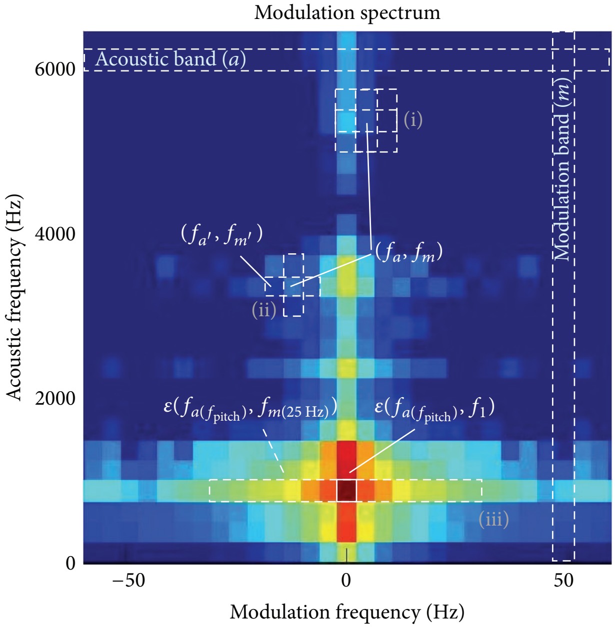


FIGURE 3: Points of the MS matrix used to obtain MSH (i), MSW (ii), and LMR (iii).

these points can be interpreted as the dispersion of the energy present in the central modulation band (0 Hz) towards side bands respecting the case of a normal voice.

With this in mind, three Morphological Parameters are proposed to measure such dispersion effect: *Cumulative Intersection Level (CIL)*, *Normalized Number of Points above Linear Average (PALA)*, and *Ratio of Points above Linear Average (RALA)*. CIL is the intersection between the histogram increasing and decreasing cumulative curves. Histogram is processed from MS modulus in logarithmic units (dB). As shown in Figure 4, CIL tends to be higher in pathological than in healthy voices. In that case, the difference is 19 dB. On the other hand, PALA is the number of points in MS modulus which are above average (linear units) divided by the total number of points of MS. RALA is quite similar to PALA but in this case it represents the ratio of points in MS modulus which are over the average and the number of points which are above this average instead of the total number of points in $E(f_a, f_m)$. Calculation of PALA and RALA is detailed in the following expressions:

$$\text{PALA} = \frac{\text{NA}}{\text{NT}}, \qquad \text{RALA} = \frac{\text{NA}}{\text{NB}} \tag{7}$$

being

$$\text{NA} = \sum_{f_a}\sum_{f_m} \gamma(f_a, f_m), \qquad \text{NB} = \sum_{f_a}\sum_{f_m} 1 - \gamma(f_a, f_m),$$

$$\gamma(f_a, f_m) = \begin{cases} 1 & |E(f_a, f_m)| \geq \overline{|E|} \\ 0 & |E(f_a, f_m)| < \overline{|E|}, \end{cases} \tag{8}$$

where $\overline{|E|}$ is the MS modulus average, NA the number of points above $\overline{|E|}$, NB the number of points below $\overline{|E|}$, and NT the total number of points in $E(f_a, f_m)$.

In the cases in which the number of points above linear average increases, the difference between PALA and RALA increases too as the denominator in PALA stays constant and the denominator in RALA decreases. Figure 5 represents these points in a healthy and a pathological voice. It is noticeable that, as expected, the MS of dysphonic voices presents more points above the modulus average.

3. Experimental Setup

3.1. Database. The Kay Elemetrics Voice Disorders Database recorded by the Massachusetts Eye and Ear Infirmary Voice Laboratory (MEEI) was used for this study [55] due to its commercial availability. The database contains recordings of the phonation of the sustained vowel /*ah* : / (53 normal, 657 pathological) and utterances corresponding to continuous speech during the reading of the "Rainbow passage" (53 normal, 661 pathological). The sample frequency of the recordings is 25 kHz with a bit depth of 16 bits. From the original amount of speakers recorded in the database, a first corpus of 224 speakers was selected according to the criteria found in [56] being named henceforward as the *original subset*. The utterances corresponding to the sustained vowel and the continuous speech recordings were used to rate *G* and *R* for each patient according to the GRBAS scale. The degree of these traits has been estimated three times by two speech therapists. One of them evaluated the whole database once, and the second one performed the assessment twice in two different sessions. Regarding this study, only the sustained vowels are considered. With the aim of obtaining more consistent labels, two reduced subsets of 87 and 85 audio files for *G* and *R*, respectively, were considered. Those files are chosen from the initial corpus of 224 recordings on the basis of selecting only those whose labeling was in a total agreement for the three assessments making up the *G* and *R agreement subsets*. This reduction was performed to avoid modeling inter/intraraters variability inherent to the process of assigning perceptual labels to each speaker. In any case, all tests were performed for the three subsets to provide evidences about such reduction. Some statistics of database are shown in Table 1.

With the aim of sharing relevant information and to promote a more reliable comparison of techniques and results, the names of the recordings extracted from MEEI corpus that were used for this study along with their respective *G* and *R* levels are included in Appendix, Table 6.

3.2. Methodology. One of the purposes of this work is to test a new source of parameters to characterize voice perturbations by replicating clinician's *G* and *R* perceptual evaluations.

Table 1: Subsets statistics.

Subset name	Number of subjects		Age range		Average age	
	Female	Male	Female	Male	Female	Male
Original (226 files)	90	134	21–52	26–59	35.8 ± 8.2	39.9 ± 9.1
Agreement-*G* (87 files)	52	35	24–52	26–58	36.6 ± 7.6	39.5 ± 9.7
Agreement-*R* (85 Files)	51	34	22–52	26–58	35.4 ± 7.6	37.9 ± 9.2

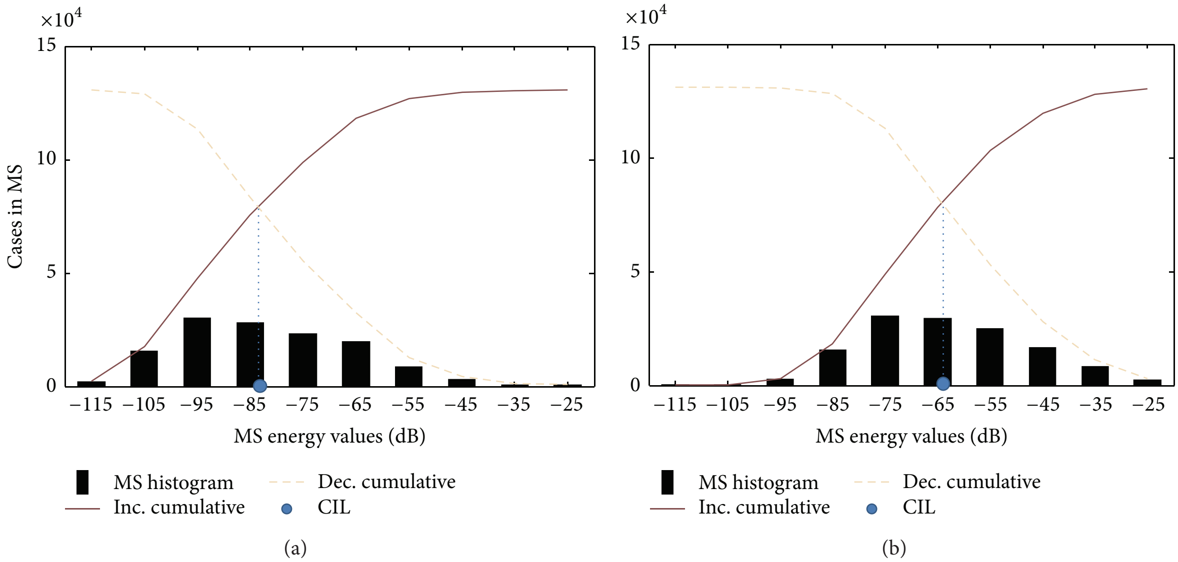


Figure 4: CIL calculation in a normal voice (top) and a pathological voice (bottom) diagnosed of bilateral laryngeal tuberculosis.

So as to quantify the contribution of this new approach, a baseline parameterization has been established to compare with the novel one. Consequently, all tests are performed using the parameters of the baseline system (MFCCs) and the MS *Morphological Parameters*. A large number of tests were accomplished to find the best setting, modifying the number of centroids or the frame duration among other degrees of freedom.

The methodology employed in this paper is shown in Figure 6, while each one of its stages is explained next. Basically, it is the classical supervised learning arrangement, which can be addressed using either classification or regression techniques. For the sake of simplicity and to concentrate on the novel parameterization approach, a simple Gaussian Mixture Model (GMM) classification back-end was employed to recognize the presence of the perturbations in the voice signal which presumably would produce high levels of *G* and *R* during perceptual analysis.

3.2.1. Characterization. Two parameterization approaches are considered in this study: MFCCs and MS Morphological Parameters. The MFCCs are the ground of the baseline system and were used for comparison due to their wide use in speech technology applications.

The MFCCs are calculated following a method based on the human auditory perception system. The mapping between the real frequency scale (Hz) and the perceived frequency scale (mels) is approximately linear below 1 kHz and logarithmic for higher frequencies. Such mapping converts real into perceived frequency. In this work MFCCs are estimated using a nonparametric FFT-based approach. Coefficients are obtained by calculating the Discrete Cosine Transform (DCT) over the logarithm of the energy in several frequency bands. The bandwidth of the critical band varies according to the perceived frequency. Each band in the frequency domain is bandwidth dependant of the filter central frequency. The higher the frequency is, the wider the bandwidth is. To obtain these parameters, a typical setup of 30 triangular filters and cepstral mean subtraction was used. Their computation is carried out over speech segments framed and windowed using Hamming windows overlapped 50%. Duration of frames oscillates from 20 to 100 ms in 20 ms steps. For the sake of comparison the number of MFCCs ranges from 10 to 22 coefficients. 0'th order cepstral coefficient is removed.

Regarding the MS Morphological Parameters, each signal is also framed and windowed using Hamming windows overlapped 50%. The window lengths are varied in the range of 20–200 ms in 20 ms steps. The feature vector extracted from MS is composed of the following: MSC, DRB, LMR, MSW, MSH, CIL, PALA, and RALA. The number of bands to obtain centroids and dynamic range features is varied in the range of $[6, 22]$ with a step size of 2. Considering that MSW and MSH provide two features each (one for modulus and other for phase), the feature vector corresponding to each frame ranges from 20 to 44 values before using data reduction

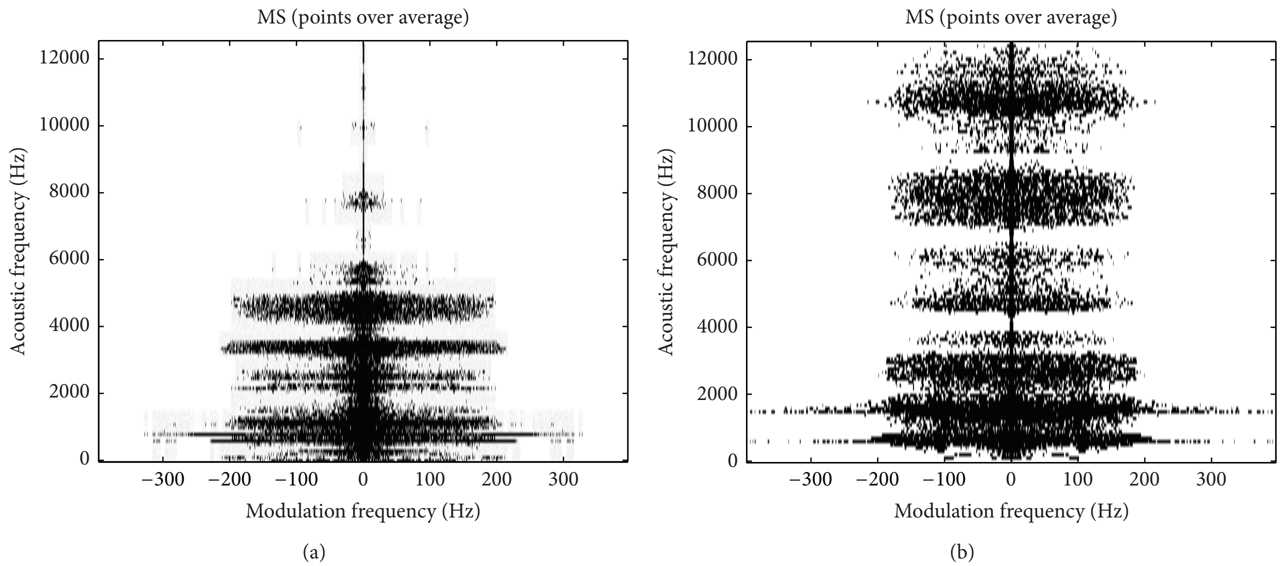


Figure 5: Points above (black) and below (white) modulus average in MS for a normal voice (a) PALA = 0.11, RALA = 0.12, and a pathological voice due to bilateral laryngeal tuberculosis (b) PALA = 0.21, RALA = 0.27.

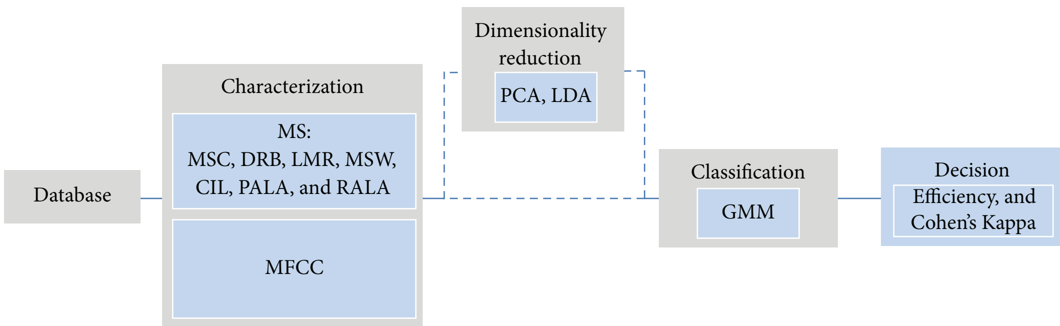


Figure 6: Outline of the automatic detector presented in the paper.

techniques. Both, coherent and noncoherent modulation (Hilbert envelope) were used for testing separately. Acoustic frequency span [0–12.5 kHz] is divided into 128 bands and maximum modulation frequency varied from 70 to 500 Hz to allow different configurations during tests.

In addition, first derivative (Δ) and second derivative ($\Delta\Delta$), representing the speed and acceleration in the changes of every characteristic, are added to the features in order to include interframe attributes [46]. The calculation of Δ and $\Delta\Delta$ was carried out employing finite impulse response filters using a length of 9 samples to calculate Δ and 3 in the case of $\Delta\Delta$.

All these features are used to feed a subsequent classification phase in two different ways depending on the test: some experiments are accomplished using features as they are obtained, and others use a reduced version to relieve *the curse of dimensionality* effect.

In the dimensionality reduction stage, PCA [57] and LDA [58] techniques are used varying the dimension of the feature vectors used for classification. In the case of LDA, all feature vectors are reduced to a 3-dimensional space. Concerning PCA, reduction ranges from 80 to 95%. With respect to these techniques, only the training data set is used to obtain the models which are employed to reshape all the data: training and test data sets. This process is repeated for every iteration of the GMM training-test process carried out for validation. The dimensionality reduction is applied for both MS Morphological Parameters and MFCCs features with and without derivatives separately.

3.2.2. Validation. Following the characterization, a *Leave-One-Out (LOO)* cross-validation scheme [33] was used for evaluating the results. On this scheme one file is considered for testing and the remaining files of the database are used as training data, generating what is called a *fold*. As a result, there are as many *folds* as number of files, and each of them will provide a classification accuracy. The global result for a certain parameterization experiment is the average of the results in all folds. In spite of having a higher computational cost, this cross-validation technique has been selected instead of other less computationally costly such as k-folds [59] due

Table 2: Results expressed as efficiency ± confidence interval and Cohen's Kappa Index using MFCCs and MS features.

Features	Original subset				Agreement subset			
	G		*R*		*G*		*R*	
	Efficiency (%)	κ	Efficiency (%)	κ	Efficiency (%)	κ	Efficiency (%)	κ
MFCC	54.5 ± 6.5	0.37	53.1 ± 6.5	0.51	75.9 ± 9.0	0.64	76.5 ± 9.0	0.64
MFCC + PCA	56.3 ± 6.5	0.39	52.2 ± 6.6	0.31	78.2 ± 8.7	0.67	74.1 ± 9.3	0.60
MFCC + LDA	45.5 ± 6.5	0.27	48.2 ± 6.6	0.29	65.5 ± 10.0	0.48	68.3 ± 9.9	0.50
MS	60.3 ± 6.4	0.45	54.9 ± 6.5	0.36	81.6 ± 8.1	0.72	76.5 ± 9.0	0.63
MS + PCA	58.5 ± 6.5	0.43	58.0 ± 6.5	0.41	79.3 ± 8.5	0.69	78.8 ± 8.7	0.68
MS + LDA	58.9 ± 6.4	0.44	59.8 ± 6.4	0.43	81.6 ± 8.1	0.72	83.5 ± 7.9	0.74

Table 3: Results expressed as efficiency ± confidence interval and Cohen's Kappa Index for MFCCs and MS features including Δ and ΔΔ.

Features	Dimensionality reduction	Agreement subset			
		G		*R*	
		Efficiency (%)	κ	Efficiency (%)	κ
MFCC + Δ	PCA	78.2 ± 8.7	0.67	74.1 ± 9.3	0.60
	LDA	72.4 ± 9.4	0.58	58.8 ± 10.5	0.35
MFCC + Δ + ΔΔ	PCA	**80.5 ± 8.3**	**0.71**	**77.7 ± 8.8**	**0.66**
	LDA	72.4 ± 9.4	0.58	62.4 ± 10.3	0.41
MS + Δ	PCA	**81.6 ± 8.1**	**0.73**	80.0 ± 8.5	0.69
	LDA	80.5 ± 8.3	0.72	81.2 ± 8.3	0.71
MS + Δ + ΔΔ	PCA	79.3 ± 8.5	0.63	80.0 ± 8.5	0.70
	LDA	80.5 ± 8.3	0.71	**84.7 ± 7.7**	**0.76**

to its suitability in view of the reduced number of recordings contained in the *agreement subsets*.

3.2.3. Classification. The features extracted during the parameterization stage are used to feed the classifier, which is based on the *Gaussian Mixture Model* (GMM) paradigm. Having a data vector $\mathbf{x}$ of dimension d resulting from the parameterization stage, a GMM is a model of the probability density function defined as a finite mixture of g multivariate Gaussian components of the form:

$$p(\mathbf{x} \mid \Theta_i) = \sum_{r=1}^{g} \boldsymbol{\lambda}_r \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \quad (9)$$

where $\boldsymbol{\lambda}_r$ are scalar mixture weights, $\mathcal{N}(\cdot)$ are Gaussian density functions with mean $\boldsymbol{\mu}_r$ of dimension d and covariances $\boldsymbol{\Sigma}_r$ of dimension $d \times d$, and $\Theta_i = \{\boldsymbol{\lambda}_r, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r\}|_{r=1}^{g}$ comprises the abovementioned set of parameters that defines the class to be modeled. Thus, for each class Θ_i to be modeled (i.e., values of the G and R perceptual levels: 0, 1, 2, or 3), a GMM is trained. Θ_i is estimated using the *expectation-maximization* algorithm (EM) [60]. The final decision about the class that a vector belongs to is taken establishing for each pair of classes i, j a threshold Γ over the likelihood ratio (LR), that in the logarithmic domain is given by

$$\mathrm{LR} = \log(p(\mathbf{x} \mid \Theta_i)) - \log(p(\mathbf{x} \mid \Theta_j)). \quad (10)$$

The threshold Γ is fixed at the Equal Error Rate (ERR) point.

In this stage, the number of Gaussian components of the GMM was varied from 4 to 48. The assessment of the classifier was performed by means of efficiency and Cohen's Kappa Index (κ) [61]. This last indicator provides information about the agreement between the results of the classifier and the clinician's perceptual labeling.

4. Results

The best results obtained for each type of test can be observed in Table 2, which disposes the outcomes taking into account the type of characterization, dimensionality reduction, and database subset used. All tests were performed using the aforementioned sets of the database with and without PCA and LDA techniques. Table 3 shows the outcomes adding first and second derivative to the original parameterizations before dimensionality reduction. All results are expressed in terms of efficiency and Cohen's Kappa Index. For the sake of simplicity, only results obtained with the third labeling of the *original subset* are shown, corresponding to columns $G3$ and $R3$ in Appendix, Table 6.

Concerning G trait, absolute best results (81.6%) are obtained in the *agreement database*, using MS + Δ in 140 ms frames, 22 centroids, Hilbert envelope, 240 Hz as max. modulation frequency, dimensionality reduction through PCA (93% reduction), and 4 GMM. Respecting MFCC, best results are obtained using MFCCs + Δ + ΔΔ, 22 coefficients, PCA, 20 ms frames, and 8 GMM.

Relating to R, as expected, absolute best results (84.7%) are also obtained in the agreement database using MS +

Table 4: Confusion matrices related to absolute best results in MS parameters and MFCCs. G_T and R_T are target labels while G_P and R_P are predicted labels.

	Grade					Roughness			
	G_P0	G_P1	G_P2	G_P3		R_P0	R_P1	R_P2	R_P3
MS Morphological Parameters									
G_T0	**28**	1	0	1	R_T0	**38**	1	0	0
G_T1	3	**3**	1	0	R_T1	3	**1**	2	0
G_T2	1	1	**11**	1	R_T2	1	0	**13**	1
G_T3	3	0	4	**29**	R_T3	3	1	1	**20**
MFCCs									
G_T0	**27**	0	2	1	R_T0	**35**	0	1	3
G_T1	2	**0**	5	0	R_T1	3	**0**	3	0
G_T2	0	0	**13**	1	R_T2	1	0	**9**	5
G_T3	0	0	6	**30**	R_T3	1	0	2	**22**

Table 5: Altman interpretation of Cohen's index.

κ	Agreement
≤0.20	Poor
0.21–0.40	Fair
0.41–0.60	Medium
0.61–0.80	Good
0.81–1.00	Excellent

$\Delta + \Delta\Delta$ calculated in 100 ms frames, 14 centroids, Hilbert envelope, 240 Hz as max. modulation frequency, dimensionality reduction through LDA, and 16 GMM. Respecting MFCC, best results are obtained using MFCCs + $\Delta + \Delta\Delta$, 22 coefficients, PCA, 20 ms frames, and 48 GMM.

Table 4 shows confusion matrices for MFCC and MS Morphological Parameters as the sum of the confusion matrices obtained at each of the test folds. They are calculated using the mentioned configurations that leaded to the best results.

5. Conclusion and Discussions

This study presents a new set of parameters based on MS being developed to characterize perturbations of the human voice. The performance of these parameters has been tested with an automatic system that emulates a perceptual assessment according to the G and R features of the GRBAS scale. The proposed automatic system follows a classical supervised learning setup, based on GMM. The outcomes have been compared to those obtained with a baseline setup using the classic MFCCs as input features. Dimensionality reduction methods as LDA and PCA have been applied to mitigate *the curse of dimensionality* effects induced by the size of the corpus used to train and validate the system. Best results are obtained with the proposed MS parameters, providing 81.6% and 84.7% of efficiency and 0.73 and 0.76 Cohen's Kappa Index for G and R, respectively, in the *agreement subset*. Having in mind Altman interpretation of Cohen's index [62], shown in Table 5, the agreement can be considered "good", almost "excellent." Likewise, most errors raised by the system correspond with adjacent classes, as it can be deduced from the confusion matrices represented in Table 4. It is noticeable that in many cases the second class (level 1 in traits G and R) is not detected properly and the main reason may be the lack of subjects of class 2 (level 1 in G and R) in the used corpus. The fact that GMM classifiers were trained with a poor quantity of class 2 frames with respect to the other classes explains the higher percentage of errors obtained for this class. In order to solve this problem in future works it might be necessary to use classification techniques for imbalanced data [63]. Another possible reason for the mismatching of intermediate classes (G and R equal to 1 or 2) is that these are the less reliable levels in GRBAS perceptual assessment as it was described by de Bodt et al. [5].

In reference to the outcomes obtained with features without dimensionality reduction, results are better for the *agreement subsets* using MS Morphological Parameters. Moreover, when applying LDA to the MS feature space, an absolute improvement of a 9% is obtained for R in comparison to MFCCs, leading to the best absolute outcome obtained and denoting that the MS Morphological Parameters are in some sense linearly separable. As a starting point, most of the *agreement subset* tests were performed with what we have called *the original subset* (224 files) using the three available label groups separately: one of them generated by one of the speech therapists and the other two created by the other specialist in two different sessions. In these cases, in spite of having a higher number of files and a more class-balanced database, results barely exceed 60% of efficiency. This demonstrates that the consistency of the database labeling (i.e., removing the noise introduced during the labeling process due to intra- and interrater's variability) is crucial to obtain more accurate results. An interesting conclusion is that further studies should utilize only consistent labels obtained in agreement with several therapists and in different evaluation sessions.

In order to search for some evidences proving that the selected cross-validation technique is not influencing the results by producing corpus-adjusted trained models, most of the tests are launched again using a 6-fold cross-validation technique as a prospecting experiment. Almost the same maximum efficiencies were obtained in all cases with a difference of around ±1%, suggesting that the selected cross-validation technique is not producing corpus-adjusted trained models.

Regarding the use of derivatives Δ and $\Delta\Delta$, they improve performance mainly when using MFCCs in 20 ms frames for G trait. This suggests that derivatives provide relevant extra information related to the short term variations occurred in pathological voices [64]. In the rest of the cases the improvements are limited; therefore, the influence of derivatives in G and R detection systems should be studied in detail in the future work.

Comparing this work with other studies mentioned in Section 1, results with MFCCs are coherent with these obtained in [17, 21, 36], although methodologies followed in them are different to the one proposed in this study. As it is stated in Section 1, previous studies seldom exceed 75% efficiency. Taking into account G and R traits, only [19, 20] surpass that value achieving 80% for G trait.

Table 6: Subsets labeling.

File	*G1*	*G2*	*G3*	*R1*	*R2*	*R3*	File	*G1*	*G2*	*G3*	*R1*	*R2*	*R3*	File	*G1*	*G2*	*G3*	*R1*	*R2*	*R3*	File	*G1*	*G2*	*G3*	*R1*	*R2*	*R3*
alb18an	2	3	3	2	3	3	gpc1nal	**0**	**0**	**0**	**0**	**0**	**0**	lba15an	2	0	0	1	0	0	pmc26an	2	3	3	**2**	**2**	**2**
amc14an	1	3	3	1	3	3	gsb11an	**3**	**3**	**3**	**3**	**3**	**3**	lba24an	2	0	0	1	0	0	pmd25an	2	3	2	2	3	2
aos21an	**3**	**3**	**3**	**3**	**3**	**3**	gxl21an	1	0	0	1	0	0	ldp1nal	1	0	0	**0**	**0**	**0**	pmf03an	2	1	1	2	1	1
axd19an	**0**	**0**	**0**	**0**	**0**	**0**	gxt10an	**3**	**3**	**3**	**3**	**3**	**3**	les15an	**3**	**3**	**3**	3	0	0	rcc11an	2	3	3	**2**	**2**	**2**
axh1nal	**0**	**0**	**0**	**0**	**0**	**0**	gzz1nal	1	0	0	1	0	0	lgm01an	1	0	0	1	0	0	rhg1nal	0	1	0	0	1	0
axt13an	**1**	**1**	**1**	1	0	1	hbl1nal	1	0	0	1	0	0	ljh06an	**2**	**2**	**2**	**2**	**2**	**2**	rhm1nal	**0**	**0**	**0**	**0**	**0**	**0**
bah13an	1	3	2	1	3	2	hjh07an	2	3	3	2	3	3	ljs31an	2	1	1	1	1	0	rhp12an	2	3	2	2	3	2
bef05an	2	3	3	2	0	0	hlm24an	1	2	2	1	2	2	lla1nal	1	0	0	**0**	**0**	**0**	rjf22an	2	3	3	2	3	3
bjb1nal	**0**	**0**	**0**	**0**	**0**	**0**	hxi29an	1	3	2	1	3	2	llm22an	3	3	2	3	3	2	rjl28an	3	3	2	2	3	2
bjv1nal	**0**	**0**	**0**	**0**	**0**	**0**	hxl58an	1	0	0	1	0	0	lmv1nal	1	0	0	1	0	0	rjr15an	**1**	**1**	**1**	**1**	**1**	**1**
bkb13an	1	0	1	2	0	1	jaf1nal	**0**	**0**	**0**	**0**	**0**	**0**	lmw1nal	1	0	0	**0**	**0**	**0**	rjs1nal	**0**	**0**	**0**	**0**	**0**	**0**
blb03an	2	3	2	2	3	2	jan1nal	**1**	**1**	**1**	**1**	**1**	**1**	lnc11an	1	0	0	**0**	**0**	**0**	rjz16an	1	0	0	1	0	0
bpf03an	1	2	1	1	2	1	jap02an	2	1	1	2	0	1	lrd21an	1	0	0	**0**	**0**	**0**	rmb07an	2	3	3	2	3	3
bsg13an	**1**	**1**	**1**	**1**	**1**	**1**	jap1nal	**0**	**0**	**0**	**0**	**0**	**0**	lvd28an	2	3	2	**2**	**2**	**2**	rpj15an	**3**	**3**	**3**	**3**	**3**	**3**
cac10an	2	3	3	2	0	0	jcc10an	**2**	**2**	**2**	**2**	**2**	**2**	lwr18an	1	3	3	1	0	0	rpq20an	2	3	3	2	3	3
cad1nal	**0**	**0**	**0**	**0**	**0**	**0**	jcr01an	**3**	**3**	**3**	**3**	**3**	**3**	lxc01an	**2**	**2**	**2**	2	0	0	rtl17an	1	0	1	1	0	1
cak25an	2	1	1	2	1	1	jeg1nal	**0**	**0**	**0**	**0**	**0**	**0**	lxc06an	**3**	**3**	**3**	**3**	**3**	**3**	rwc23an	2	3	3	2	3	3
ceb1nal	**0**	**0**	**0**	**0**	**0**	**0**	jeg29an	2	3	2	2	0	0	lxr15an	3	3	2	3	3	2	rxm15an	1	0	0	1	0	0
cls31an	2	1	1	2	1	1	jfg08an	**3**	**3**	**3**	**3**	**3**	**3**	mab06an	2	1	1	2	1	1	rxp02an	1	3	3	1	3	3
cma06an	1	2	1	1	2	1	jfn21an	3	2	2	3	2	2	mam08an	2	3	3	2	3	3	sac10an	2	3	3	2	3	3
cmr06an	**0**	**0**	**0**	**0**	**0**	**0**	jhw29an	2	1	1	2	1	1	mam1nal	**0**	**0**	**0**	**0**	**0**	**0**	sae01an	1	0	0	1	0	0
crm12an	**3**	**3**	**3**	3	2	2	jkr1nal	1	0	0	1	0	0	mas1nal	**0**	**0**	**0**	**0**	**0**	**0**	sav18an	**3**	**3**	**3**	**3**	**3**	**3**
ctb30an	2	2	1	2	1	1	jld24an	2	3	2	2	3	2	mcb1nal	1	0	0	1	0	0	sbf11an	**0**	**0**	**0**	**0**	**0**	**0**
daj1nal	**0**	**0**	**0**	**0**	**0**	**0**	jls11an	2	1	1	2	1	1	mcw21an	2	1	1	2	1	1	scc15an	**3**	**3**	**3**	3	0	1
dap17an	2	2	1	2	2	1	jmc18an	1	0	0	1	0	0	mec06an	2	0	0	2	0	0	sck1nal	1	0	0	1	0	0
das30an	1	—	1	1	—	1	jmc1nal	1	0	0	1	0	0	mec28an	2	2	1	1	2	1	sct1nal	1	0	0	1	0	0
dbf18an	2	2	1	2	2	1	jpp27an	**3**	**3**	**3**	3	0	0	mfc20an	**3**	**3**	**3**	2	3	3	seb1nal	1	0	0	1	0	0
dfp1nal	**0**	**0**	**0**	**0**	**0**	**0**	jrf30an	1	0	0	1	0	0	mfm1nal	1	0	0	1	0	0	sec02an	2	0	1	2	0	1
djg1nal	**0**	**0**	**0**	**0**	**0**	**0**	jth1nal	1	0	0	**0**	**0**	**0**	mju1nal	**0**	**0**	**0**	**0**	**0**	**0**	sef10an	1	0	0	1	0	0
djp04an	**2**	**2**	**2**	**2**	**2**	**2**	jtm05an	1	2	1	1	2	1	mpb23an	**3**	**3**	**3**	**3**	**3**	**3**	seg18an	2	1	1	2	1	1
dma1nal	**0**	**0**	**0**	**0**	**0**	**0**	jxc1nal	1	0	0	**0**	**0**	**0**	mpf25an	1	2	1	1	2	1	sek06an	2	2	1	2	2	1
dmc03an	2	3	2	2	0	0	jxc21an	2	0	0	2	0	0	mps09an	2	0	1	2	0	1	shd04an	**3**	**3**	**3**	**3**	**3**	**3**
dmp04an	**2**	**2**	**2**	**2**	**2**	**2**	jxd30an	**1**	**1**	**1**	0	1	1	mrb11an	1	3	2	1	3	2	sis1nal	**0**	**0**	**0**	**0**	**0**	**0**
drc15an	**2**	**2**	**2**	2	0	0	jxf11an	**3**	**3**	**3**	**3**	**3**	**3**	mrc20an	2	3	2	2	3	2	sjd28an	**1**	**1**	**1**	**1**	**1**	**1**
dsc25an	**3**	**3**	**3**	3	0	0	kab03an	1	0	0	1	0	0	mwd28an	2	3	2	2	3	2	slc1nal	2	0	0	1	0	0
dsw14an	2	—	1	2	—	1	kac07an	1	0	0	1	0	0	mxb1nal	**0**	**0**	**0**	**0**	**0**	**0**	slc23an	2	0	1	1	0	1
dvd19an	**3**	**3**	**3**	**3**	**3**	**3**	kan1nal	**0**	**0**	**0**	**0**	**0**	**0**	mxc10an	**2**	**2**	**2**	**2**	**2**	**2**	slg05an	1	0	0	1	0	0
dwk04an	**2**	**2**	**2**	**2**	**2**	**2**	kcg23an	2	2	1	2	2	1	mxn24an	1	0	0	1	0	0	sma08an	**3**	**3**	**3**	**3**	**3**	**3**
dws1nal	**0**	**0**	**0**	**0**	**0**	**0**	kcg25an	1	0	0	1	0	0	mxz1nal	1	0	0	**0**	**0**	**0**	sws04an	3	2	2	3	2	2
eab27an	**3**	**3**	**3**	**3**	**3**	**3**	kdb23an	**2**	**2**	**2**	**2**	**2**	**2**	nfg08an	2	2	1	1	2	1	sxv1nal	1	0	0	1	0	0
eas11an	1	2	1	1	2	1	kjb19an	2	3	3	2	3	3	njs06an	2	2	1	2	2	1	tab21an	**3**	**3**	**3**	**3**	**3**	**3**
eas15an	**3**	**3**	**3**	**3**	**3**	**3**	klc06an	2	3	3	2	0	0	njs1nal	1	0	0	1	0	0	tdh12an	2	2	1	2	2	1
edc1nal	1	0	0	1	0	0	klc09an	2	3	2	2	3	2	nkr03an	1	0	0	1	0	0	tlp13an	**2**	**2**	**2**	**2**	**2**	**2**
eec04an	**3**	**3**	**3**	3	2	2	kld26an	2	2	1	2	2	1	nlc08an	2	2	1	2	2	1	tls09an	**2**	**2**	**2**	**2**	**2**	**2**
eed07an	**3**	**3**	**3**	**3**	**3**	**3**	kmc22an	2	1	1	2	1	0	nmb28an	2	3	2	1	3	2	tpp24an	1	0	0	1	0	0
ejc1nal	**0**	**0**	**0**	**0**	**0**	**0**	kms29an	2	3	3	2	3	3	nmc22an	2	2	1	**1**	**1**	**1**	tps16an	1	—	1	1	—	1
ejh24an	**3**	**3**	**3**	3	0	1	kmw05an	**3**	**3**	**3**	3	1	1	nml15an	0	1	1	0	1	1	txn1nal	1	—	0	1	—	0
emp27an	**2**	**2**	**2**	**2**	**2**	**2**	kps25an	0	2	1	0	2	1	nmv07an	**3**	**3**	**3**	3	0	0	vaw07an	**3**	**3**	**3**	**3**	**3**	**3**
ess05an	1	0	0	1	0	0	ktj26an	**3**	**3**	**3**	**3**	**3**	**3**	oab28an	1	3	3	2	3	3	vmc1nal	1	0	0	1	0	0
eww05an	**2**	**2**	**2**	**2**	**2**	**2**	kxb17an	**3**	**3**	**3**	**3**	**3**	**3**	ovk1nal	**0**	**0**	**0**	**0**	**0**	**0**	wcb24an	1	0	0	1	0	0
fmb1nal	**0**	**0**	**0**	**0**	**0**	**0**	kxh30an	**3**	**3**	**3**	**3**	**3**	**3**	pat10an	2	3	3	1	3	3	wdk1nal	**0**	**0**	**0**	**0**	**0**	**0**
fmr17an	**3**	**3**	**3**	**3**	**3**	**3**	lac02an	2	1	1	2	0	1	pbd1nal	1	0	0	1	0	0	wfc07an	2	3	2	2	3	2
fxc12an	1	2	2	1	2	2	lad13an	1	2	1	1	2	1	pca1nal	1	0	0	**0**	**0**	**0**	wjb06an	2	0	0	2	0	0
gdr15an	**3**	**3**	**3**	3	0	0	lad1nal	**0**	**0**	**0**	**0**	**0**	**0**	pdo11an	2	3	3	2	3	3	wjp20an	**3**	**3**	**3**	**3**	**3**	**3**
gmm09an	**3**	**3**	**3**	**3**	**3**	**3**	lai04an	1	3	2	1	3	2	pgb16an	**1**	**1**	**1**	**1**	**1**	**1**	wpb30an	2	1	1	2	1	1
gms05an	2	3	3	2	3	3	lap05an	1	3	3	1	3	3	plw14an	**2**	**2**	**2**	**2**	**2**	**2**	wxe04an	**3**	**3**	**3**	**3**	**3**	**3**

Despite the promising results, an accurate comparison with the studies found in the state of the art is difficult since, as stated in [65], different works tend to use different types of corpus and methodologies, and results are unfortunately dependant of the corpus used for training and validation. Furthermore, those cases on which different studies utilize the same database, labeling is usually different. In this sense, the definition of a standardized database with a consistent and known labeling would lead to comparable results. For this reason, with the aim of providing the scientific community with a benchmark labeling and to promote a more solid comparative estimate of future techniques and studies, the labeling of the G and R features used on this work has been included in Table 6. Despite its known limitations [65], the fact that MEEI database is commercially available for researchers is also an advantage in this sense.

On the other hand, other approaches such as [11, 23] have already demonstrated that MS is a good source of information to detect pathological voices or to perform an automatic pathology classification. The main difference with respect to Markaki's approach is that in this study MS is used to evaluate the speech according to a 4-level scale in two different features of the speech: Grade and Roughness. On the other hand, the parameters used in the present study are less abstract and have an easier physical interpretation, opening the possibility of using them in a clinical setting.

In spite of the good figures, MS has a weakness which could make it a nonviable parameterization in some applications: computational cost. Depending on the configuration and frequency margins, to calculate a MS matrix can take around 400 times more than to calculate MFCCs on the same signal frame.

Regarding the future work, all MS parameters must be studied and adjusted separately to find the adequate frequency margins of operation to optimize results. In addition, the use of the proposed MS Morphological Parameters in combination with some other features such as complexity and noise measurements or cepstral-based coefficients to characterize GRBAS traits would be advisable. Moreover, the study of regression methods like Support Vector Regression [66] and other feature selection techniques such as Least Absolute Shrinkage and Selection Operator (LASSO) [67] is of interest. In respect of the classification stage, the stratification of the speakers according to her/his sex, age, or emotional state could increase performance as suggested in [68]. For this purpose, a priori categorization of speaker's characteristics using hierarchical methods might be used to simplify the statistical models behind to automatically assess the quality of speech.

Summarizing, results suggest that the proposed MS Morphological Parameters are an objective basis to help clinicians to assess Grade and Roughness according the GRBAS scale, reducing uncertainty and making the assessment easier to replicate. It would be advisable to study the synthesis of a new parameter combining the proposed MS Morphological Parameters, being suitable for therapists and physicians. In view of the experiments carried out in this work, there are evidences that suggest that the use of these parameters provides better results than the classic MFCCs, traditionally used to characterize voice signals. On the other hand, its main drawback is the initial difficulty of applying the proposed MS-based parameters to the study of running speech.

Appendix

On Table 6 perceptual assessment of GRBAS Grade and Roughness traits for 224 recordings of MEEI corpus [55] can be found. $G1$ and $R1$ are assessments of Therapist 1. The rest of evaluations are performed by Therapist 2 in two different sessions. Numbers in bold represent the *agreement subsets* while all evaluations are the *original subsets*.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors of this paper have developed their work under the grant of the project *TEC2012-38630-C04-01* from the Ministry of Economy and Competitiveness of Spain and *Ayudas para la realizacion del doctorado (RR01/2011)* from Universidad Politecnica de Madrid, Spain.

References

[1] C. Sapienza and B. Hoffman-Ruddy, *Voice Disorders*, Plural Publishing, 2009.

[2] D. K. Wilson, *Voice Problems of Children*, Williams & Wilkins, Baltimore, Md, USA, 1987.

[3] G. B. Kempster, B. R. Gerratt, K. V. Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, "Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol," *American Journal of Speech-Language Pathology*, vol. 18, no. 2, pp. 124–132, 2009.

[4] M. Hirano, *Clinical Examination of Voice*, Springer, 1981.

[5] M. S. De Bodt, F. L. Wuyts, P. H. van de Heyning, and C. Croux, "Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality," *Journal of Voice*, vol. 11, no. 1, pp. 74–80, 1997.

[6] I. V. Bele, "Reliability in perceptual analysis of voice quality," *Journal of Voice*, vol. 19, no. 4, pp. 555–573, 2005.

[7] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[8] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 181–190, 2014.

[9] C. Fredouille, G. Pouchoulin, A. Ghio, J. Revis, J.-F. Bonastre, and A. Giovanni, "Back-and-forth methodology for objective voice quality assessment: from/to expert knowledge to/from automatic classification of dysphonia," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 982102, 13 pages, 2009.

[10] P. Gómez-Vilda, R. Fernández-Baillo, V. Rodellar-Biarge et al., "Glottal Source biometrical signature for voice pathology

detection," *Speech Communication*, vol. 51, no. 9, pp. 759–781, 2009.

[11] M. Markaki and Y. Stylianou, "Voice pathology detection and discrimination based on modulation spectral features," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 1938–1948, 2011.

[12] J. D. Arias-Londoño, J. I. Godino-Llorente, M. Markaki, and Y. Stylianou, "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices," *Logopedics Phoniatrics Vocology*, vol. 36, no. 2, pp. 60–69, 2011.

[13] R. Wielgat, T. P. Zieliński, T. Woźniak, S. Grabias, and D. Król, "Automatic recognition of pathological phoneme production," *Folia Phoniatrica et Logopaedica*, vol. 60, no. 6, pp. 323–331, 2009.

[14] L. Salhi and A. Cherif, "Robustness of auditory teager energy cepstrum coefficients for classification of pathological and normal voices in noisy environments," *The Scientific World Journal*, vol. 2013, Article ID 435729, 8 pages, 2013.

[15] M. Rosa, J. Pereira, M. Greller, and A. Carvalho, "Signal processing and statistical procedures to identify laryngeal pathologies," in *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems (ICECS '99)*, vol. 1, pp. 423–426, Pafos, Cyprus, 1999.

[16] G. Muhammad, M. Alsulaiman, A. Mahmood, and Z. Ali, "Automatic voice disorder classification using vowel formants," in *Proceedings of the 12th IEEE International Conference on Multimedia and Expo (ICME '11)*, pp. 1–6, IEEE, July 2011.

[17] P. Yu, Z. Wang, S. Liu, N. Yan, L. Wang, and M. Ng, "Multidimensional acoustic analysis for voice quality assessment based on the GRBAS scale," in *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing (ISCSLP '14)*, pp. 321–325, IEEE, Singapore, September 2014.

[18] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *Journal of the Royal Society Interface*, vol. 8, no. 59, pp. 842–855, 2011.

[19] G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, and A. Giovanni, "Frequency study for the characterization of the dysphonic voices," in *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech '07)*, pp. 1198–1201, Antwerp, Belgium, August 2007.

[20] G. Pouchoulin, C. Fredouille, J. Bonastre et al., "Dysphonic voices and the 0–3000 Hz frequency band," in *Proceedings of the 9th Annual Conference of the International Speech Communication Association (Interspeech '08)*, pp. 2214–2217, ISCA, Brisbane, Australia, September 2008.

[21] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco, and F. Cruz-Roldán, "Automatic assessment of voice quality according to the GRBAS scale," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 2478–2481, September 2006.

[22] A. Stráník, R. Čmejla, and J. Vokřál, "Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS scale," *Journal of Voice*, vol. 28, no. 5, pp. 653.e9–653.e17, 2014.

[23] M. Markaki and Y. Stylianou, "Modulation spectral features for objective voice quality assessment: the breathiness case," in *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, December 2009.

[24] W. S. Winholtz and L. O. Ramig, "Vocal tremor analysis with the vocal demodulator," *Journal of Speech and Hearing Research*, vol. 35, no. 3, pp. 562–573, 1992.

[25] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering Online*, vol. 6, article 23, 2007.

[26] C. Peng, W. Chen, X. Zhu, B. Wan, and D. Wei, "Pathological voice classification based on a single Vowel's acoustic features," in *Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT '07)*, pp. 1106–1110, October 2007.

[27] J. I. Godino-Llorente and P. Gómez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, 2004.

[28] C. Maguire, P. de Chazal, R. B. Reilly, and P. D. Lacy, "Identification of voice pathology using automated speech analysis," in *Proceedings of the 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA '03)*, pp. 259–262, Florence, Italy, December 2003.

[29] M. Marinaki and C. Kotropoulos, "Automatic detection of vocal fold paralysis and edema," in *Proceedings of the ICSLP*, Jeju Island, Republic of Korea, 2004.

[30] J. Godino-Llorente, P. Gómez-Vilda, N. Sáenz-Lechón, M. Blanco-Velasco, F. Cruz-Roldán, and M. A. Ferrer, "Discriminative methods for the detection of voice disorders," in *Proceedings of the International Conference on Non-Linear Speech Processing (NOLISP '05)*, pp. 158–167, Barcelona, Spain, April 2005.

[31] K. Shama, A. Krishna, and N. U. Cholayya, "Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, Article ID 85286, 2007.

[32] A. I. Fontes, P. T. Souza, A. D. Neto, A. d. Martins, and L. F. Silveira, "Classification system of pathological voices using correntropy," *Mathematical Problems in Engineering*, vol. 2014, Article ID 924786, 7 pages, 2014.

[33] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI '95)*, vol. 2, pp. 1137–1145, 1995.

[34] S. Jannetts and A. Lowit, "Cepstral analysis of hypokinetic and ataxic voices: correlations with perceptual and other acoustic measures," *Journal of Voice*, vol. 28, no. 6, pp. 673–680, 2014.

[35] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9-10, pp. 341–345, 2002.

[36] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón et al., "Automatic GRBAS assessment using complexity measures and a multiclass GMM-based detector," in *Proceedings of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 2011.

[37] N. C. Singha and F. E. Theunissen, "Modulation spectra of natural sounds and ethological theories of auditory processing," *The Journal of the Acoustical Society of America*, vol. 114, no. 6, pp. 3394–3411, 2003.

[38] L. Atlas and S. A. Shamma, "Joint acoustic and modulation frequency," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 7, pp. 668–675, 2003.

[39] S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, “Music genre/mood classification using a feature-based modulation spectrum,” in *Proceedings of the International Conference on Mobile IT-Convergence (ICMIC ’11)*, pp. 133–136, IEEE, September 2011.

[40] W.-Y. Chu, J.-W. Hung, and B. Chen, “Modulation spectrum factorization for robust speech recognition,” in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC ’11)*, pp. 1–6, October 2011.

[41] H.-T. Fan, Y.-C. Tsai, and J.-W. Hung, “Enhancing the sub-band modulation spectra of speech features via nonnegative matrix factorization for robust speech recognition,” in *Proceedings of the International Conference on System Science and Engineering (ICSSE ’12)*, pp. 179–182, July 2012.

[42] E. Bozkurt, O. Toledo-Ronen, A. Sorin, and R. Hoory, “Exploring modulation spectrum features for speech-based depression level classification,” in *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, Singapore, September 2014.

[43] K. M. Carbonell, R. A. Lester, B. H. Story, and A. J. Lotto, “Discriminating simulated vocal tremor source using amplitude modulation spectra,” *Journal of Voice*, vol. 29, no. 2, pp. 140–147, 2015.

[44] P. H. Dejonckere, C. Obbens, G. M. de Moor, and G. H. Wieneke, “Perceptual evaluation of dysphonia: reliability and relevance,” *Folia Phoniatrica*, vol. 45, no. 2, pp. 76–83, 1993.

[45] M. P. Karnell, S. D. Melton, J. M. Childes, T. C. Coleman, S. A. Dailey, and H. T. Hoffman, “Reliability of clinician-based (grbas and cape-v) and patientbased (v-rqol and ipvi) documentation of voice disorders,” *Journal of Voice*, vol. 21, no. 5, pp. 576–590, 2007.

[46] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, New York, NY, USA, 1993.

[47] S. M. Schimmel, L. E. Atlas, and K. Nie, “Feasibility of single channel speaker separation based on modulation frequency analysis,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’07)*, vol. 4, pp. IV605–IV608, April 2007.

[48] L. Atlas, P. Clark, and S. Schimmel, “Modulation Toolbox Version 2.1 for MATLAB,” 2010, http://isdl.ee.washington.edu/projects/modulationtoolbox/.

[49] S. M. Schimmel and L. E. Atlas, “Coherent envelope detection for modulation filtering of speech,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’05)*, vol. 1, pp. I221–I224, IEEE, March 2005.

[50] R. Cusack and N. Papadakis, “New robust 3-D phase unwrapping algorithms: application to magnetic field mapping and undistorting echoplanar images,” *NeuroImage*, vol. 16, no. 3, pp. 754–764, 2002.

[51] B. Gajić and K. K. Paliwal, “Robust speech recognition in noisy environments based on subband spectral centroid histograms,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 2, pp. 600–608, 2006.

[52] R. J. Elble, “Central mechanisms of tremor,” *Journal of Clinical Neurophysiology*, vol. 13, no. 2, pp. 133–144, 1996.

[53] M. Bové, N. Daamen, C. Rosen, C.-C. Wang, L. Sulica, and J. Gartner-Schmidt, “Development and validation of the vocal tremor scoring system,” *The Laryngoscope*, vol. 116, no. 9, pp. 1662–1667, 2006.

[54] R. Peters and R. Strickland, “Image complexity metrics for automatic target recognizers,” in *Proceedings of the Automatic Target Recognizer System and Technology Conference*, October 1990.

[55] *Voice Disorders Database*, Kay Elemetrics Corporation, Lincoln Park, NJ, USA, 1994.

[56] V. Parsa and D. G. Jamieson, “Identification of pathological voices using glottal noise measures,” *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 2, pp. 469–485, 2000.

[57] L. Smith, *A Tutorial on Principal Components Analysis*, vol. 51, Cornell University, Ithaca, NY, USA, 2002.

[58] R. Haeb-Umbach and H. Ney, “Linear discriminant analysis for improved large vocabulary continuous speech recognition,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 1, pp. 13–16, San Francisco, Calif, USA, March 1992.

[59] B. Efron and G. Gong, “A leisurely look at the bootstrap, the jackknife, and cross-validation,” *The American Statistician*, vol. 37, no. 1, pp. 36–48, 1983.

[60] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[61] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[62] D. G. Altman, *Practical Statistics for Medical Research*, CRC Press, 1990.

[63] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[64] D. G. Childers, “Detection of laryngeal function using speech and electroglottographic data,” *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 1, pp. 19–25, 1992.

[65] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, “Methodological issues in the development of automatic systems for voice pathology detection,” *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.

[66] V. N. Vapnik and V. Vapnik, *Statistical Learning Theory*, vol. 2 of *Adaptive and Learning Systems for Signal Processing, Communications, and Control*, John Wiley & Sons, 1998.

[67] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B: Methodological*, vol. 58, no. 1, pp. 267–288, 1996.

[68] R. Fraile, N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and C. Fredouille, “Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex,” *Folia Phoniatrica et Logopaedica*, vol. 61, no. 3, pp. 146–152, 2009.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 683279, 11 pages
http://dx.doi.org/10.1155/2015/683279

Research Article

Nonword Repetition and Speech Motor Control in Children

Christina Reuterskiöld and Maria I. Grigos

Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Suite 922, New York, NY 10012, USA

Correspondence should be addressed to Christina Reuterskiöld; christina.reuterskiold@nyu.edu

Received 23 January 2015; Accepted 23 April 2015

Academic Editor: Markus Hess

This study examined how familiarity of word structures influenced articulatory control in children and adolescents during repetition of real words (RWs) and nonwords (NWs). A passive reflective marker system was used to track articulator movement. Measures of accuracy were obtained during repetition of RWs and NWs, and kinematic analysis of movement duration and variability was conducted. Participants showed greater consonant and vowel accuracy during RW than NW repetition. Jaw movement duration was longer in NWs compared to RWs across age groups, and younger children produced utterances with longer jaw movement duration compared to older children. Jaw movement variability was consistently greater during repetition of NWs than RWs in both groups of participants. The results indicate that increases in phonological short-term memory demands affect articulator movement. This effect is most pronounced in younger children. A range of skills may develop during childhood, which supports NW repetition skills.

1. Introduction

Language acquisition is often studied in isolation from neuromotor development [1, 2]. As a result, the relationship between higher-level language and cognitive skills and lower-level abilities related to speech output is poorly understood. There is a need to better understand this relationship not only from a theoretical standpoint but also from a clinical perspective as many diagnostic measures of language ability, such as nonword repetition, rely on speech production as a response mode. The purpose of the present study was to explore the interaction between cognitive/linguistic and speech motor processes by studying how children and adolescents modify articulatory control during the repetition of real words (RWs) and nonwords (NWs) that vary in length.

NW repetition is a task that has been widely used in assessment of children with language and literacy impairments and it has been suggested as a marker of the behavioral phenotype of Specific Language Impairment (SLI) [3, 4]. Children listen to pseudowords and are asked to repeat them as accurately as possible [5–9]. The notion underlying the use of NW repetition is that using unknown words (e.g., *mustrefalj*) makes it difficult to access lexical knowledge in long-term memory to support performance. There is a large body of research demonstrating a link between NW repetition skills and language and literacy abilities in children and adolescents with and without language impairments (e.g., [10–16]).

NW repetition has been considered a measure of phonological short-term memory (PSTM) [8], but it is currently viewed as a task with high psycholinguistic complexity [11, 14, 17–19] taxing a range of input and output processes. The working memory model [20, 21] has been a widely used theoretical concept describing a limited capacity system supporting cognitive processes in children and adults. The model has been extensively used over the last few decades as a framework to explain behavior in terms of both development and disorders. The original theoretical model is comprised of three parts, the central executive and two slave systems. One of the slave systems is the phonological loop, which is responsible for short-term storage of recently presented unknown phonological information. The phonological loop is comprised of a storage unit, which retains phonological representations of language, and a subvocal rehearsal unit, which is a tool that aids in retention of novel phonological information [22]. According to Gathercole [11], in a short-term memory model there is a temporal decay of phonological representations. Longer stimuli will therefore be more vulnerable to decay than shorter stimuli due to the increased time of presentation and repetition [23]. The difference in performance between children with language impairment and children with typical development (TD) increases with length of the NWs [8, 24]. As Leonard [25] discusses, citing a meta-analysis by Estes and colleagues [26], there is much evidence that children with language impairment perform worse than children with TD also on shorter tokens even at a length of one syllable. This, according to Leonard [25], is evidence that processes other than PSTM contribute to the differences in performance between groups.

While it is obvious that speech motor processes are involved in NW repetition, it is not clear how different levels of articulation proficiency affect the formation of phonological representations. A few older studies have examined PSTM in individuals with severe motor speech disorders [27–29]. Such studies are interesting since they may help clarify whether articulation skills support phonological processing skills, including phonological memory. Bishop and colleagues studied individuals with anarthria (the inability to produce speech) who were diagnosed with cerebral palsy. The investigators showed that their participants were able to retain RWs long enough in memory to perform a judgment of same-different, but this was not the case when the stimuli were NWs. It was concluded that the retention of unfamiliar phonological word forms is supported by overt or covert articulation, a strategy that may not be available to individuals with speech impairment, which results in difficulty remembering NWs.

Although there appears to be a link between speech motor control and PSTM, few studies of NW repetition have also examined oral motor skills. In one of our earlier studies of five-year-old Swedish-speaking children with language impairment [14, 30], we found a correlation between NW repetition and expressive phonology, but not with performance on a test of oral motor skills. This result suggests that NW repetition taps into the representational level of phonology but is not linked to oral motor skills per se. More recently, Krishnan et al. [31] reported that differences in oral motor control in children contributed significantly to NW repetition scores independent of age and general language or cognitive skills. The authors suggested that poor oral motor control might be one of several risk factors for language impairment in children. The contrasting findings from these studies highlight the complex relationship between language and motor processing, which merits further investigation.

Developmental changes in articulatory control with respect to linguistic complexity are well documented. Typically, developing speakers modify spatial and temporal features of articulator movement throughout childhood and adolescence [32–40]. Maturational changes in lip and jaw movement include decreases in duration and increases in velocity [32, 33, 37, 39, 41]. Movement variability also decreases with age [32, 37, 42, 43]. Several studies have explored the influence of linguistic complexity on articulatory control. Maner et al. [44] examined lower lip movement changes in five- and eight-year-old children and adults during RW phrases that increased in length and syntactic complexity. The children consistently produced longer and more variable lower lip movements as utterances increased in length and complexity. Dromey and Benson [45] reported decreases in lip movement stability in adults when linguistic demands were placed on participants (i.e., verb generation during a sentence completion task). Further, they showed that speakers produced slower lip movements during taxing cognitive tasks (i.e., counting backward from 100 by 7). Walsh et al. [46] examined articulator movement in nine- to ten-year-olds and adults during the production of NWs that increased in syllable length. Both the adults and children showed a tendency for lip aperture variability, as well as lower lip/jaw variability, to increase as syllable length increased. Taken together, results from these studies demonstrate that speech motor control is influenced by cognitive and linguistic processing demands.

Several recent studies involving adult speakers have focused on the interaction between higher levels of processing and speech output [47–50]. McMillan et al. [47] investigated elicited slips of the tongue where participants were asked to repeat word pairs, which appeared on a screen for a brief period of time. In some of the word-pair cases subjects were cued to repeat the words in reversed order (*tum gop* resulting in *gop tum*). The outcome of the study showed that substitutions were more likely to occur when RW competitors were present. Further, electropalatographic measures revealed greater variability when stimuli pairs consisted of NWs only. From these findings, the researchers concluded that there is a clear lexical bias in articulation and a relationship between cognitive and motor movements involved in speech processing and production [47]. To tap into this relationship in children, Heisler et al. [51] examined the influence of word learning on speech production during a novel word-learning task. They compared phonetic accuracy and movement pattern stability during the production of phonetic forms with and without lexical representation (a visual referent and/or object function). The results showed that production of a novel phonetic sequence was less variable in terms of articulatory movement when paired with a visual and/or functional lexical referent.

There is a need for research that can lead to a deeper understanding of the perceptual, cognitive, and motor processes that are involved in NW repetition, as well as their relationship with linguistic processes [52–56]. The focus of the current work is to examine the influence of increased PSTM demands (repetition of RWs versus NWs) on speech motor control in typically developing children and adolescents. We explored the hypothesis that articulator movement duration and variability will increase during tasks with greater PSTM demands. Moreover, we investigated whether there are age-related differences in articulator movements related to word type, hypothesizing that differences between RWs and NWs would be larger in children than in adolescents. Specific questions guiding this study were as follows. (1) Does jaw movement duration and stability differ between children and adolescents during the repetition of RWs versus NWs with similar phonetic properties? (2) Is articulatory control for the production of RWs versus NWs influenced by increases in stimuli length in children and adolescents?

2. Method

2.1. Participants. Sixteen participants completed the study and were categorized into either a younger or older age group (eight participants per group (four males/four females)). Mean age (standard deviation) was 6.10 (1.6) in the younger group and 14.4 (1.8) in the older group. These age groupings were selected to compare the effects of increased cognitive demands on articulator movements between children and adolescents. It is well documented that articulatory control differs between children and adolescents [40] and that children's performance on working memory measures improves with age [57]. It is not clear, however, how the interaction between cognitive and speech motor skills changes with increased age. All participants were monolingual speakers of American English, with no reported histories of speech, language, hearing problems, or neurological disorders. The study was approved by Institutional Review Board at New York University and informed consent was obtained from all participants and their parents.

Speech and language skills were formally and informally assessed. Speech production skills were examined using the Goldman Fristoe Test of Articulation-2 (GFTA-2) [58] and through a conversational speech sample. Receptive and expressive language skills were evaluated using the Clinical Evaluation of Language Fundamentals (CELF-P; CELF-3) [59, 60]. The Verbal Motor Production Assessment in Children (VMPAC) [61] was used to examine oral motor skills. The participants demonstrated age appropriate speech, language, and oral motor skills on these measures. All participants passed a pure-tone hearing screening presented bilaterally at 25 dB at .5, 1, 2, and 4 kHz.

In order to rule out any major difficulties with NW repetition, we included a NW repetition task in our pretest procedure. NW repetition was assessed using the Children's Nonword Repetition Test (CNrep) [57]. NWs were recorded by a speaker of American English and the task was completed in a sound treated booth. Participants were instructed to listen to and repeat each NW presented through speakers. Responses were recorded and percent consonants correct (PCC) [62] scores were calculated separately by two trained graduate students in speech language pathology. Mean PCC for the repeated targets in the CNrep test were 88.88% for younger group and 98.1% for the older group.

2.2. Signal Recording and Processing. Jaw movement was tracked in three dimensions using a motion capture system, Vicon 460 [63]. Ten reflective markers (each 3 mm in diameter) were placed on the face. Five markers were used to track lip and jaw movement and were placed on the midline of the vermilion border of the upper lip, midline of the vermilion border of the lower lip, superior to the mental protuberance of the mandible, and on the corners of the mouth. Five reference markers were used to account for head movement and rotation, which were placed on the nose, nasion, and forehead. Jaw movement was calculated by subtracting y coordinates from the stationary points on the forehead.

Kinematic data analysis was conducted using MATLAB, version 7.2 [64]. The system tracked reflective markers at a sampling rate of 120 frames per second. Audio recordings were made using a digital minidisc recorder, M-Audio, MicroTrack 2496. Participants wore a lapel microphone, Audio-Technica, Model AT831W, which was placed on the shirt approximately 6 inches from the mouth. All recordings were made in a sound attenuated audiometric booth at New York University.

Table 1: Phonotactic probability.

	Real words	Nonwords
PAIR 1		
Phonotactic probability	0.44	0.42
Biphonotactic probability	0.01	0.02
PAIR 2		
Phonotactic probability	0.36	0.37
Biphonotactic probability	0.01	0.3
PAIR 3		
Phonotactic probability	0.44	0.35
Biphonotactic probability	0.01	0.02
PAIR 4		
Phonotactic probability	0.55	0.44
Biphonotactic probability	0.02	0.03

2.3. Data Collection and Procedures. Participants listened to recordings of a monolingual American-English-speaking adult producing RWs and NWs. They were told that they would be hearing "real words" and "funny, made-up words" and were asked to repeat the structures exactly as they heard them using their habitual speaking rate and loudness. Referents were not provided for the RWs or NWs. Eight practice items (four RWs and four NWs) were administered. If a participant requested additional practice items or if the experimenters felt that they did not completely understand the task, the practice items were repeated. This occurred in two of the younger participants, who did not have any difficulty completing the experimental protocol after additional practice. The tokens included two RWs (i.e., "baby muppet"/bebi mʌpɪt/ and "peppy mama muppet"/pɛpi mamə mʌpɪt/) and two NWs (i.e., "babu mepid"/bæbə mɛpɪd/ and "bebu pupu bepid"/bɛbə pʌpə bɛpɪd/), which were presented in a randomized order in terms of word type (RW versus NW) and length (four versus six syllables). By the end of the session, fifteen productions of each token were obtained from the subjects. These RW and NW structures were selected because they include bilabial phonemes, /p/, /b/, and /m/, that allowed lip and jaw movements to be visualized. NWs did not contain any syllables that constituted real words. RWs and NWs were matched in number of syllables, stress pattern, linguistic complexity, and phonotactic probability (Table 1) [14, 65–67]. The latter is an index of the probability of a segment occurring in combination with one or two other segments in the sequential arrangement in the word. A higher value means a higher probability of occurrence or the combinations of segments included in the targets.

2.4. Analyses

2.4.1. Perceptual Judgments. A graduate student in speech language pathology, naïve to the purpose of the experiment, listened to and transcribed all of the productions of each token from each speaker. A second graduate student transcribed 10% of speaker productions from randomly chosen participants. Interrater agreement on PCC scores was computed on a segment to segment basis and reached 99%. The perceptual analyses included measures of percent consonant correct (PCC) [62] and percent vowel correct (PVC). PCC and PVC were calculated separately for each participant (1) for the first production of each token (total of 64 utterances: 4 tokens × 1 production × 16 participants) and (2) across all productions of each token (total of 960 utterances: 4 tokens × 15 productions × 16 participants). The rationale for examining first productions was to calculate segmental accuracy of the productions of each NW the first time they were produced. Segmental accuracy was also examined across all productions to obtain a more comprehensive index of articulation performance.

2.4.2. Kinematic Analysis. The kinematic analysis was based upon accurate productions identified through the perceptual analysis. Only accurate productions were included to ensure that any observed kinematic differences are due to underlying changes in speech motor control that are independent of articulation errors. Given the high variability of children's articulator movements, multiple productions of the same token from each child were analyzed, rather than also examining first productions of tokens. The first eight productions, in which segmental and suprasegmental components were judged to be accurate and in which all reflective markers were visible, were included in the analyses. Eight productions were selected as this was the greatest number of productions across all participants that met the criteria mentioned above. Productions were eliminated due to one or more of the following factors: consonant/vowel error, suprasegmental error (e.g., equal stress), and missing reflective markers. Segmental and suprasegmental errors were more prevalent in the younger children. A total of 512 utterances were included in the kinematic analyses (4 targets × 8 productions × 16 participants).

The acoustic and kinematic signals from each production were aligned. The acoustic signal was used to help identify articulator movement associated with each word. Kinematic records of the jaw were then analyzed. The onset and offset of movement were based upon velocity minima in the jaw kinematic trace. The onset of movement was selected as the point of minimum velocity into oral opening for the first vowel in the word. The point of minimum velocity into opening for the word final vowel was chosen as the movement offset. Total movement duration was calculated as the time between movement onset and movement offset in the jaw velocity trajectory (Figure 1).

Movement trajectory stability was examined to explore whether there are changes in stability of the underlying movement pattern associated with the production of real words and nonwords across development once differences in

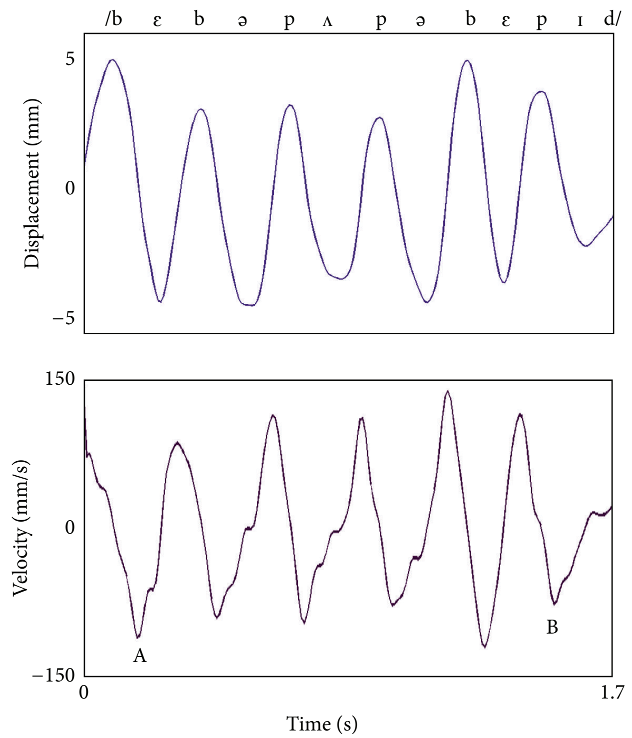


Figure 1: Kinematic traces of jaw velocity and jaw displacement corresponding to the utterance /bɛbə pʌpə bɛpɪd/. Duration measures are based upon velocity points. Total movement duration is calculated as the time between movement onset and movement offset. Movement onset is the velocity minima associated with oral opening for the first vowel in the word (i.e., /ɛ/) and movement offset is the velocity minima associated with oral opening for the final vowel in the word (i.e., /ɪ/) (points A to B).

absolute time and amplitude were removed. The onset and offset of movement identified in the velocity trace for the total duration measure was also used to segment displacement data for the movement stability analysis. Segmented displacement traces were normalized for amplitude and time. For each displacement trace, amplitude normalization was achieved by subtracting the mean of the displacement record and dividing by its standard deviation. Time normalization was achieved by using a cubic spline procedure to interpolate each waveform onto a time base of 1,000 points. The spatiotemporal index (STI) was then calculated to examine stability in movement trajectories across repeated productions of target utterances [68]. The STI was computed by calculating standard deviations at 2% intervals across repetitions of the time and amplitude normalized displacement traces. The STI is the cumulative sum of these 50 standard deviations. The STI indicates the degree to which the set of trajectories converge onto one fundamental movement pattern [69].

2.4.3. Statistical Analyses. Means and standard deviations were calculated for PCC, PVC, jaw movement duration (DUR), and STI, for each participant. Repeated measures analyses of variance (ANOVAs) were performed to examine the effects of the between-subjects variable *Group* (younger or older) and the within-subjects variables *Word Type* (RW or NW) and *Length* (four or six syllables) on PCC, PVC, DUR, and STI. Interactions between *Word Type*, *Length*, and *Group* were also measured. Each dependent measure was examined separately. When the main effect of *Word Type* was significant, pairwise contrasts were performed to explore differences between RWs and NWs within each experimental group. A Bonferroni correction factor was used to account for multiple comparisons within each variable (RW versus NW in the younger group and in the older group), which adjusted the alpha level to 0.025.

3. Results

3.1. Perceptual Accuracy. PCC and PVC scores for each RW and NW for first repetitions, as well as across all repetitions for each participant, are shown in Table 2.

3.2. First Productions. Comparisons of first productions revealed a trend of greater consonant and vowel accuracy in RWs than NWs. Consonant and vowel accuracy were higher for RWs than NWs as evidenced by a significant main effect of *Word Type* on PCC, $F(1,14) = 13.75$, $p = 0.002$, $\eta_p^2 = 0.495$, and PVC, $F(1,14) = 9.95$, $p = 0.007$, $\eta_p^2 = 0.415$. PCCs and PVCs were similar between four- and six-syllable structures and between experimental groups. Thus, there were no significant main effects of *Length* or *Group* on PCC or PVC. Further, there were no significant interactions between *Word Type*, *Length*, and *Group*.

3.2.1. All Productions. When all productions were examined, consonant accuracy was significantly higher in the RWs as compared to the NWs. This observation was supported by a significant main effect of *Word Type* on PCC, $F(1,14) = 18.283$, $p = 0.001$, $\eta_p^2 = 0.574$. There were no significant main effects of *Length* or *Group* on PCC as consonant accuracy was similar between four- and six-syllable tokens for both groups. There were no significant two- or three-way interactions between *Word Type*, *Length*, and *Group*.

All participants produced vowels with greater accuracy in RWs than in NWs. This finding was supported by a significant main effect of *Word Type* on PVC, $F(1,14) = 10.45$, $p = 0.006$, $\eta_p^2 = 0.427$. The difference in vowel accuracy between RWs and NWs was evident in both four- and six-syllable structures, as well as in the younger and older groups. Thus, there were no significant main effects of *Group* or *Length*. Interactions between *Word Type*, *Length,* and *Group* were not significant.

3.3. Articulator Movement. Total jaw movement duration (DUR) and movement stability (STI) were calculated from all accurate productions of RWs and NWs.

3.3.1. Movement Duration. Jaw movement duration was longer in NWs than RWs during the production of four- and six-syllable structures across both groups (Figure 2).

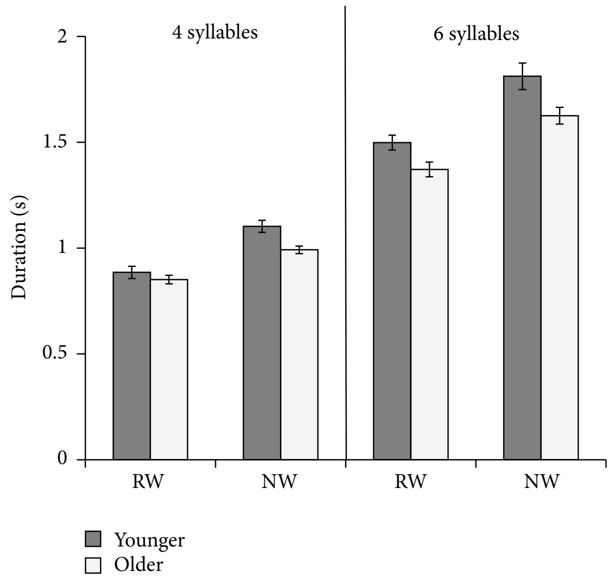


Figure 2: Mean total jaw movement duration and standard error in the younger and older groups producing four-syllable (left column) and six-syllable (right column) real words (RWs) and nonwords (NWs).

As expected, movement duration was longer for six- than four-syllable structures. These findings were supported by significant main effects of *Word Type*, $F(1,14) = 79.49$, $p < 0.001$, $\eta_p^2 = 0.850$, and *Length*, $F(1,14) = 1082.32$, $p < 0.001$, $\eta_p^2 = 0.987$, on jaw movement duration. There was also a significant main effect of *Group* on duration, $F(1,14) = 11.95$, $p = 0.004$, $\eta_p^2 = 0.460$. Participants in the younger group produced utterances with significantly longer jaw movement durations as compared to the older group. A significant interaction between *Length* and *Group* was found, $F(1,14) = 5.04$, $p = 0.041$, $\eta_p^2 = 0.265$, where the difference in duration between four- and six-syllable tokens was greater in the younger than in the older group. Further, the interaction between *Length* and *Word Type* was significant, $F(1,14) = 6.24$, $p = 0.026$, $\eta_p^2 = 0.308$, as differences between NWs and RWs were larger in the six-syllable than in the four-syllable tokens. Post hoc tests revealed significantly longer movement duration in NWs than RWs in both the younger group (mean difference = 0.265, $p < 0.001$) and the older group (mean difference = 0.198, $p < 0.001$). A three-way interaction between *Word Type*, *Length,* and *Group* was not significant.

3.3.2. Movement Stability. Comparisons of spatiotemporal stability were performed by examining changes in the jaw STI across age groups in four- and six-syllable RWs and NWs. High STIs indicate greater spatiotemporal variability and low STIs represent more stability across movement trajectories. STIs were higher in NWs than RWs in both groups (Figure 3). This finding was supported by a significant main effect of *Word Type* on jaw STI, $F(1,14) = 8.20$, $p = 0.013$, $\eta_p^2 = 0.369$. There was no main effect of *Length* on STI as movement

Table 2: Percent consonant correct (PCC) and percent vowel correct (PVC) from first repetitions and all repetitions.

Group	First repetition				All repetitions				First repetition				All repetitions			
	4-syllable RW /bebi mʌpɪt/		4-syllable NW /bæbə mɛpɪd/		4-syllable RW /bebi mʌpɪt/		4-syllable NW /bæbə mɛpɪd/		6-syllable RW /pɛpi mamə mʌpɪt/		6-syllable NW /bɛbə pʌpə bɛpɪd/		6-syllable RW /pɛpi mamə mʌpɪt/		6-syllable NW /bɛbə pʌpə bɛpɪd/	
	PCC	PVC	PCC	PVC	PCC	PVC	PCC	PVC	PCC	PVC	PCC	PVC	PCC	PVC	PCC	PVC
Younger																
S1	100.0	100.0	80.0	100.0	100.0	100.0	80.0	100.0	86.0	67.0	86.0	67.0	100.0	83.0	100.0	100.0
S2	100.0	100.0	80.0	100.0	100.0	100.0	60.0	100.0	100.0	83.0	86.0	67.0	100.0	100.0	71.0	63.0
S3	100.0	100.0	60.0	75.0	100.0	100.0	100.0	100.0	86.0	100.0	86.0	100.0	100.0	100.0	100.0	100.0
S4	60.0	100.0	80.0	100.0	60.0	100.0	14.0	67.0	86.0	100.0	57.0	83.0	71.0	100.0	60.0	75.0
S5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S6	100.0	100.0	60.0	100.0	86.0	100.0	80.0	100.0	100.0	100.0	28.0	100.0	100.0	100.0	29.0	83.0
S7	100.0	100.0	80.0	100.0	100.0	100.0	80.0	100.0	100.0	100.0	86.0	100.0	100.0	100.0	86.0	100.0
S8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	74.0	85.0
Mean	**95.0**	**100.0**	**80.0**	**96.9**	**93.3**	**100.0**	**76.8**	**95.9**	**94.8**	**93.8**	**78.6**	**89.6**	**96.4**	**97.9**	**77.5**	**88.3**
SD	**14.1**	**0.0**	**15.1**	**8.8**	**14.3**	**0.0**	**29.0**	**11.7**	**7.2**	**12.3**	**24.4**	**15.1**	**10.3**	**6.0**	**24.8**	**14.2**
Older																
S9	100.0	100.0	100.0	100.0	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	86.0	83.0
S10	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	71.0	100.0	100.0	100.0	100.0	67.0
S11	100.0	100.0	100.0	100.0	100.0	100.0	80.0	75.0	100.0	100.0	43.0	67.0	100.0	100.0	57.0	100.0
S12	100.0	100.0	80.0	100.0	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
S13	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	83.0	100.0	100.0	100.0	83.0
S14	100.0	100.0	80.0	75.0	100.0	100.0	80.0	75.0	100.0	100.0	100.0	100.0	100.0	100.0	86.0	100.0
S15	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	86.0	83.0	100.0	100.0	86.0	83.0
S16	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	71.0	100.0	100.0	100.0	87.0	100.0
Mean	**100.0**	**100.0**	**92.5**	**96.9**	**100.0**	**100.0**	**90.0**	**93.8**	**100.0**	**100.0**	**83.9**	**91.6**	**100.0**	**100.0**	**87.8**	**89.5**
SD	**0.0**	**0.0**	**10.4**	**8.8**	**0.0**	**0.0**	**10.7**	**11.6**	**0.0**	**0.0**	**20.9**	**12.6**	**0.0**	**0.0**	**14.2**	**12.4**

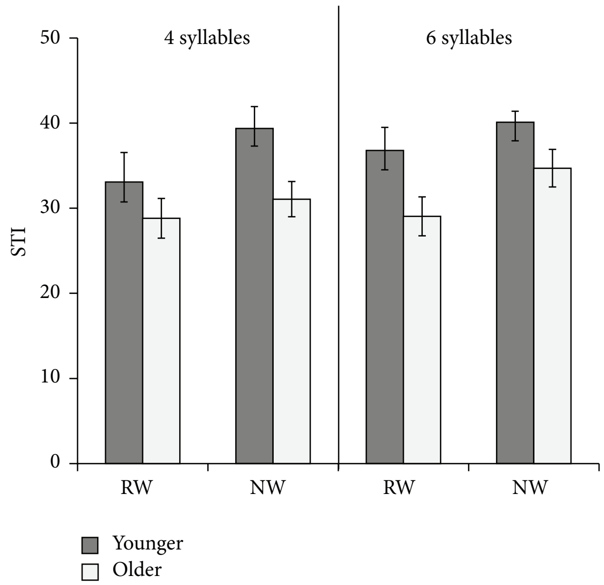


FIGURE 3: Mean jaw STI and standard error in the younger and older groups producing four-syllable (left column) and six-syllable (right column) real words (RWs) and nonwords (NWs).

stability was similar between four- and six-syllable structures with each word type. Group effects were evident, however, as STIs were higher in the younger than the older participants, $F(1,14) = 5.75$, $p = 0.031$, $\eta_p^2 = 0.291$. There were no significant interactions between *Word Type*, *Length*, and *Group*.

4. Discussion

Our main interest was to determine whether speech motor patterns of children and adolescents were equally vulnerable to increases in cognitive demands. Specifically, we examined patterns of movement duration and variability across multiple accurate productions of RWs and NWs. Our findings illustrate that both jaw movement duration and stability differed between RWs and NWs, although to varying degrees in the children and adolescents based on word length. These results support the hypothesis that articulator movement duration and variability will increase during tasks with greater PSTM demands.

4.1. Nonword Repetition Accuracy. Accuracy of production has been the traditional measure of NW repetition and considered an index of phonological short-term memory skills provided one production only was permitted in order to avoid a practice effect. In the present work, first productions were found to have higher consonant and vowel accuracy in RWs than in NWs across word lengths and groups, as expected. These results are consistent with numerous studies showing that accuracy scores are higher during repetition of RWs compared with NWs in children with and without language impairment [8, 14, 19]. Our results showing similar consonant accuracy between younger and older children did not correspond with results from past research, which reported increased NW repetition accuracy with age [4, 12]. These researchers viewed a developmental trend as evidence that PSTM skills continue to develop with increased age during childhood and adolescence. A careful inspection of the data in Table 2 illustrates a trend for the differences between RWs and NWs to be greater in the younger than the older children for the four-syllable tokens yet similar between the six-syllable tokens. Thus, it is plausible that our findings may have more closely mirrored those from earlier studies if a larger participant pool and more complex NWs were examined. These limitations are discussed in greater detail below.

Differences between RW and NW accuracy that were seen during first productions were also evident when all productions were analyzed. Across all productions, consonant accuracy remained higher in RWs compared to NWs in both groups. Vowel accuracy was also higher in RWs than NWs in both groups of children. Overall, the consistency of accuracy between first productions and all productions suggests a sustained effect of a more cognitively challenging production task. Nonetheless, both groups of participants were able to achieve many accurate productions of both the RWs and NWs.

4.2. Influence of Word Type on Articulator Movement. Jaw movement duration was significantly longer in NWs compared to RWs across both word lengths and age groups, illustrating that all participants were taxed by NW production and increased movement duration to complete this task. These results suggest that both children and adolescents may compensate for the increase in cognitive demand by modifying the temporal control of speech movements. Our findings also support past research that has shown that articulatory control follows a protracted course of development [40, 43].

Age-related differences in temporal control influenced patterns seen between the younger and older children. The older children consistently produced both RWs and NWs with shorter movement durations than the younger children, which is consistent with past research showing that movement duration decreases with age [32, 33, 37, 39, 41]. Duration differences between RWs and NWs in both the four- and six-syllable tokens were significantly greater in the younger than in the older children. This suggests that the articulatory patterns of the younger children were more vulnerable to task demands. The significant interaction between *Length* and *Word Type* resulted from the differences between RWs and NWs being greater in the six-syllable tokens than in the four-syllable tokens, a finding that was more pronounced in the younger than the older participants. Taken together, these results illustrate that more mature temporal control seen in the older as compared to the younger group may facilitate greater consonant and vowel accuracy during the NW task. These findings are consistent with evidence that children modify temporal control during other speaking tasks, such as marking linguistic [37] and prosodic contrasts [35, 36].

Jaw movement variability, as measured by the spatiotemporal index (STI), was also examined across multiple accurate productions of tokens. A high spatiotemporal index (STI) indicates more jaw movement variability across productions.

Movement variability was higher in the younger than older group for both NWs and RWs. Comparisons by *Word Type*, however, revealed that movement variability was greater during repetition of NWs compared with repetition of RWs in both groups of children. Thus, both the younger and older participants were challenged by the NWs even though they achieved perceptually accurate productions of the NWs. This result suggests that developing speakers may continue to alter speech motor planning and execution processes to meet the cognitively taxing demands of a NW task even during adolescence. Greater variability in the productions of NWs may be attributed to their less mature speech motor skills. An alternate explanation is that greater variability may reflect more movement flexibility required for children and adolescents to achieve accurate NW production. Taken together, these results are consistent with past research showing that articulatory control becomes more stable with maturation [32, 37, 42, 43] and continues to stabilize into adolescence [40].

It is important to highlight that differences in movement patterns observed between RWs and NWs were seen even though consonant and vowel accuracy were at 100% for these tokens. This suggests that demands on PSTM influence articulator movement although changes may not lead to perceptually detectable differences. This finding is particularly interesting given that the NW tokens were relatively simple in phonetic make-up. This observation leads us to speculate that there may be differences in children and adolescents pertaining to how the system reorganizes itself in order to complete more challenging speech production tasks. Such differences may lead to trade-off effects between linguistic processing and speech motor control that are measurable although not necessarily perceptually detectable. Just as others have suggested that there are trade-offs between linguistic levels during language production [70–74], the current findings suggest reciprocity between linguistic processing and speech motor control in children and adolescents.

Together, these findings suggest that when PSTM is taxed, children and adolescents compensate differently in order to achieve accurate word production. Both groups modified temporal control and movement variability to achieve accurate productions of RW and NW tokens. Our results showing that task demands influence articulatory control even during adolescence are not surprising as several studies have reported that adults are also sensitive to increases in cognitive demands during speaking tasks [45, 46]. It could be argued that examining articulatory control across repeated productions of NWs (rather than first productions of NWs) is a practice exercise rather than a word repetition task. We examined results from first productions in terms of accuracy but, in order to obtain measures from articulatory patterns, multiple productions were necessary. Our results show differences between articulatory patterns during production of RWs versus NWs in spite of multiple productions indicating an effect of cognitive load on speech motor control in spite of a possible practice effect.

4.3. Subskills Involved in Nonword Repetition. Factors, such as subvocal rehearsal and vocabulary growth, have been proposed to play a role in NW repetition. According to Gathercole [11], subvocal rehearsal, examined during first productions of a novel target, is not typically present in children younger than seven years. We found a significant difference in accuracy of first productions of RWs versus NWs but no significant differences between the age groups. This result indicates that both groups may have used this strategy to support PSTM, which might be expected since the mean age of our younger group was almost seven.

NW repetition is a complex psycholinguistic task and there is more to be learned about the contributing processes involved in performance in speakers across ages and clinical categories. Our results lend support to the view that there is a mutually dependent relationship between speech motor processes and the lexical level of processing of a target for production. It is interesting to observe that even when targets are well matched and are not containing phoneme combinations of high complexity or low probability, word type seems to affect children's accuracy of production and the processes involved in articulatory control. Adams and Gathercole [75] stated that output processes should not be used to "explain away" the relationship between phonological memory and language development. The contribution of speech motor processes to NW repetition skills in children is an area, which needs to be further investigated, however.

4.4. Methodological Limitations. Findings of the present work may be influenced by the construction of RW and NW targets. Although construction of NWs is known to be challenging (for a discussion, see [14, 65]), we made every attempt to match the RWs and NWs by syllable number, phonetic make-up, and phonotactic probability. Our longest structures were six syllables long but did not contain any clusters or phoneme combinations with low phonotactic probability. These factors have been described in the literature to tax NW repetition skills [7, 76]. Edwards and colleagues [77] reported higher accuracy in NW repetition for high frequency phonological patterns compared with low-frequency patterns in children. This effect decreased with increasing age, a result the authors attributed to larger vocabularies in the older children.

A second issue that should be pointed out is that construction of NWs matched to a selection of RWs known to young children and suitable for the kinematic procedures proved to be a significant challenge. We avoided syllables consisting of RWs when constructing our NWs. Further, bilabials were used in the tokens because facial tracking technology records movements from visual structures, such as the lips and jaw, which are involved in bilabial production. Using movement tracking was advantageous as this was a direct approach to investigating output processes involved in NW repetition. In comparison, many previous studies have employed indirect approaches, such as speed of articulation (i.e., speech rate) in producing sets of words [8, 22, 78, 79] or correlations between accuracy measures and results on oral motor tasks [14, 31].

5. Conclusion

In the present study, children and adolescents showed a lower level of consonant and vowel accuracy during NW repetition compared with repetition of RWs. Jaw movement duration was longer and variability of articulator movements was greater in NWs compared to RWs. Young children showed longer duration of jaw movements than adolescents. It is possible that a range of skills develop between the ages of six and fourteen, which support NW repetition skills. NW repetition is a complex task requiring a range of subskills. Future studies should continue to examine the role of less studied skills supporting NW repetition in children, such as speech motor control and orthographic skills. In addition, changes in articulator movement related to practice and learning during production of novel items should be further investigated.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to the participants and their parents. They also wish to thank Nicole Barish for her involvement in the data collection as well as Guemali Viera and Jessica Jones for their assistance with data analysis.

References

[1] J. M. Iverson, "Developing language in a developing body: the relationship between motor development and language development," *Journal of Child Language*, vol. 37, no. 2, pp. 229–261, 2010.

[2] J. M. Iverson and B. A. Braddock, "Gesture and motor skill in relation to language in children with language impairment," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 1, pp. 72–86, 2011.

[3] D. V. M. Bishop, T. North, and C. Donlan, "Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 37, no. 4, pp. 391–403, 1996.

[4] G. Conti-Ramsden and Z. Simkin, "Non-word repetition and grammatical morphology: normative data for children in their final year of primary school," *International Journal of Language & Communication Disorders*, vol. 36, no. 3, pp. 395–404, 2001.

[5] A. G. Kamhi and H. W. Catts, "Toward an understanding of developmental language and reading disorders," *The Journal of Speech and Hearing Disorders*, vol. 51, no. 4, pp. 337–347, 1986.

[6] A. G. Kamhi, H. W. Catts, D. Mauer, K. Apel, and B. F. Gentry, "Phonological and spatial processing abilities in language- and reading-impaired children," *Journal of Speech and Hearing Disorders*, vol. 53, no. 3, pp. 316–327, 1988.

[7] S. E. Gathercole and A. D. Baddeley, "The role of phonological memory in vocabulary acquisition: a study of young children learning new names," *The British Journal of Psychology*, vol. 81, pp. 439–454, 1990.

[8] S. E. Gathercole and A. D. Baddeley, "Phonological memory deficits in language disordered children: is there a causal connection?" *Journal of Memory and Language*, vol. 29, no. 3, pp. 336–360, 1990.

[9] S. E. Gathercole and A. D. Baddeley, "Phonological working memory: a critical building block for reading development and vocabulary acquisition?" *European Journal of Psychology of Education*, vol. 8, no. 3, pp. 259–272, 1993.

[10] H. W. Catts, S. M. Adlof, T. P. Hogan, and S. E. Weismer, "Are specific language impairment and dyslexia distinct disorders?" *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 6, pp. 1378–1396, 2005.

[11] S. E. Gathercole, "Nonword repetition and word learning: the nature of the relationship," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 513–543, 2006.

[12] S. E. Gathercole, C. S. Willis, H. Emslie, and A. D. Baddeley, "Phonological Memory and Vocabulary Development During the Early School Years: A Longitudinal Study," *Developmental Psychology*, vol. 28, no. 5, pp. 887–898, 1992.

[13] J. W. Montgomery, "Sentence comprehension in children with specific language impairment: the role of phonological working memory," *Journal of Speech and Hearing Research*, vol. 38, no. 1, pp. 187–199, 1995.

[14] B. Sahlén, C. Reuterskiöld-Wagner, U. Nettelbladt, and K. Radeborg, "Non-word repetition in children with language impairment—pitfalls and possibilities," *International Journal of Language and Communication Disorders*, vol. 34, no. 3, pp. 337–352, 1999.

[15] B. Sahlén, C. Reuterskiöld-Wagner, U. Nettelbladt, and K. Radeborg, "Language comprehension and non-word repetition in children with language impairment," *Clinical Linguistics and Phonetics*, vol. 13, no. 5, pp. 369–380, 1999.

[16] M. J. Snowling, "Phonemic deficits in developmental dyslexia," *Psychological Research*, vol. 43, no. 2, pp. 219–234, 1981.

[17] J. W. Montgomery, "Working memory and comprehension in children with specific language impairment: what we know so far," *Journal of Communication Disorders*, vol. 36, no. 3, pp. 221–231, 2003.

[18] A. J. O. Whitehouse, J. G. Barry, and D. V. M. Bishop, "Further defining the language impairment of autism: is there a specific language impairment subtype?" *Journal of Communication Disorders*, vol. 41, no. 4, pp. 319–336, 2008.

[19] J. W. Montgomery, B. M. Magimairaj, and M. C. Finney, "Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment," *American Journal of Speech-Language Pathology*, vol. 19, no. 1, pp. 78–94, 2010.

[20] A. D. Baddeley, *Working Memory*, vol. 11 of *Oxford Series*, Clarendon Press, Oxford, UK, 1986.

[21] A. D. Baddeley and G. J. Hitch, "Working memory," in *The Psychology of Learning and Motivation*, G. Bower, Ed., pp. 47–90, Academic Press, San Diego, Calif, USA, 1974.

[22] S. E. Gathercole, A.-M. Adams, and G. J. Hitch, "Do young children rehearse? An individual-differences analysis," *Memory and Cognition*, vol. 22, no. 2, pp. 201–207, 1994.

[23] A. D. Baddeley, N. Thomson, and M. Buchanan, "Word length and the structure of short-term memory," *Journal of Verbal Learning and Verbal Behavior*, vol. 14, no. 6, pp. 575–589, 1975.

[24] K. S. Marton and R. G. Schwartz, "Working memory capacity and language processes in children with specific language impairment," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 5, pp. 1138–1153, 2003.

[25] L. Leonard, *Children with Specific Language Impairment*, MIT Press, Cambridge, Mass, USA, 2nd edition, 2014.

[26] K. G. Estes, J. L. Evans, M. W. Alibali, and J. R. Saffran, "Can infants map meaning to newly segmented words? Statistical segmentation and word learning," *Psychological Science*, vol. 18, no. 3, pp. 254–260, 2007.

[27] D. V. M. Bishop, B. B. Brown, and J. Robson, "The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals," *Journal of Speech and Hearing Research*, vol. 33, no. 2, pp. 210–219, 1990.

[28] D. V. Bishop and J. Robson, "Accurate non-word spelling despite congenital inability to speak: phoneme-grapheme conversion does not require subvocal articulation," *The British Journal of Psychology*, vol. 80, pp. 1–13, 1989.

[29] D. V. Bishop and J. Robson, "Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of 'articulatory coding'," *The Quarterly Journal of Experimental Psychology Section A*, vol. 41, no. 1, pp. 123–140, 1989.

[30] D. V. M. Bishop, C. V. Adams, and C. F. Norbury, "Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins," *Genes, Brain and Behavior*, vol. 5, no. 2, pp. 158–169, 2006.

[31] S. Krishnan, K. J. Alcock, E. Mercure et al., "Articulating novel words: children's oromotor skills predict non-word repetition abilities," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 6, pp. 1800–1812, 2013.

[32] L. Goffman and A. Smith, "Development and phonetic differentiation of speech movement patterns," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, no. 3, pp. 649–660, 1999.

[33] J. R. Green, C. A. Moore, M. Higashikawa, and R. W. Steeve, "The physiologic development of speech motor control: lip and jaw coordination," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 1, pp. 239–255, 2000.

[34] J. R. Green, C. A. Moore, and K. J. Reilly, "The sequential development of jaw and lip control for speech," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 1, pp. 66–79, 2002.

[35] M. I. Grigos and R. Patel, "Articulator movement associated with the development of prosodic control in children," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 4, pp. 706–715, 2007.

[36] M. I. Grigos and R. Patel, "Acquisition of articulatory control for sentential focus in children," *Journal of Phonetics*, vol. 38, no. 4, pp. 706–715, 2010.

[37] M. I. Grigos, J. H. Saxman, and A. M. Gordon, "Speech motor development during acquisition of the voicing contrast," *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 4, pp. 739–752, 2005.

[38] J. Sasisekaran, A. Smith, N. Sadagopan, and C. Weber-Fox, "Nonword repetition in children and adults: effects on movement coordination," *Developmental Science*, vol. 13, no. 3, pp. 521–532, 2010.

[39] A. Smith and L. Goffman, "Stability and patterning of speech movement sequences in children and adults," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 1, pp. 18–30, 1998.

[40] B. Walsh and A. Smith, "Articulatory movements in adolescents: evidence for protracted development of speech motor control processes," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 6, pp. 1119–1133, 2002.

[41] S. G. Sharkey and J. W. Folkins, "Variability of lip and jaw movements in children and adults: implications for the development of speech motor control," *Journal of Speech and Hearing Research*, vol. 28, no. 1, pp. 8–15, 1985.

[42] M. I. Grigos, "Changes in articulator movement variability during phonemic development: a longitudinal study," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 1, pp. 164–177, 2009.

[43] A. Smith and H. N. Zelaznik, "Development of functional synergies for speech motor coordination in childhood and adolescence," *Developmental Psychobiology*, vol. 45, no. 1, pp. 22–33, 2004.

[44] K. J. Maner, A. Smith, and L. Grayson, "Influences of utterance length and complexity on speech motor performance in children and adults," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 2, pp. 560–573, 2000.

[45] C. Dromey and A. Benson, "Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 5, pp. 1234–1246, 2003.

[46] B. Walsh, A. Smith, and C. Weber-Fox, "Short-term plasticity in children's speech motor systems," *Developmental Psychobiology*, vol. 48, no. 8, pp. 660–674, 2006.

[47] C. T. McMillan, M. Corley, and R. J. Lickley, "Articulatory evidence for feedback and competition in speech production," *Language and Cognitive Processes*, vol. 24, no. 1, pp. 44–66, 2009.

[48] M. Goldrick and S. E. Blumstein, "Cascading activation from phonological planning to articulatory processes: evidence from tongue twisters," *Language and Cognitive Processes*, vol. 21, no. 6, pp. 649–683, 2006.

[49] B. S. Munson and N. P. Solomon, "The effect of phonological neighborhood density on vowel articulation," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 5, pp. 1048–1058, 2004.

[50] M. Baese-Berk and M. Goldrick, "Mechanisms of interaction in speech production," *Language and Cognitive Processes*, vol. 24, no. 4, pp. 527–554, 2009.

[51] L. Heisler, L. Goffman, and B. Younger, "Lexical and articulatory interactions in children's language production," *Developmental Science*, vol. 13, no. 5, pp. 722–730, 2010.

[52] D. V. M. Bishop, "Phonological short-term memory and syntactic impairment in specific language impairment," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 545–547, 2006.

[53] J. A. Bowey, "Clarifying the phonological processing account of nonword repetition," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 548–552, 2006.

[54] E. Service, "Phonological networks and new word learning," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 581–584, 2006.

[55] S. E. Gathercole, "Nonword repetition and word learning: the nature of the relationship," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 513–543, 2006.

[56] A. Smith, "Speech motor development: integrating muscles, movements, and linguistic units," *Journal of Communication Disorders*, vol. 39, no. 5, pp. 331–349, 2006.

[57] S. E. Gathercole, C. S. Willis, A. D. Baddeley, and H. Emslie, "The children's test of nonword repetition: a test of phonological working memory," *Memory*, vol. 2, no. 2, pp. 103–127, 1994.

[58] R. Goldman and M. Fristoe, *Goldman-Fristoe Test of Articulation*, American Guidance Service, Circle Pines, Minn, USA, 2nd edition, 2002.

[59] E. H. Wiig, W. Secord, and E. Semel, *Clinical Evaluation of Language Fundamentals-Preschool*, The Psychological Corporation, San Antonio, Tex, USA, 1992.

[60] E. Semel, E. H. Wiig, and W. A. Secord, *Clinical Evaluation of Language Fundamentals: CELF-3 Screening Test*, The Psychological Corporation, San Antonio, Tex, USA, 1995.

[61] D. Hayden and P. Square, *VMPAC Manual*, The Psychological Corporation, San Antonio, Tex, USA, 1999.

[62] L. D. Shriberg and J. Kwiatkowski, "Phonological disorders III: a procedure for assessing severity of involvement," *Journal of Speech and Hearing Disorders*, vol. 47, no. 3, pp. 256–270, 1982.

[63] Vicon Motion Analysis, 2001.

[64] *Matlab, Version 7.2 (Computer Software)*, MathWorks, Natick, Mass, USA, 2007.

[65] C. Reuterskiöld-Wagner, B. Sahlén, and A. Nyman, "Non-word repetition and non-word discrimination in Swedish preschool children," *Clinical Linguistics and Phonetics*, vol. 19, no. 8, pp. 681–699, 2005.

[66] M. S. L. Vitevitch and P. A. Luce, "A Web-based interface to calculate phonotactic probability for words and nonwords in English," *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 3, pp. 481–487, 2004.

[67] M. S. Vitevitch, "Nonword repetition and language learning disorders. A developmental contingency framework. [Peer commentary on Gathercole, S. Nonword repetition and word learning: the nature of the relationship]," *Applied Psycholinguistics*, vol. 27, no. 4, pp. 588–598, 2006.

[68] A. Smith, L. Goffman, H. N. Zelaznik, G. Ying, and C. McGillem, "Spatiotemporal stability and patterning of speech movement sequences," *Experimental Brain Research*, vol. 104, no. 3, pp. 493–501, 1995.

[69] A. Smith, M. Johnson, C. McGillem, and L. Goffman, "On the assessment of stability and patterning of speech movements," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 1, pp. 277–286, 2000.

[70] P. Menyuk and P. L. Looney, "A problem of language disorder: length versus structure," *Journal of Speech and Hearing Research*, vol. 15, no. 2, pp. 264–279, 1972.

[71] P. Menyuk and P. L. Looney, "Relationships among components of the grammar in language disorder," *Journal of Speech and Hearing Research*, vol. 15, no. 2, pp. 395–406, 1972.

[72] J. M. Panagos, M. E. Quine, and R. J. Klich, "Syntactic and phonological influences on children's articulation," *Journal of Speech and Hearing Research*, vol. 22, no. 4, pp. 841–848, 1979.

[73] J. M. Panagos and P. A. Prelock, "Phonological constraints on the sentence productions of language-disordered children," *Journal of Speech and Hearing Research*, vol. 25, no. 2, pp. 171–177, 1982.

[74] V. A. Schmauch, J. M. Panagos, and R. J. Klich, "Syntax influences the accuracy of consonant production in language-disordered children," *Journal of Communication Disorders*, vol. 11, no. 4, pp. 315–323, 1978.

[75] A.-M. Adams and S. E. Gathercole, "Limitations in working memory: implications for language development," *International Journal of Language and Communication Disorders*, vol. 35, no. 1, pp. 95–116, 2000.

[76] J. A. Coady and J. L. Evans, "Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI)," *International Journal of Language and Communication Disorders*, vol. 43, no. 1, pp. 1–40, 2008.

[77] J. Edwards, M. E. Beckman, and B. Munson, "The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 2, pp. 421–436, 2004.

[78] C. Hulme, N. Thomson, C. Muir, and A. Lawrence, "Speech rate and the development of short-term memory span," *Journal of Experimental Child Psychology*, vol. 38, no. 2, pp. 241–253, 1984.

[79] C. Hulme and V. Tordoff, "Working memory development: the effects of speech rate, word length, and acoustic similarity on serial recall," *Journal of Experimental Child Psychology*, vol. 47, no. 1, pp. 72–87, 1989.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 827463, 15 pages
http://dx.doi.org/10.1155/2015/827463

Research Article

Over-the-Counter Hearing Aids: A Lost Decade for Change

Zoe Yee Ting Chan and Bradley McPherson

Division of Speech and Hearing Sciences, Faculty of Education, University of Hong Kong, Pokfulam, Hong Kong

Correspondence should be addressed to Bradley McPherson; dbmcpher@hku.hk

Received 15 January 2015; Accepted 30 March 2015

Academic Editor: Haldun Oguz

Background. Hearing aids sold directly to consumers in retail stores or through the internet, without individual prescription by audiological professionals, are termed over-the-counter (OTC) devices. This study aimed to determine whether there was any change in the electroacoustic characteristics of OTC devices compared to research carried out a decade earlier. The previous results indicated that most OTC devices were low-frequency-emphasis devices and were unsuitable for elderly people with presbycusis, who were likely to be the major consumers of these products. *Methods*. Ten OTC devices were selected and their electroacoustic performance was measured. Appropriate clients for the OTC devices were derived, using four linear prescription formulae, and OTC suitability for elderly persons with presbycusis was investigated. *Results*. OTC electroacoustic characteristics were similar to those in the earlier study. Most OTC devices were not acoustically appropriate for potential consumers with presbycusis. Although several of the devices could match prescriptive targets for individuals with presbycusis, their poor electroacoustic performance—including ineffective volume control function, high equivalent input noise, and irregular frequency response—may override their potential benefit. *Conclusion*. The low-cost OTC devices were generally not suitable for the main consumers of these products, and there has been little improvement in the appropriateness of these devices over the past decade.

1. Background

Hearing aids that are sold directly to consumers in retail shops or through the internet, without customized prescription by audiological professionals, are termed over-the-counter (OTC) hearing aids [1, 2]. People who purchase OTC hearing aids do not receive the potential benefits provided by professional service, which include audiological assessment, counseling, hearing aid selection, hearing aid fitting, and hearing aid orientation. Without any prior audiological assessment, unnecessary amplification or delay in diagnosis of otologic problems may result [3]. In addition, the amplification characteristics of a hearing aid may not be appropriate for the client if the hearing aid is not programmed according to the individual's hearing loss.

In many developed economies, the sale of hearing aids is regulated. For example, in the United States, the provision of hearing aids is under the regulation of the United States Food and Drug Administration. Only licensed hearing healthcare professionals can provide hearing aids. Purchasers have to show a recent medical statement proving that they are hearing aid candidates or sign a waiver stating that they declined medical evaluation of their hearing loss before receiving the hearing aids [4]. These regulations intend to protect hearing aid users from any undiagnosed ear disorders and inappropriate amplification [3]. However, in numerous jurisdictions, there is no regulation of hearing aid sales, and this is the case in Hong Kong, as in many Asian localities. In Hong Kong, OTC hearing aids can be purchased in rehabilitation aid shops, electrical appliance stores, and general department stores and through the internet.

1.1. Elderly People as Potential Consumers of OTC Hearing Aids. An informal survey conducted in Hong Kong by Cheng [5] indicated that customers who purchased OTC hearing aids were primarily elderly people. The main reason for purchasing OTC hearing devices is probably their low cost as they are more affordable than conventional custom hearing aids. The cost of OTC hearing devices is variable but very often less than $US250; in contrast, the cost of conventional custom hearing aids is often above $US700 in Hong Kong.

In Hong Kong, the income of elderly people with low socioeconomic status mainly comes from government allowances for elderly individuals, and monthly income is about US$145 to US$282 [6]. Cost has been noted to be a major barrier to hearing aid use amongst elderly people in Hong Kong [7]. Although the Public Hospital Authority in Hong Kong provides subsidized, conventional hearing aids that often cost less than $US160, patient's first appointment waiting time is lengthy—approximately 23 to 85 weeks for a new case [8]. After the first physician visit, patients require further appointments for audiological assessment and hearing aid prescription/fitting, and this long waiting period may be another factor encouraging purchase of OTC devices.

1.2. Prevalence of Presbycusis and Its Common Audiometric Configuration. Presbycusis is a very common problem in the elderly population in both developed and developing countries [9]. According to the World Health Organization, the prevalence of hearing loss is approximately 33% among the global elderly population aged above 65 years [10]. In Hong Kong, it was estimated that the prevalence of presbycusis with moderate to profound hearing loss was 37.1% [11]. The common pattern of presbycusis is high frequency hearing loss and the degree of sensorineural hearing loss generally ranges from mild to moderately severe [12–15].

1.3. Electroacoustic Characteristics of OTC Hearing Aids in Previous Studies. Elderly people are likely to be the major users of OTC hearing aids. However, previous literature indicates that the quality of the OTC hearing aids and their effectiveness in matching the amplification needs of elderly people with presbycusis may be questionable.

1.3.1. Input-Output Characteristics. Two previous studies, one undertaken in Hong Kong (Cheng and McPherson) [2] and one in North America (Callaway and Punch) [1], found that all low-cost OTC hearing aids sampled were linear hearing aids. Linear amplification has an advantage in preserving the natural loudness difference in input signals. However, it is not suitable to people with sensorineural hearing loss who have recruitment [16]. Quiet sounds may not have enough gain while loud sounds may have too much gain and it is impractical for users to adjust a volume control continuously in order to maintain an appropriate gain in an acoustically dynamic environment. In addition, Cheng and McPherson [2] found that the volume control range was limited in some OTC hearing aids, making it difficult to adjust gain to appropriate levels. The peak clipping system associated with linear hearing aids also causes distortion to high level input signals [2].

1.3.2. Frequency Response. Most of the low-cost OTC hearing aids investigated in previous studies were low-frequency emphasis hearing aids with little gain or even no gain in high frequencies [1, 2]. Therefore, the OTC hearing aids tested are not suitable for elderly people with presbycusis, who typically have high-frequency, sloping hearing loss [12–15]. Speech understanding for individuals with presbycusis is not likely to be improved with low-frequency-emphasis hearing devices since consonants may remain inaudible with little gain, and amplified low-frequency background noise upwardly masks higher frequency sounds. In addition, overamplified low-frequency sounds may cause a perception of increased noise and potentially be harmful to residual hearing [1, 2]. Moreover, around half of the hearing devices tested in the two known OTC hearing aid studies showed sharp peaks of 8 dB or more in their frequency response [1, 2]. Both sound quality and speech intelligibility may be degraded with this type of response and the likelihood of feedback will also increase [16].

1.3.3. Equivalent Input Noise. EIN is the internal electronic noise of the hearing aid and it becomes audible and disturbing if it is too high [17]. Some of the hearing aids examined in previous studies had very high EIN that exceeded the 28 dB SPL target maximum set by the ANSI S3.22-1987 standard [1, 2]. As a result, some OTC hearing aids can generate internal noise that is perceptible to users.

1.3.4. Total Harmonic Distortion. THD reflects the amount of harmonic distortion generated by a hearing aid [16]. The THD of the OTC hearing aids tested in the two previous studies was generally within acceptable levels [1, 2].

1.3.5. Acoustic Feedback. Cheng and McPherson [1, 2] reported that the sampled OTC hearing aids could generally be turned to maximum output without feedback. However, this was probably related to their poor high frequency amplification characteristics and also because the hearing devices tested at that time were mainly body worn and had a long feedback path [2].

All of the above electroacoustic characteristics of OTC hearing aids were based on low-cost models retailing for less than US$100 each. The electroacoustic performance of some higher cost OTC hearing aids may be more suitable to people with presbycusis [1, 18].

1.4. Aims of the Present Study. Previously studied low-cost OTC hearing aids generally were incapable of providing sufficient appropriate gain to elderly people with presbycusis and their overall electroacoustic performance was not satisfactory [1, 2]. However, these studies were conducted over 14 years ago in Hong Kong and 6 years ago in North America. The performance of OTC devices may be improved with advancing technologies and nowadays such instruments may be more suitable for elderly consumers.

The present study aimed to determine whether there was any change in the electroacoustic characteristics of OTC hearing aids available in Hong Kong over the past decade. The electroacoustic performance of the current generation of OTC hearing aids in Hong Kong was examined. In addition, the study also aimed to determine the potential client groups for the OTC hearing aids and whether recent OTC hearing aids are appropriate for elderly people with presbycusis. The results obtained may have important implications for those who plan to purchase OTC hearing aids and for audiologists and other hearing health professionals advising patients with hearing impairment.

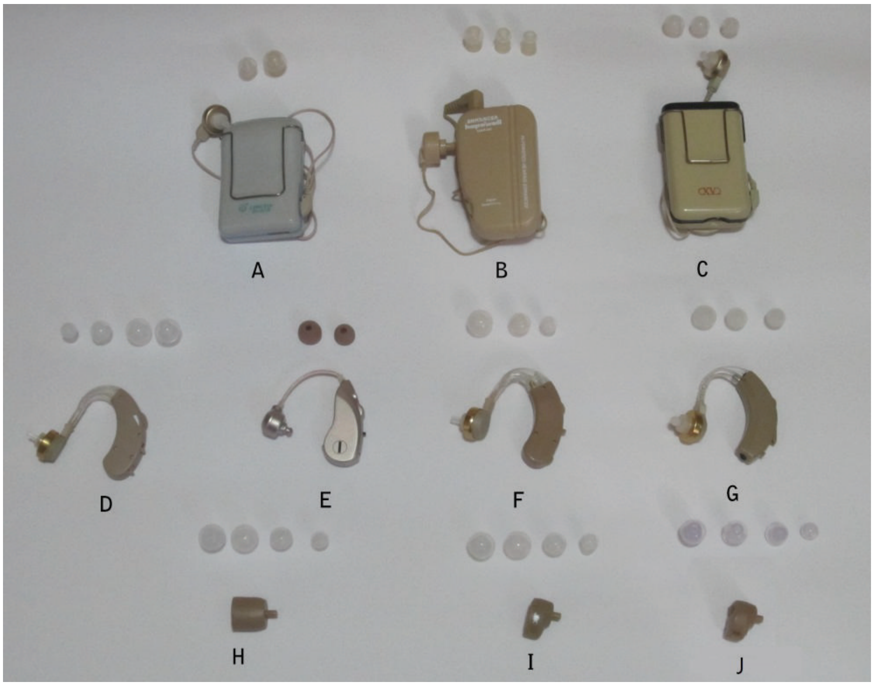


Figure 1: Over-the-counter hearing aids used in the present study. Top row: A: LingYin HA 611B; B: Hopewell HAP-40; C: Axwa EX-12D. Middle row: D: JNC-MHA-BTE130; E: UP-6411; F: ShengDe V-163; G: Axwa OM-188. Bottom row: H: Powertone HAP-F883; I: JNC-MHA-ITE 110; J: Axon K-80.

2. Methods

Ten low-cost OTC hearing aids were investigated in present study. Selection was based on their wide availability to consumers in rehabilitation aid stores, electrical appliance shops and department stores in Hong Kong. The study had three aspects: (1) Measurement of the electroacoustic performance of the OTC hearing aids. Both 2-cc coupler measurement and simulated real-ear measurement were conducted to examine the performance of the OTC hearing aids; (2) Estimation of the hearing loss that could be appropriately fitted with the OTC hearing aids using four prescriptive formulae. The four prescriptive formulas used in present study were the National Acoustic Laboratories-Revised (NAL-R), prescription of gain and output (POGO), Libby one-third gain (Libby 1/3) and Desired Sensation Level 4.0 (DSL 4.0); and (3) Determination of whether the OTC hearing aids could be appropriately fit to adults with typical sensorineural hearing loss caused by presbycusis, using the NAL-R prescriptive formula. The NAL-R formula was selected because this formula is a relatively widely used method among linear prescriptive formulae [1]. It was used in both Hong Kong [2] and North American [1] research. Moreover, the NAL-R formula was designed for fitting amplification to people with mild to moderately severe hearing loss, so that it is appropriate for fitting people with typical presbycusis hearing loss configurations [19]. For comparison purposes, the methods used generally followed those of previous studies [1, 2], but an updated American National Standards Institute (ANSI) hearing aid specification standard, ANSI S3.22-2009, was used.

2.1. Equipment

2.1.1. Over-the-Counter Hearing Aids. The ten OTC hearing aids investigated in the present study were (A) LingYin HA 611B; (B) Hopewell HAP-40; (C) Axwa EX-12D; (D) JNC-MHA-BTE130; (E) UP-6411; (F) ShengDe V-163; (G) Axwa OM-188; (H) Powertone HAP-F883; (I) JNC-MHA-ITE 110; and (J) Axon K-80. All of them were low-cost OTC hearing aids costing less than US$115 each. Moreover, they were not investigated in a previous OTC hearing aid study in Hong Kong [2]. The ten OTC hearing aids and their characteristics are shown in Figure 1 and Table 1, respectively. Three body-worn (BW), four behind-the-ear (BTE), and three in-the-ear (ITE) OTC hearing aids were included in the present study. Two of the hearing aids, Ling Yin HA 611B and Axwa EX-12D, have two and three tone controls, respectively. Therefore, there were thirteen testing conditions in the present study.

Each OTC hearing aid was provided with, on average, three stock ear domes of different sizes by the manufacturer. The outer diameter of the domes ranged from 0.7 to 1.3 cm, with a dome with outer diameter of 1.0 cm being the most commonly available size. To standardize the measurements, domes with outer diameter of 1.0 cm or the nearest available size were used in the electroacoustic measurements.

2.1.2. Measurement Equipment. All of the electroacoustic measurements were conducted with a Fonix 7000 Hearing Aid Test System in a Fonix 7020 sound chamber (Frye Electronics, Tigard, OR). An HA-1 coupler was used in the

Table 1: Summary of the characteristics of the ten OTC hearing aids.

	Models	Style	Cost ($US)	Country of manufacturer	Volume range	Special features	Operation manual	Technical specification	Battery
A	LingYin HA 611B	BW	41	China	1–5	2 tone controls (N; H)	Yes (Chinese)	Yes	AAA
B	Hopewell HAP-40	BW	22	Unknown	No marking	—	Yes (English and Spanish)	No	AAA
C	Axwa EX-12D	BW	49	China	1–8	3 tone controls (N; H; L)	Yes (English)	Yes	AA
D	JNC-MHA-BTE130	BTE	47	Korea	1–3	—	Yes (Chinese and English)	Yes	675
E	UP-6411	BTE	52	Japan[#]	1–6	—	Yes (English)	Yes	675
F	ShengDe V-163	BTE	51	China	1–4	—	Yes (Chinese)	Yes	675
G	Axwa OM-188	BTE	55	China	1–4	—	Yes (English)	Yes	675
H	Powertone HAP-F883	ITE	114	Unknown	1–5	—	Yes (English)	Yes	13
I	JNC-MHA-ITE 110	ITE	47	Korea	No marking	—	Yes (Chinese and English)	Yes	312
J	Axon K-80	ITE	37	Unknown	No marking	—	Yes (English)	Yes	312

Note. #: no information of manufacturer is printed on packaging, but the salesperson claimed that it was a Japanese brand; BW: body-worn; BTE: behind-the-ear; ITE: in-the-ear; N: normal; H: high; L: low; technical specification: manufacturer's information on electroacoustic characteristics of the hearing aid.

Table 2: Estimated hearing thresholds of elderly people based on Stenklev and Laukli data [23].

	Frequency (Hz)					
	250	500	1000	2000	3000	4000
Estimated hearing threshold (dB HL)	23.4	23.6	27.1	38.1	51.8	55.8

Note. These values show mean hearing thresholds of elderly people aged 60 or above, including both male and female and left and right ears.

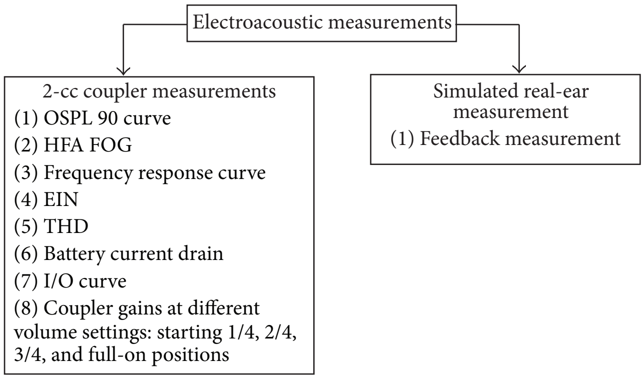


Figure 2: Electroacoustic measurements conducted on OTC hearing aids. OSPL 90: maximum sound pressure level output; HFA FOG; high frequency average full-on gain; EIN: equivalent input noise; THD: total harmonic distortion; I/O: input/output.

2-cc coupler measurement, as an HA-1 coupler is recommended to measure hearing aids with attached molds [1, 20]. In a simulated real-ear measurement component of the study, a Knowles Electronic Manikin for Acoustic Research (KEMAR; Knowles, Elk Grove, IL) was used to simulate the real-ear condition. KEMAR measures provide similar acoustic characteristics to measurement on a real person because the manikin can provide pinna, head, and torso effects; also it can simulate the impedance characteristics of the real ear, which changes with frequency [21]. Using KEMAR in measurements also avoids potentially loud intensity sound exposure to real listeners when an OTC hearing aid volume control is turned to a high output level or when feedback occurs.

2.2. Procedure

2.2.1. Electroacoustic Measurements. Both 2-cc coupler and simulated real-ear measurements were performed. Figure 2 shows the full range of measures conducted.

In the 2-cc coupler measurements, the OTC hearing aids were tested according to the ANSI S3.22-2009 hearing aid specification standard [22]. Leveling of the test equipment with the equivalent substitution method was carried out before all measurements [22]. The dome portion of the hearing aid was attached to an HA-1 2-cc coupler for measuring the electroacoustic performance of hearing aid. Output sound pressure level-90 (OSPL-90) curves, high-frequency average full-on gain (HFA FOG), frequency response curves, equivalent input noise (EIN), total harmonic distortion (THD), battery current drain, and input-output curves (I/O curve) were measured with the OTC hearing aids. The battery life for each hearing aid was estimated based on the battery current drain and the capacity of battery used [1].

In addition, the output sound pressure level and the gain at different volume settings were measured. The volume control wheels of the hearing aids were divided into 4 equal portions and the gains were measured at the starting, 1/4, 1/2, 3/4, and the full-on positions [2]. Composite noise of 50 dB SPL was used as test signal to avoid saturation of the OTC hearing aids [2].

In the simulated real-ear feedback measurements, the hearing aids were tested on KEMAR's right pinna. The loudspeaker was located at an azimuth angle of 45° and 30 cm from KEMAR [20]. The center of the loudspeaker was at the same level as the midpoint of the KEMAR pinna. To simulate a normal conversational situation, the input signal used was digital speech at 60 dB SPL [2]. The volume control was rotated until feedback was detected by the normal hearing first author who stood at 25 cm behind KEMAR or when abnormal peaks began to appear in the frequency response during testing [2]. The volume settings that the OTC hearing aids could achieve before audible or visible feedback occurred were measured.

2.2.2. Appropriate Hearing Loss for OTC Hearing Aids. The hypothetical hearing losses that could be appropriately prescribed with the OTC hearing aids were estimated. These estimates were based on the 2-cc coupler gain at the full-on position for each device and were derived using four prescription formulae [2]. The reserve gains recommended by the four selected fitting formulae were allowed for [19].

2.2.3. Fitting OTC Hearing Aids for Presbycusis Using NAL-R Formula. The average hearing thresholds of elderly people were estimated based on the study by Stenklev and Laukli [23], who surveyed the hearing levels of elderly people aged 60 or above in Norway. The ratio of male to female participants was 1.1 : 1, and average hearing thresholds were estimated by averaging the mean pure-tone hearing thresholds of both ears and both genders. The estimated hearing thresholds are showed in Table 2. These data were chosen because (1) the sample size was reasonably large, having included 232 subjects; (2) the data were relatively up-to-date when compared with the data used in previous work [2]; and (3) the data were collected in a sound-attenuating room meeting international standards.

Target 2-cc coupler full-on gains were generated by the NAL-R prescription formula based on the estimated average hearing thresholds of an elderly person shown in Table 2. The calculated target 2-cc coupler full-on gains were compared with the measured 2-cc coupler full-on gains for the OTC hearing aids to determine whether the amplification

Table 3: Summary of the results of OTC hearing aids: 2-cc coupler measurements.

OTC	OSPL 90		HFA FOG (dB)	Frequency range (Hz)	EIN (dB)	THD (%)			Battery life (hours)
	Peak frequency (Hz)	Peak SPL (dB SPL)				500 Hz	800 Hz	1.6 kHz	
A tone N	1600	127.6	44.0	375–4000	28.4	1.9	N/A	0.4	DNT
A tone L	1600	128.5	39.7	667–4667	28.5	N/A	N/A	0.7	DNT
B	700	129.8	29.0	354–>8000	35.2	0.3	0.1	0.7	DNT
C tone N	1400	126.6	52.8	<200–3667	26.4	2.3	N/A	0.1	DNT
C tone H	1400	126.3	52.5	<200–3667	25.5	2.4	N/A	0.3	DNT
C tone L	1400	126.1	41.8	396–3667	29.7	3.3	3.1	0.1	DNT
D	1400	129.3	37.2	<200–3667	30.6	2.1	1.2	0.5	142
E	1700	118.8	19.1	<200–4667	38.1	4.2	N/A	0.1	182
F	1400	125.9	32.6	<200–3833	24.9	1.4	N/A	0.3	233
G	1600	126.8	30.6	<200–5333	33.2	2.7	N/A	0.2	235
H	800	124.4	14.2	<200–4333	45.6	6.6	1.2	10.1	307
I	2000	113.1	20.9	<200–5000	31.4	23.5	46.5	10.8	154
J	700	118.4	7.6	<200–4667	52.9	4.8	0.8	4.6	212

Note. Peak SPL: peak sound pressure level; N/A: not applicable. According to the 12 dB rule, THD does not need to be measured at that frequency when its second harmonic was amplified 12 dB more than the first harmonic in the frequency response curve (Frye, 2010 [20]).
DNT: did not test. Measurement of battery current drain was not conducted since no battery substitution pills for AA and AAA battery size were available.

characteristics of the OTC hearing aids could appropriately fit people with presbycusis [1]. In the present study, the tolerances for matching the prescriptive targets were set to ±5 dB at 250 Hz, 500 Hz, 1 kHz, and 2 kHz and ±8 dB at 3 kHz and 4 kHz [24]. If the OTC hearing aid matched the prescriptive target for four frequencies or more, that hearing aid was judged to satisfactorily meet the amplification needs of elderly people with a typical hearing loss associated with presbycusis [1]. If the OTC hearing aid failed to match the prescription target under the above criterion, a higher tolerance of ±10 dB at all frequencies was used to determine whether they could meet this less strict criterion [25].

3. Results

3.1. Electroacoustic Measurements

3.1.1. 2-cc Coupler Measurements. 2-cc coupler measurement results for the ten OTC hearing aids are summarized in Table 3. The maximum OSPL 90 was 125 to 130 dB SPL for most of the OTC hearing aids, except for OTC E, I, and J which had their maximum OSPL 90 at approximately 115 dB SPL. Most of the OTC hearing aids showed their peak response at around 1400 Hz to 2000 Hz, while OTC B, H, and J had their maximum responses at 700 Hz to 800 Hz.

All OTC hearing aid frequency response curves showed high frequency limits up to 4000 Hz, except for OTC C (all tones), D, and F. The shape of the frequency response curve for OTC B differed from that usually found in hearing aids. The OSPL 90 curves and frequency response curves of three OTC hearing aids (OTC C tone N, OTC J, and OTC B) are shown in Figure 3 and represent the group of OTC hearing aids with a peak response at mid frequencies, the group with a peak response at low frequencies, and the device with an irregular frequency response, respectively.

All of the OTC hearing aids investigated were linear hearing aids and most of them showed peak clipping at high input levels. Figure 4 displays the I/O curve of OTC G, which shows this typical peak clipping effect. The output level of the OTC devices was generally limited to around 110 dB to 120 dB SPL. However, some of the low-gain hearing aids were not saturated even at a 90 dB input level, such as OTC E, H, I, and J. Figure 5 shows the I/O curve of OTC E and it can be noted that no saturation occurred.

Volume control characteristics are shown in Table 4. The volume range between different volume settings was measured and the percentage of total gain at different volume settings was calculated.

3.1.2. Simulated Real-Ear Feedback Measurement. In the feedback measurement, none of the OTC hearing aids exhibited feedback problems even at a full-on volume position.

3.1.3. Statistical Analysis. The peak frequency and peak sound pressure level at OSPL 90, the HFA FOG, and EIN in the present study were compared with those parameters in Cheng and McPherson's study [2] using an independent t-test analysis. In these four parameters, the input signals used in measuring HFA FOG and EIN are different between ANSI 3.22-1987 and ANSI 3.22-2009 standards. Nevertheless, the gain measured with either 60 dB in ANSI 3.22-1987 or 50 dB in ANSI 3.22-2009 is the same for linear hearing aids. Therefore, comparison can be made between the two studies. The results are presented in Table 5 and reveal no significant difference ($p > 0.05$) in these parameters between the present study and the previous Hong Kong OTC hearing aid study.

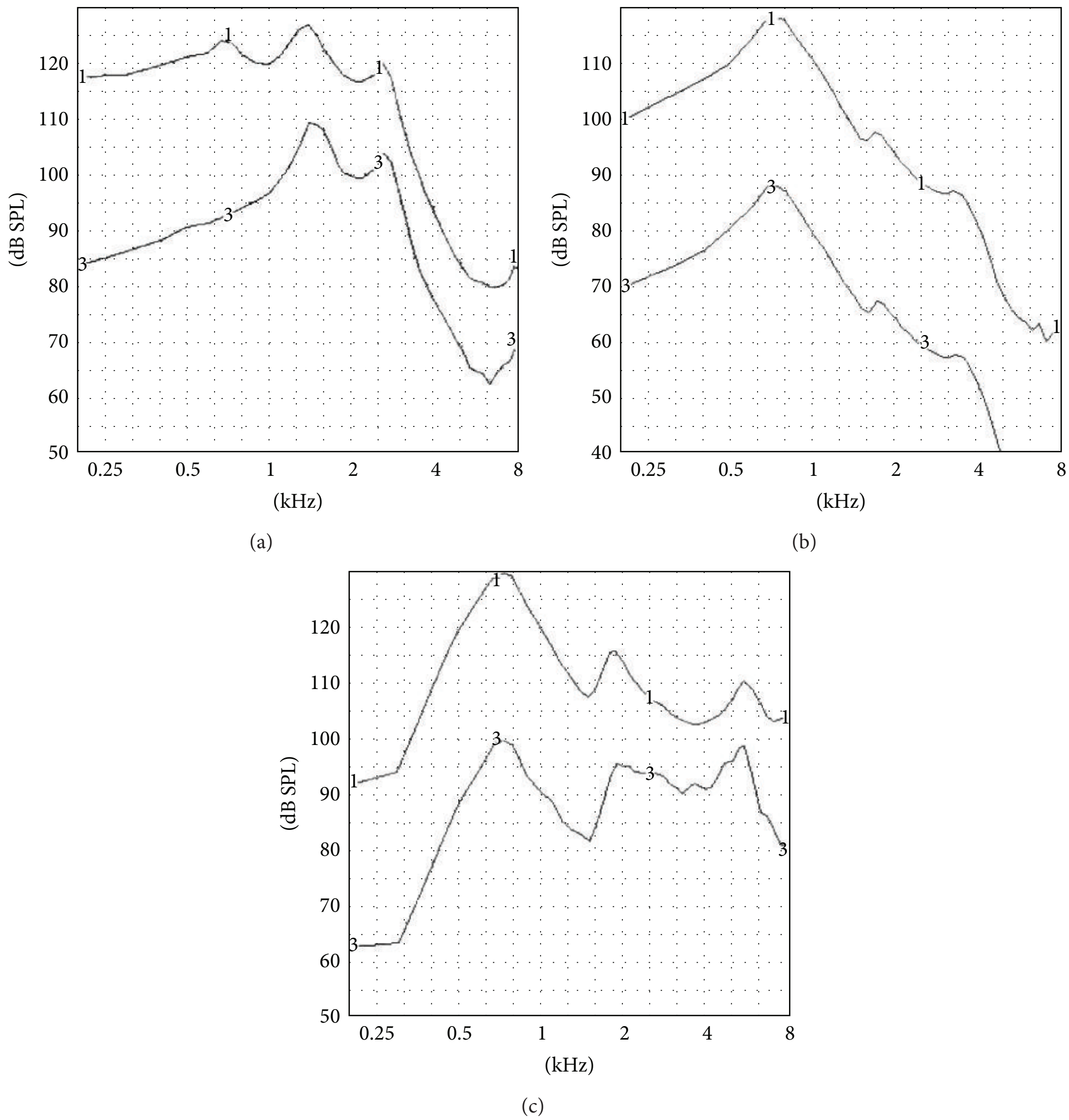


Figure 3: OSPL 90 curves (1) and frequency response curves (3) of OTC hearing aids: (a) C tone N; (b) J; (c) B.

3.2. Appropriate Hearing Loss for OTC Hearing Aids. The hypothetical hearing losses estimated by the four prescription formulae when using the 2-cc coupler full-on gain data were plotted for each OTC hearing aid and are shown in Figure 7. Figure 6, showing the mean Stenklev and Laukli audiogram for presbycusis [23], is displayed for comparison purposes. Any negative values for derived hearing thresholds were assumed to equate to normal hearing with 0 dB HL threshold [2]. The types of hearing loss that could be appropriately fit with the OTC hearing aids were generally divided into four categories: (1) sloping hearing loss up to 3000 Hz (OTC A tones N and L); (2) reverse sloping hearing loss (OTC B, C all tones, H, and J); (3) flat loss up to 3000 Hz (OTC D, F, and G); and (4) normal or minimal hearing loss (OTC E and I).

3.3. Fitting OTC Hearing Aids for Presbycusis Using NAL-R Formula. Table 6 summarizes findings for the stricter prescription matching criterion and Table 7 for the looser criterion. OTC A (tone L), F, and G were the only hearing aids that could meet the amplification needs of elderly people with presbycusis if the stricter criterion was used. If the looser criterion for matching targets was used, OTC B also matched the amplification needs for presbycusis.

4. Discussion

4.1. Electroacoustic Measurement Findings

4.1.1. OSPL 90 Curves and Frequency Response Curves. Most of the OTC hearing aids showed their peak response at the mid frequencies, while OTC B, H, and J had their maximum response at the low frequencies. Excessive amplification in the low frequencies will increase the adverse effects of background noise and additionally increase the possibility of upward spread of masking by low-frequency speech components [21]. Hence, speech intelligibility will be reduced with this pattern of amplification.

Table 4: OTC hearing aids: volume range and percentage of total gain at different volume control settings.

OTC	Volume range between different volume settings (dB SPL)				Percentage of total gain at different volume settings (%)			
	Starting to full-on	1/4 to full-on	2/4 to full-on	3/4 to full-on	1/4	2/4	3/4	Full-on
A tone N	30	6	1	0	81	97	100	100
A tone L	27	6	1	0	78	97	100	100
B	36	25	18	3	29	49	90	100
C tone N	36	20	15	2	44	59	94	100
C tone H	36	22	15	3	40	57	93	100
C tone L	35	20	14	3	44	61	92	100
D	40	19	7	2	53	82	95	100
E	19	12	4	2	40	81	90	100
F	36	9	5	2	75	87	95	100
G	33	14	6	1	58	82	93	100
H	30	18	5	1	41	83	95	100
I	22	15	6	2	31	74	90	100
J	27	12	5	2	54	80	94	100

Table 5: Statistical analysis of the OSPL90, HFA FOG, and EIN data between present study and Cheng and McPherson [2] study.

Parameter	Study	Number of testing conditions	Means	p value
OSPL 90-peak frequency	Present study	13	1361.5 Hz	0.357
	Cheng and McPherson	16	1193.8 Hz	
OSPL 90-peak sound pressure level	Present study	13	124.7 dB SPL	0.653
	Cheng and McPherson	16	125.9 dB SPL	
HFA FOG	Present study	13	32.5 dB	0.422
	Cheng and McPherson	16	36.1 dB	
EIN	Present study	13	33.1 dB SPL	0.124
	Cheng and McPherson	16	28.4 dB SPL	

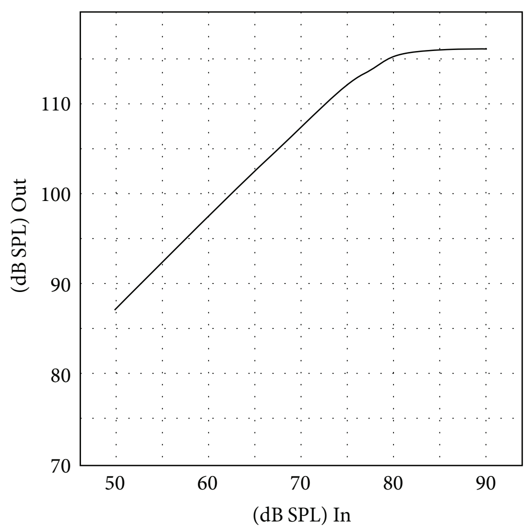


Figure 4: I/O curve of OTC hearing aid G.

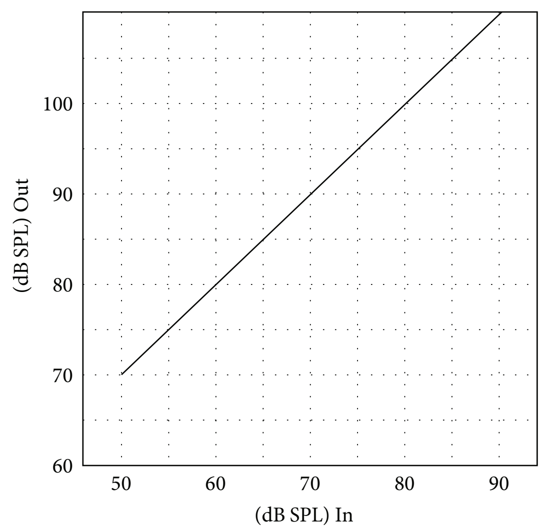


Figure 5: I/O curve of OTC hearing aid E.

Despite most OTC hearing aids having a frequency range up to 4000 Hz, there was little or no usable gain at or above 4000 Hz. Some frequency responses dropped abruptly at about 4000 Hz, such as OTC A (tone N and tone L), OTC E, and OTC I. In the present study, all the BTE and ITE hearing aids examined could not provide adequate high frequency

Table 6: Judgment of matching prescriptive targets for presbycusis with stricter criterion.

OTC	Matching the prescriptive targets for presbycusis?						Match the targets at four or more frequencies?
	0.25 Hz	0.5 Hz	1 kHz	2 kHz	3 kHz	4 kHz	
A tone N	✗	✗	✗	✗	✗	√	✗
A tone L	✗	√	√	✗	√	√	√
B	✗	✗	√	√	✗	√	✗
C tone N	✗	✗	✗	✗	√	✗	✗
C tone H	✗	✗	✗	✗	√	√	✗
C tone L	✗	✗	✗	√	√	✗	✗
D	✗	✗	✗	✗	√	✗	✗
E	√	✗	✗	✗	✗	✗	✗
F	✗	√	√	√	√	✗	√
G	√	√	√	√	√	✗	√
H	✗	√	√	✗	✗	✗	✗
I	✗	✗	✗	√	✗	✗	✗
J	√	√	✗	✗	✗	✗	✗

Note. √: OTC hearing aid matched prescriptive target.

Table 7: Judgment of matching prescriptive targets for presbycusis with looser criterion.

OTC	Matching the prescriptive targets for presbycusis?						Match the targets at four or more frequencies?
	0.25 Hz	0.5 Hz	1 kHz	2 kHz	3 kHz	4 kHz	
A tone N	√	✗	✗	✗	√	√	✗
A tone L	√	√	√	✗	√	√	√
B	√	✗	√	√	√	√	√
C tone N	✗	✗	✗	√	√	√	✗
C tone H	✗	✗	✗	✗	√	√	✗
C tone L	✗	✗	✗	√	√	✗	✗
D	✗	✗	√	√	√	✗	✗
E	√	√	✗	√	✗	✗	✗
F	√	√	√	√	√	✗	√
G	√	√	√	√	√	✗	√
H	√	√	√	✗	✗	✗	✗
I	√	√	✗	√	✗	✗	✗
J	√	√	√	✗	✗	✗	✗

Note. √: OTC hearing aid matched the prescriptive target.

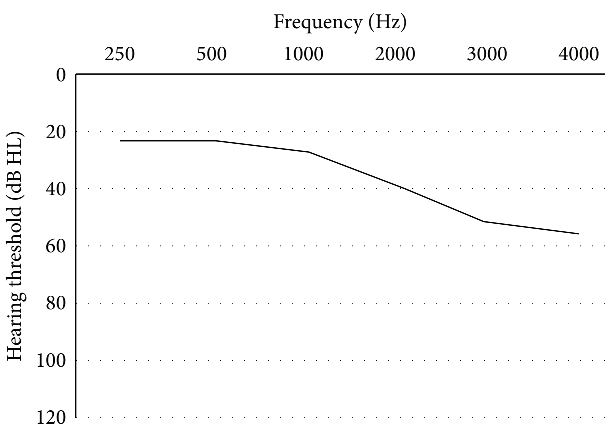


Figure 6: Estimated mean audiogram for presbycusis based on Stenklev and Laukli's data [23].

amplification. BW was the only style that could provide 15 dB or more gain at 4000 Hz.

Narrow peaks with approximately 10 dB amplitude were observed in the high frequency region of the frequency responses for OTC A (tone N and tone L), OTC D, OTC E, and OTC F. According to Dillon and Macrae [26] and van Buuren et al. [27], narrow peaks with 6 dB amplitude or more in the frequency response evoke a negative response from hearing aid users. Although OTC B did not show narrow peaks in the frequency response, its frequency response was quite irregular with several broad peaks. The smoothness of a hearing aid frequency response has been found to have a positive relationship with speech intelligibility and sound quality [27–29].

4.1.2. High-Frequency Average Full-on Gain (HFA FOG). The HFA FOG was generally higher in the BW styles and lower

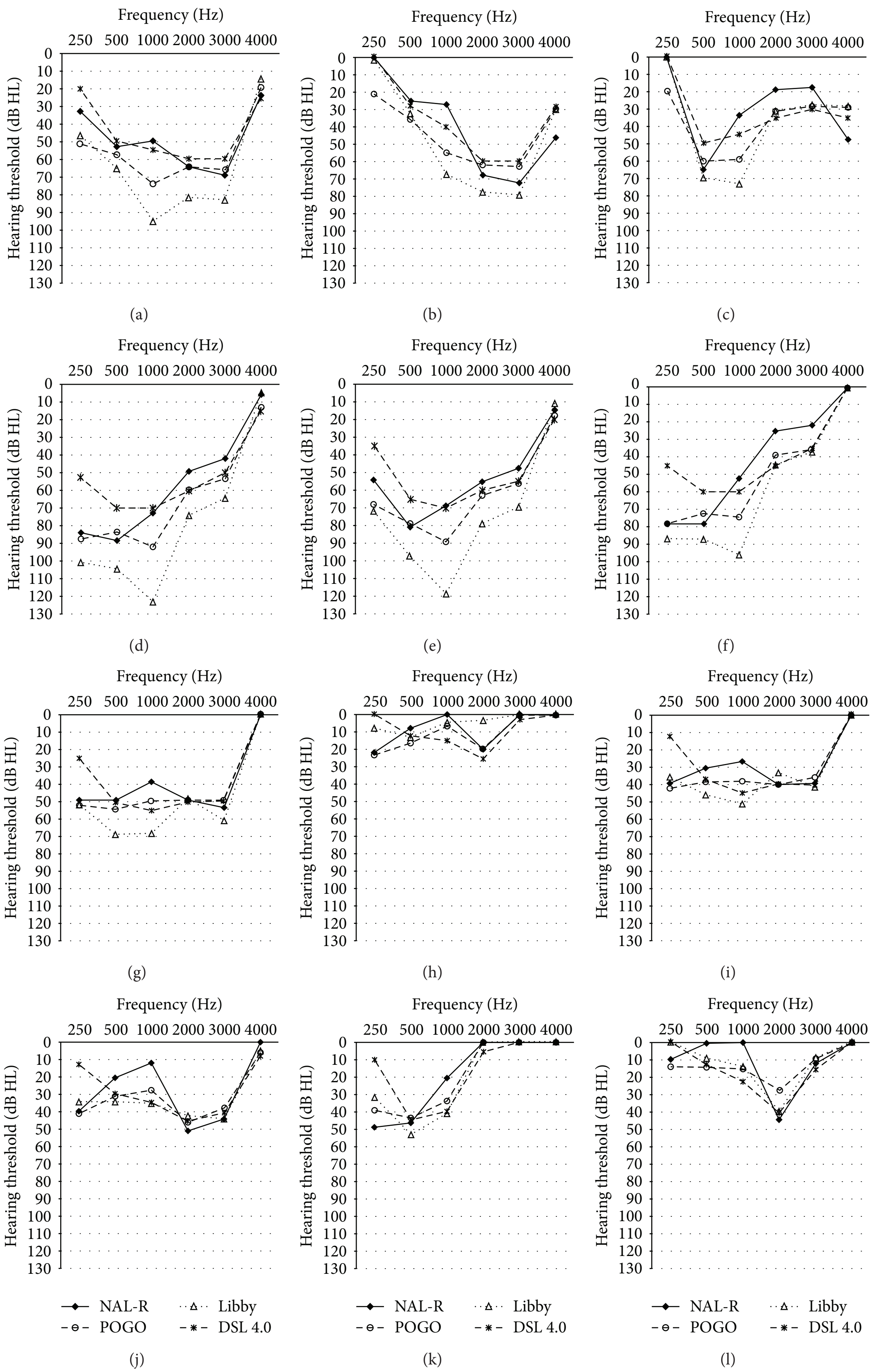
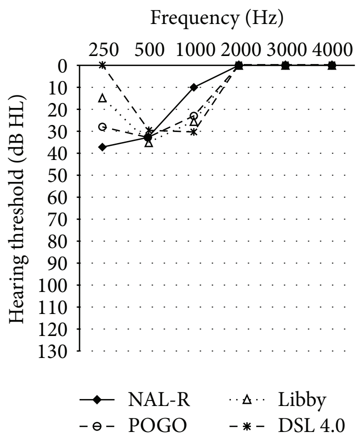


Figure 7: Continued.



(m)

Figure 7: (a) Appropriate hearing loss for OTC A (tone N). (b) Appropriate hearing loss for OTC A (tone L). (c) Appropriate hearing loss for OTC B. (d) Appropriate hearing loss for OTC C (tone N). (e) Appropriate hearing loss for OTC C (tone H). (f) Appropriate hearing loss for OTC C (tone L). (g) Appropriate hearing loss for OTC D. (h) Appropriate hearing loss for OTC E. (i) Appropriate hearing loss for OTC F. (j) Appropriate hearing loss for OTC G. (k) Appropriate hearing loss for OTC H. (l) Appropriate hearing loss for OTC I. (m) Appropriate hearing loss for OTC J.

in the ITE styles. The mean HFA FOG was 43.3 dB in BW, 29.9 dB in BTE, and 14.2 dB in ITE hearing aids. HFA FOG is the average FOG at 1000 Hz, 1600 Hz, and 2500 Hz and is used in the ANSI standard because these frequencies are very important for speech intelligibility and because most hearing aids generate significant amount of gain at these frequencies [22]. However, more than half of the OTC hearing aids exhibited a characteristic of "special-purpose hearing aids" in that their peak FOGs at any frequency were 15 dB higher than the FOG at any of the HFA frequencies [22]. For example, OTC J had a very low HFA FOG of 7.6 dB, but its peak FOG at 700 Hz was 29.3 dB. Thus, OTC J was not a low-gain hearing aid, but it amplified particular low frequencies rather than the typical HFA frequencies. Some of the OTC hearing aids may be designed for specific purposes, but these special purposes were not stated in the packaging or user manuals. This is similar to the results for OTC hearing aids studied by Callaway and Punch [1], who found that all of the low-cost OTC hearing aids they tested could be classified as special-purpose hearing aids as defined by ANSI S3.22-2009.

4.1.3. Input-Output Characteristics. All of the OTC hearing aids investigated were linear hearing aids and most of them showed peak clipping at high input levels. As mentioned before, linear hearing aids are often not suitable for people with sensorineural hearing loss [16]. Moreover, a peak clipping system will introduce more distortion when compared with a compression limiting system [30]. Peak clipping will degrade both the sound quality and speech intelligibility of loud inputs to a greater extent than a compression limiting system [30, 31].

4.1.4. Equivalent Input Noise. In the present results, the EIN ranged from 24.9 to 52.9 dB SPL, with a mean value of 33.1 dB SPL. The extremely high EINs found in OTC J and OTC H were probably measurement artifacts due to their low HAF gain [16]. According to ANSI S3.22-1987 standards [1], EIN should be limited at 28 dB SPL or less. Only three of the hearing aid measurement conditions were within the EIN limit set by ANSI 3.22-1987, which were OTC C (tones N and H) and OTC F. As a result, some OTC hearing aids will generate internal noise that is perceptible to users and may be high enough to elicit user rejection [32].

4.1.5. Total Harmonic Distortion. THD values below 5% and not more than 10% were recommended by Dillon and Macrae [26]. Most of the OTC hearing aids examined therefore had acceptable THD levels, except for OTC I. The manufacturer of OTC I specified the THD level to be 10% or lower without stating the tested frequencies. However, the THD levels at 500 Hz, 800 Hz, and 1600 Hz were all higher than 10% and THD level was 46.5% at 800 Hz. THD levels should not exceed manufacturer specification by more than 3%; otherwise, the hearing aid may have malfunctioned [22]. Moreover, another hearing aid (OTC G) had an intermittent response and its volume control was hard to rotate. Quality control is a potential problem for OTC hearing aid purchasers since salespeople may not have sufficient knowledge to check hearing aid function prior to sale.

4.1.6. Battery Life. The estimated battery life of all the OTC hearing aids was acceptable, ranging from 142 hours to 307 hours. This was probably because they used a relatively larger size of battery than conventional hearing aids. All BTE hearing aids investigated in the present study used a 675 zinc air battery, which is rarely used in modern, conventional BTE hearing aids except for high power BTE hearing aids. Therefore, the BTE style OTC hearing aids were larger in size when compared to modern conventional BTE hearing aids of similar gain.

4.1.7. Volume Control Characteristics. The mean volume range of the OTC hearing aids was 31.3 dB, indicating that

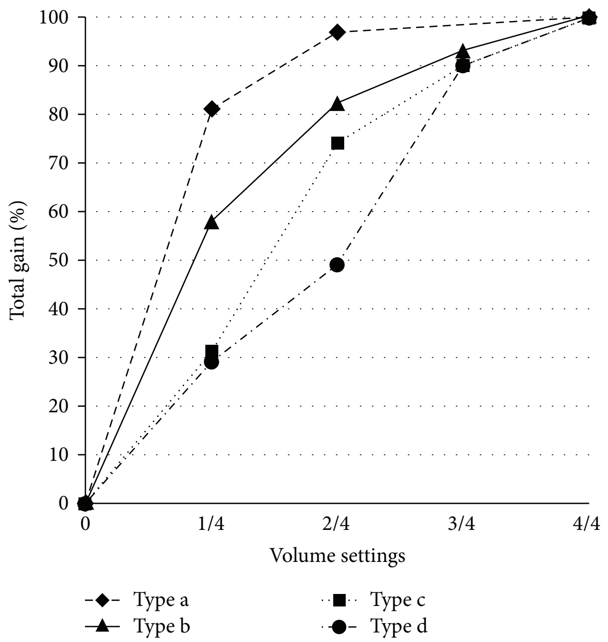


Figure 8: OTC hearing aid volume control gain characteristics.

there was about 31 dB flexibility in volume adjustment. However, the volume control (VC) was often not very effective as gain did not increase proportionally with volume control wheel adjustment. Half of the OTC hearing aids attained more than 50% of total gain at 1/4 volume setting. The VC characteristics could be divided into four types: (a) OTC A and F attained at least 75% of total gain at 1/4 volume setting; (b) OTC D, G, and J attained approximately 50% of total gain at 1/4 volume setting and the gain increased rapidly to at least 80% of total gain at 2/4 volume setting; (c) OTC E, H, and I attained less than 50% of total gain at 1/4 volume settings, but the gain increased rapidly to at least 70% of total gain at 2/4 volume setting; (d) OTC B and C attained approximately 50% of total gain at 2/4 volume setting. All OTC hearing aids reached nearly maximum gain at 3/4 volume settings. The percentages of total gain at different volume settings were plotted in Figure 8 for the four types of VC; one representative was chosen from each VC type.

Type d was the most effective VC and type a was the least effective among the four types of VC. OTC A had the poorest VC among all OTC hearing aids. Although OTC A (tone N) had a volume range of 30 dB between the 0/4 to 4/4 volume settings, there was only 6 dB difference between 1/4 and 4/4 volume settings. Consequently, it could be difficult for a hearing aid user to adjust the volume to an optimum listening level as rotation of the volume control will result in either too much or too little change in gain.

In addition, hearing aids are not typically designed to operate at the full-on position in usual situations. Reserve gain should be allowed for hearing aid users, so that they have the freedom to adjust for different listening environments. Most prescription formulae suggest a reserve gain of 10 to 15 dB [19]. It would be optimal if the hearing aid can provide appropriate gain when it is operated at the mid VC setting. However, only OTC B and C were able to provide a reserve gain of 10 to 15 dB at the mid VC setting.

4.1.8. Acoustic Feedback. In the simulated real-ear measurement, all of the OTC hearing aids could turn to full-on gain without feedback occurring. This was probably because the BTE hearing aids and ITE hearing aids were low-gain hearing aids with little output at high frequencies. Moreover, the long feedback pathway in the BW style may reduce the likelihood of feedback [2].

4.1.9. Comparison of the OTC Hearing Aids in the Present Study with Those in Previous Studies. In the present study, the performances of the OTC hearing aids were generally similar to those investigated by Cheng and McPherson [2] and the low-cost OTC hearing aids investigated by Callaway and Punch [1]. All were linear hearing aids and most of them had little usable high frequency gain. Problematic peaks were observed in the high frequency region of the frequency response curve in some of the hearing aids. EIN was still a concern for many OTC hearing aids. THD and battery life were generally acceptable in the present study and results were comparable with previous studies. Statistical analysis on the peak frequency and peak sound pressure level of OSPL 90, the HFA FOG, and EIN revealed no significant differences between the OTC hearing aids in the present study and those in the Cheng and McPherson [2] study.

4.2. Appropriate Hearing Loss for OTC Hearing Aids. The appropriate hearing loss for each OTC hearing aid was estimated using the four prescription formulae and illustrated in Figure 7. Only OTC A (tone L) and OTC B were able to provide enough amplification for a mild hearing loss at 4000 Hz; all other OTC hearing aids were only suitable for people with normal hearing thresholds at 4000 Hz. In the present study, approximately half of the hearing aid test conditions (OTC B, OTC C tone N, tone H, and tone L, OTC H, and OTC J) were suitable for people with reverse sloping hearing loss. Similarly, the majority of the OTC hearing aids investigated by Cheng and McPherson [2] revealed this phenomenon. Individuals with Meniere's disease or early stage otosclerosis may be potential clients for these OTC hearing aids [33]. The hearing loss estimated for OTC A showed a sloping configuration, especially when operated at tone L, and thus OTC A was the most appropriate device to fit presbycusis. The hypothetical appropriate hearing loss estimated from the gain of the three BTE hearing aids (OTC D, F, and G) revealed a flat hearing loss up to 3000 Hz. Therefore, these BTE hearing aids could improve the loudness of speech for people with mild to moderate flat hearing loss, but they might not provide sufficient improvement in speech intelligibility since they did not provide adequate gain at 4000 Hz. OTC I was suitable for people with mild to moderate hearing loss at 2000 Hz, but this type of hearing loss is comparatively rare. OTC E was suitable for people with relatively normal hearing sensitivity only due to its low gain.

Although the variety of appropriate hearing losses for OTC hearing aids was greater in present study than in that by Cheng and McPherson [2], it may be the result of random product selection in the two studies. Some of the OTC hearing aids in the present study were suitable for people with sloping or flat hearing loss. Nevertheless, reverse sloping hearing loss was still the most common audiological configuration appropriate for the OTC hearing aids. However, the major users of OTC hearing aids are elderly people who have high frequency sloping hearing loss [14, 18]. These individuals typically experience hearing difficulties in noisy environments since they have reduced frequency resolution ability [21]. If the elderly person with presbycusis wears a low-frequency emphasis OTC hearing aid, their speech understanding in noise will further deteriorate due to upward spread of masking caused by amplified background noise and amplified low-frequency speech components [2].

There was a limitation in the estimation of hypothetical hearing loss in the present study since the prescription formulae included reserve gain of 10 to 15 dB and assumed the hearing aids were operated at mid volume settings [21]. However, most of the OTC hearing aids did not have enough reserve gain when operated at the mid volume settings. Therefore, Figure 7 reflects the hearing loss that can be fit appropriately with the hearing aid when the hearing aid was operated at the volume setting which is 10 to 15 dB lower than the full-on gain. Some of the devices were therefore operating at 0/4 to 1/4 volume settings.

4.3. Fitting OTC Hearing Aids for Presbycusis Using NAL-R Formula. OTC A (tone L), F, and G were the only hearing aids that could meet the amplification needs of elderly people with presbycusis if a strict fitting outcome criterion was used. If a looser criterion of matching targets was used, OTC B also could be considered to match the amplification needs for presbycusis. Generally, BW styles could match the prescriptive targets at the high frequencies, but they provided too much low-frequency gain at the same time; thus they usually failed to match the targets at low frequencies; in contrast, BTE and ITE styles could match the fitting targets at low frequencies, but they usually failed to provide sufficient amplification in the high frequencies.

Although OTC A (tone L), F, and G could fit a hypothetical presbycusis using the NAL-R formula, they provided 15 dB reserve gain only when they were operated at 0/4 to 1/4 VC settings. These three hearing aids had type a and type b VC characteristics, such that the gains increase abruptly between 0/4 and 1/4 VC settings. This attribute would make it difficult for wearers to adjust the gain to an appropriate level, especially for elderly people with poor manual dexterity. Overamplification or underamplification would be the likely result due to the poor VC range in these hearing aids [2]. OTC B had relatively more effective VC, but its frequency response curve was quite irregular which may affect sound quality and speech intelligibility. Moreover, OTC B failed to meet the prescriptive target at 500 Hz due to overamplification at that frequency. In summary, some of the OTC hearing aids could match the prescriptive targets of presbycusis, but, on the other hand, their poor VC performance and inappropriate frequency response made them unsuitable for elderly people with typical presbycusis audiometric configuration. Poor benefit, increased background noise, poor sound quality, and the need to adjust the volume control are some of the major reasons that hearing aid users do not use their hearing aids [34]. Negative experience with an OTC hearing aid that has these characteristics may keep wearers from further amplification device purchase and users may label all "hearing aids" as useless. Consumers also require information on how to successfully wear and adapt to a hearing instrument. In the present study, a Chinese operation manual version was not available with half of the OTC hearing aids, making it difficult for many elderly people in Hong Kong to learn correct fitting procedures.

5. Conclusion

The present findings indicate that the electroacoustic characteristics of the selected, current generation OTC hearing aids are similar to those in previous studies over the past decades. There is no major improvement shown in the performance of OTC hearing aids over the years. All were linear hearing aids with less than optimal volume controls. Most of them showed unacceptable electroacoustic performance, such as sharp peaks in the high frequency region of frequency response, low HFA gain, poor amplification in high frequencies, and/or high EIN. However, THD levels, battery life, and feedback were generally not problematic in the OTC hearing aids.

Reverse sloping hearing loss was still the most prevalent type of hearing loss that could be appropriately fit with the OTC hearing aids. This type of low-frequency emphasis hearing aid is not suitable for most people with presbycusis, who typically show high frequency sloping hearing loss. The prescriptive targets for presbycusis were generated using the NAL-R formula. Four of the hearing aid test conditions could match the prescriptive targets at four or more frequencies, and they were judged to meet the amplification needs of elderly people with presbycusis. However, their electroacoustic performance, such as ineffective volume control, high EIN, and irregular frequency response, may override their benefit in matching the amplification needs of clients with presbycusis. For example, the appropriate hearing loss derived from OTC A was a sloping hearing loss; thus OTC A was potentially suitable for elderly individuals with hearing loss associated with presbycusis. Nevertheless, OTC A had inadequate volume control parameters and attained approximately 80% of total gain at 1/4 volume setting, making it difficult for elderly wearers to adjust gain to an appropriate level.

In summary, the low-cost OTC hearing aids investigated in the present study were not considered suitable for elderly people with presbycusis, who are likely to be the major users of OTC hearing aids. The inadequate performance of such OTC hearing aids may cause wearers to decline to adopt hearing aid use. Manufacturers should consider ways to improve VC effectiveness, lower EIN, and smooth the frequency response of OTC hearing aids. Moreover, they may consider increasing the gain at high frequencies and reducing the gain at low frequencies, in line with prescription formulae guidelines. Future OTC hearing aids then may be more suitable for their major client group—elderly people with presbycusis. On the other hand, most conventional hearing aids have advanced technology, such as directional microphones, noise-reduction algorithms, and automatic volume controls, which can improve the listening experience of wearers. In addition, the hearing aid prescribed by an audiological professional can appropriately fit the user because the electroacoustic parameters are specifically adjusted according to their individual hearing loss. Although conventional hearing aids are more expensive than OTC hearing aids, the benefits brought by conventional hearing aids may far outweigh their cost. To clarify this point further research should be done, using subjective rating procedures, on the fitting experience of current OTC hearing aid users [18]. Qualitative opinions are important because objective measurements cannot fully reflect the actual performance of a device on users [35].

Abbreviations

OTC:	Over-the-counter
EIN:	Equivalent input noise
THD:	Total harmonic distortion
NAL-R:	National Acoustic Laboratories-Revised
POGO:	Prescription of gain and output
DSL 4.0:	Desired Sensation Level version 4.0
ANSI:	American National Standards Institute
BW:	Body-worn
BTE:	Behind-the-ear
ITE:	In-the-ear
KEMAR:	Knowles Electronic Manikin for Acoustic Research
OSPL-90:	Output sound pressure level-90 dB input
HFA FOG:	High frequency average full-on gain
I/O:	Input/output
VC:	Volume control.

Conflict of Interests

There are no competing interests. The authors have no financial relationship with any organization or company mentioned in the paper.

Authors' Contribution

Bradley McPherson developed the initial idea for the study. Zoe Yee Ting Chan and Bradley McPherson designed the study. Zoe Yee Ting Chan performed the data collection and drafted the initial paper. Both authors read and approved the final paper.

Acknowledgments

The authors wish to thank Kit T. Y. Chan and Otto K. C. Fung for their assistance with the statistical analysis. The support of the Faculty Research Fund, Faculty of Education, the University of Hong Kong, is gratefully acknowledged.

References

[1] S. L. Callaway and J. L. Punch, "An electroacoustic analysis of over-the-counter hearing aids," *American Journal of Audiology*, vol. 17, no. 1, pp. 14–24, 2008.

[2] C. M. Cheng and B. McPherson, "Over-the-counter hearing aids: electroacoustic characteristics and possible target client groups," *Audiology*, vol. 39, no. 2, pp. 110–116, 2000.

[3] S. Boswell, "FDA rejects citizen petitions for over-the-counter hearing aids," *ASHA Leader*, vol. 9, no. 8, article 1, 2004.

[4] United States Food and Drug Administration: How to get hearing aids, http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/HomeHealthandConsumer/ConsumerProducts/HearingAids/ucm181479.htm.

[5] C. M. Cheng, *Over-the-counter hearing aids: electroacoustic characteristics and possible target client groups [M.S. thesis]*, Department of Speech and Hearing Sciences, University of Hong Kong, Hong Kong, 1998.

[6] Hong Kong Social Welfare Department: Social security, http://www.swd.gov.hk/en/index/site_pubsvc/page_socsecu/sub_socialsecurity/#SSAla.

[7] P. W. Y. Wong and B. McPherson, "Reasons for non-adoption of a hearing aid among elderly Chinese," *Asian Journal of Gerontology & Geriatrics*, vol. 5, pp. 62–68, 2010.

[8] Hospital Authority: Waiting time for new case booking at Ear, Nose, Throat specialist out-patient clinic, http://www.ha.org.hk/visitor/ha_visitor_index.asp?Parent_ID=10042&Content_ID=10053&Ver=HTML.

[9] L. P. Emerson, A. Job, and V. Abraham, "A model for provision of ENT health care service at primary and secondary hospital level in a developing country," *BioMed Research International*, vol. 2013, Article ID 562643, 5 pages, 2013.

[10] World Health Organization, *Hearing Loss in Persons 65 Years and Older Based on WHO Global Estimates on Prevalence of Hearing Loss*, World Health Organization, Geneva, Switzerland, 2012, http://www.who.int/pbd/deafness/news/GE_65years.pdf.

[11] Hear Talk Foundation and The Chinese University of Hong Kong, "Data from the 'Ear & Hearing Assessment Project for the Elderly' suggest '37.1% of the elderly in Hong Kong suffer from moderate to profound hearing loss,'" http://www.ihcr.cuhk.edu.hk/eng/events/pdf/pr_Eng22Nov04.pdf.

[12] K. Demeester, A. van Wieringen, J.-J. Hendrickx et al., "Audiometric shape and presbycusis," *International Journal of Audiology*, vol. 48, no. 4, pp. 222–232, 2009.

[13] S. Gordon-Salant, "Hearing loss and aging: new research findings and clinical implications," *Journal of Rehabilitation Research and Development*, vol. 42, supplement 2, pp. 9–24, 2005.

[14] S. Hannula, R. Bloigu, K. Majamaa, M. Sorri, and E. Mki-Torkko, "Audiogram configurations among older adults: prevalence and relation to self-reported hearing problems," *International Journal of Audiology*, vol. 50, no. 11, pp. 793–801, 2011.

[15] A. L. Pittman and P. G. Stelmachowicz, "Hearing loss in children and adults: audiometric configuration, asymmetry, and progression," *Ear and Hearing*, vol. 24, no. 3, pp. 198–205, 2003.

[16] M. Valente, Ed., *Hearing Aids: Standards, Options, and Limitations*, Thieme, New York, NY, USA, 2002.

[17] J. H. Macrae and H. Dillon, "An equivalent input noise level criterion for hearing aids," *Journal of Rehabilitation Research and Development*, vol. 33, no. 4, pp. 355–362, 1996.

[18] B. McPherson and E. T. L. Wong, "Effectiveness of an affordable hearing aid with elderly persons," *Disability and Rehabilitation*, vol. 27, no. 11, pp. 601–609, 2005.

[19] H. G. Mueller, D. B. Hawkins, and J. L. Northern, Eds., *Probe Microphone Measurements: Hearing Aid Selection and Assessment*, Singular, San Diego, Calif, USA, 1992.

[20] Frye Electronics, *Fonix 7000 Hearing Aid Test System Operator's Manual Version 1.70*, Frye Electronics, Tigard, Ore, USA, 2010.

[21] H. Dillon, *Hearing Aids*, Boomerang Press, Sydney, Australia, 2nd edition, 2012.

[22] Frye Electronics, *Fonix ANSI'09 Workbook*, Frye Electronics, 2013, http://www.frye.com/wp/wp-content/uploads/2013/08/ANSI09_workbook.pdf.

[23] N. C. Stenklev and E. Laukli, "Presbyacusis—hearing thresholds and the ISO 7029," *International Journal of Audiology*, vol. 43, no. 5, pp. 295–306, 2004.

[24] British Society of Audiolog and British Academy of Audiology, *Guidance on the Use of Real Ear Measurement to Verify the Fitting of Digital Signal Processing Hearing Aids*, British Society of Audiology, British Academy of Audiology, 2007, http://www.thebsa.org.uk/docs/RecPro/REM.pdf.

[25] H. Aazh, B. C. J. Moore, and D. Prasher, "The accuracy of matching target insertion gains with open-fit hearing aids," *American Journal of Audiology*, vol. 21, no. 2, pp. 175–180, 2012.

[26] H. Dillon and J. Macrae, "Derivation of design specifications for hearing aids," National Acoustics Laboratory Report 102, National Acoustics Laboratory, Sydney, Australia, 1984.

[27] R. A. van Buuren, J. M. Festen, and T. Houtgast, "Peaks in the frequency response of hearing aids: evaluation of the effects on speech intelligibility and sound quality," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 2, pp. 239–250, 1996.

[28] L. A. Davis and S. A. Davidson, "Preference for and performance with damped and undamped hearing aids by listeners with sensorineural hearing loss," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 3, pp. 483–493, 1996.

[29] J. Jerger and J. Thelin, "Effects of electroacoustic characteristics of hearing aids on speech understanding," *Bulletin of Prosthetics Research*, vol. 9, pp. 159–197, 1968.

[30] T. R. Crain and D. J. Van Tasell, "Effect of peak clipping on speech recognition threshold," *Ear and Hearing*, vol. 15, no. 6, pp. 443–453, 1994.

[31] L. Kozma-Spytek, J. M. Kates, and S. G. Revoile, "Quality ratings for frequency-shaped peak-clipped speech: results for listeners with hearing loss," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 6, pp. 1115–1123, 1996.

[32] J. Agnew, "Audible circuit noise in hearing aid amplifiers," *Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 2793–2799, 1997.

[33] J. Shanks and J. Shohet, "Tympanometry in clinical practice," in *Handbook of Clinical Audiology*, Katz J, Medwetsky L, Burkard R, and Hood L, Eds., pp. 157–188, Lippincott Williams & Wilkins, Philadelphia, Pa, USA, 6th edition, 2009.

[34] S. Kochkin and V. MarkeTrak, "'Why my hearing aids are in the drawer': the consumers' perspective," *The Hearing Journal*, vol. 53, no. 2, pp. 34–41, 2000.

[35] L. L. N. Wong, L. Hickson, and B. McPherson, "Hearing aid satisfaction: what does research from the past 20 years say?" *Trends in Amplification*, vol. 7, no. 4, pp. 117–161, 2003.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 820654, 6 pages
http://dx.doi.org/10.1155/2015/820654

Clinical Study

Treatment of Hemorrhagic Vocal Polyps by Pulsed Dye Laser-Assisted Laryngomicrosurgery

Hyung Kwon Byeon,[1,2] Ji Hyuk Han,[1,2] Byeong Il Choi,[1,2] Hye Jin Hwang,[1,2] Ji-Hoon Kim,[1,2] and Hong-Shik Choi[1,2]

[1] *Department of Otorhinolaryngology, Yonsei University College of Medicine, Gangnam Severance Hospital, 211 Eonju-ro, Gangnam-gu, Seoul 135-720, Republic of Korea*
[2] *Institute of Logopedics and Phoniatrics, Yonsei University College of Medicine, Seoul, Republic of Korea*

Correspondence should be addressed to Hong-Shik Choi; hschoi@yuhs.ac

Received 8 January 2015; Revised 19 March 2015; Accepted 19 March 2015

Academic Editor: Adam Klein

Objective. Conventional surgical techniques of laryngomicrosurgery (LMS) on hemorrhagic vocal polyps are often difficult due to obscuration of the surgical field by inadvertent bleeding from the lesion, and there are often significant amounts of mucosal epithelium loss. Here, we introduce our surgical technique using pulsed dye laser (PDL), which can effectively resect the polyp with vocal fold mucosa preservation. *Methods*. Patients who were diagnosed with hemorrhagic vocal polyp and who were surgically managed using PDL from March 2013 to October 2014 were retrospectively reviewed. Preoperative and postoperative clinical outcomes and surgical findings were evaluated. *Results*. A total of 39 patients were treated with PDL-assisted enucleation LMS. The average age was 43.7 years (range 20–73), and there were 20 males and 19 females (17 professional voice users). In all cases, the hemorrhagic polyp was successfully enucleated after application of PDL, thereby preserving the overlying epithelium. Postoperative voice outcomes were favorable with clear preservation of the vocal fold mucosal wave. *Conclusion*. PDL-assisted enucleation LMS for the treatment of hemorrhagic vocal polyps can be a safe and effective surgical technique. It can be considered a promising treatment option for hemorrhagic vocal polyps.

1. Introduction

Hemorrhagic vocal polyps are the most commonly encountered benign lesions of the vocal fold. They are more often found in men than in women and are usually located on the anterior two-thirds of the vocal fold [1, 2]. The main causative mechanism for hemorrhagic vocal polyps is acute or chronic mechanical phonotrauma to the microvasculature of the superficial layer of lamina propria (SLP). The hyperfunctional glottal sound production causes shearing stress, leading to capillary bleeding within the SLP. The resulting malformed neovascularized mass can be easily ruptured, leading to hemorrhagic events within the vocal fold proper, which in turn hinder normal propagation of vocal fold mucosal waves and cause increased vocal effort. This creates further phonotrauma to the vocal fold, which eventually exacerbates the evolving polyp. The main involvement site is the SLP, and the overlying mucosal epithelium is typically normal; however, the latter may also be involved, having either a keratotic appearance from increased frictional force between the vocal folds or a mucosa-overabundant appearance from the volumetric expansion off the polyp within the SLP [3–5].

Conventionally, hemorrhagic vocal polyps are surgically resected either by the amputation technique using cold microinstruments or CO_2 laser or by the subepithelial microflap resection technique [6]. The amputation technique results in a significant amount of epithelial loss, especially if the polyp is broad-based, and postoperative vocal outcomes can be undesirable. Additionally, although CO_2 laser can achieve optimal hemostasis and improved precision, it frequently results in mucosal scarring [7]. The transmitted heat energy from the CO_2 laser inevitably affects both the epithelium and the SLP, which can lead to fibrosis and increased mucosal stiffness, thereby ultimately resulting in deterioration of voice quality. In the subepithelial microflap resection technique, although the overlying epithelium is raised by epinephrine-saline infusion, the operation can often be cumbersome due to a certain amount of inevitable bleeding owing to the nature of the hemorrhagic polyp. The 585 nm pulsed dye laser (PDL) is an angiolytic laser causing selective photothermolysis. The chromophore of the PDL is oxyhemoglobin, resulting in the laser inducing photocoagulation of microvascular lesions with minimal damage to the surrounding normal tissue. Based on our recent successful reports of PDL application in treating sulcus vocalis and glottic leukoplakia [8, 9], we have managed to use PDL in surgically removing hemorrhagic vocal polyps. This study aims to report our surgical technique, PDL-assisted enucleation laryngomicrosurgery (LMS), which can effectively resect the polyp with vocal fold mucosa preservation.

2. Methods

2.1. Patients. This study was conducted on 39 patients who were admitted and received PDL-assisted enucleation LMS under general anesthesia for hemorrhagic vocal polyps despite persistent conservative treatment at the Department of Otorhinolaryngology, Gangnam Severance Hospital, from March 2013 to October 2014. Inclusion criteria of the patients were as follows: (1) pathologically proven hemorrhagic vocal polyps and (2) evident hemorrhagic vocal polyps on preoperative videostrobolaryngoscopy (Laryngograph Ltd., London, UK). Exclusion criteria were as follows: (1) simple, nonhemorrhagic vocal polyps and (2) other accompanying structural vocal fold abnormalities. All patients provided written informed consent before the surgery, and the Institutional Review Board of Yonsei University College of Medicine approved this retrospective study.

2.2. Surgical Technique. All operations were performed solely by the senior author (Hong-Shik Choi). Each step of the operation is illustrated (Figure 1). After clearly exposing the lesion under suspension laryngoscopy, the 585 nm PDL (Cynosure Inc., Chelmsford, MA, USA) was applied above the polyp, though without direct contact (450 μs pulse width, 2.0 J/pulse maximum output, and 2 Hz repetition rate). The PDL was delivered via a 0.6 mm flexible fiber with a constantly delivered energy of 0.75 J/pulse, and an average of 20 pulses (range 8–49) was applied for each vocal polyp. After confirming the blanching change of the overlying epithelium, an incision was made, and, with careful dissection, the epithelium could be easily "peeled away" from the polyp lesion. Further dissection was performed with microinstruments, and eventually only the polyp was extracted, leaving the overlying epithelium unharmed. The epithelium was repositioned carefully to cover the surgical defect, and the operation was completed. Following the operation, the patients were prescribed strict voice rest for 7 to 10 days and were also counseled on vocal hygiene and behavioral vocal changes.

2.3. Clinical Outcome Assessment. For evaluation of treatment outcomes, all patients were given voice assessments and laryngeal stroboscopic examinations before and two months after the operation. Voice analysis including aerodynamic measures (Phonatory Aerodynamic System; KayPENTAX, Montvale, NJ, USA), acoustic analysis, and electroglottographic analysis (EGG) (Lx speech studio program; Laryngograph Ltd., London, UK) was conducted, and voice handicap index (VHI) was examined. Auditory perceptual judgment was carried out in a recorded sample called "autumn" [10], which was approximately two minutes in length, with the use of the GRBAS scale (G, grade; R, roughness; B, breathiness; A, asthenia; S, strain). The recorded samples were evaluated in a blinded manner by two speech pathologists. The point scale of each category was 0 to 3 points, and the mean of two values each obtained from two speech pathologists was used for the analysis. The patients were followed up for a minimum of 4 months to a maximum of 9 months until it was considered that there was no need for further followup.

The demographic, clinicopathological, and treatment characteristics were retrospectively reviewed and analyzed. The following voice parameters were evaluated: maximum phonation time (MPT), mean airflow rate (MFR), subglottic pressure (Psub), closed quotient (CQ), irregularity of frequency percentage (CFx), irregularity of amplitude percentage (CAx), average fundamental frequency (F0), noise-to-harmonic ratio (NHR), jitter percentage, and shimmer percentage. Preoperative and postoperative results of each voice parameter were statistically compared using paired t-test with SPSS 18.0 software for Windows (SPSS, Chicago, IL, USA), with statistical significance defined as a P value less than 0.05.

3. Results

The mean age of the studied population was 43.7 years (range 20–73 years), and there were 20 males and 19 females. Among these patients, 17 (43.6%) were professional voice users. The hemorrhagic vocal polyp was located on the right vocal fold in 17 patients, on the left in 19, and on both in 3 patients.

All operations were successfully accomplished without any significant intraoperative complications. There was almost no bleeding during each operation, and after removal of the polyp, there were minimal amounts of mucosal loss (Figure 1). Preservation of the mucosal wave was noted in all cases on postoperative stroboscopic examination (Figure 2). During the follow-up period of the patients, there was no recurrence of the polyp or any other laryngeal lesion development.

Changes of each voice parameter after the operation are summarized in Table 1. VHI score and GRBAS scale scores from auditory perceptual judgment showed significant improvements after the operation. Other objective measures of voice analysis showed a general improvement, but only significant differences were noted in MFR and jitter percentage.

4. Discussion

The normal vocal fold is composed of five layers: the epithelium, the three layers of the lamina propria (superficial,

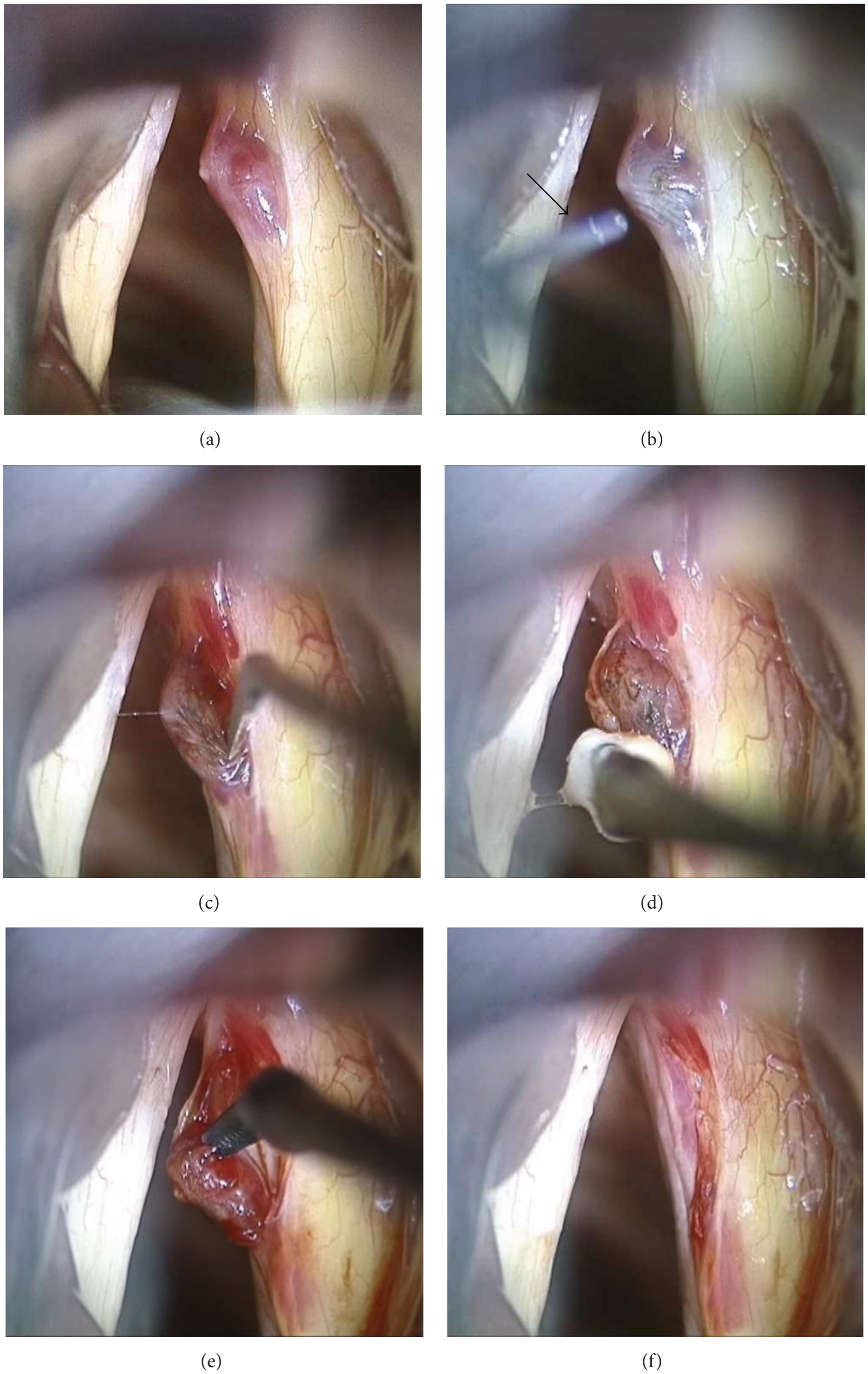
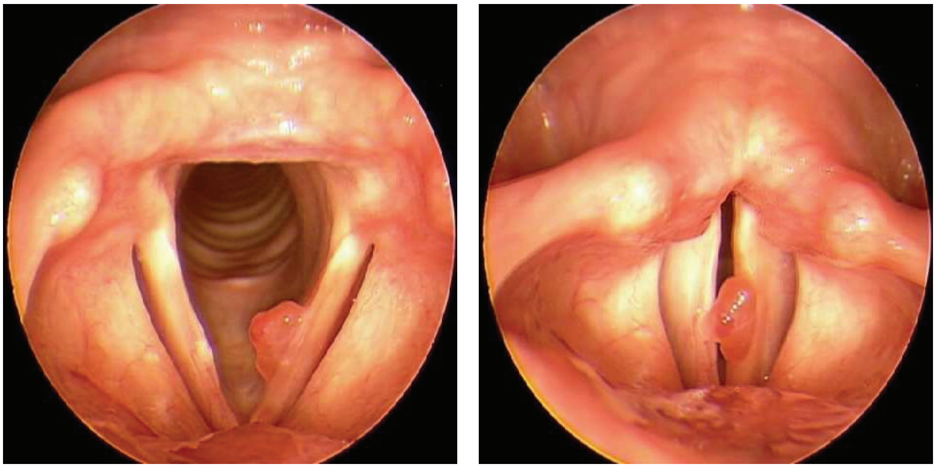
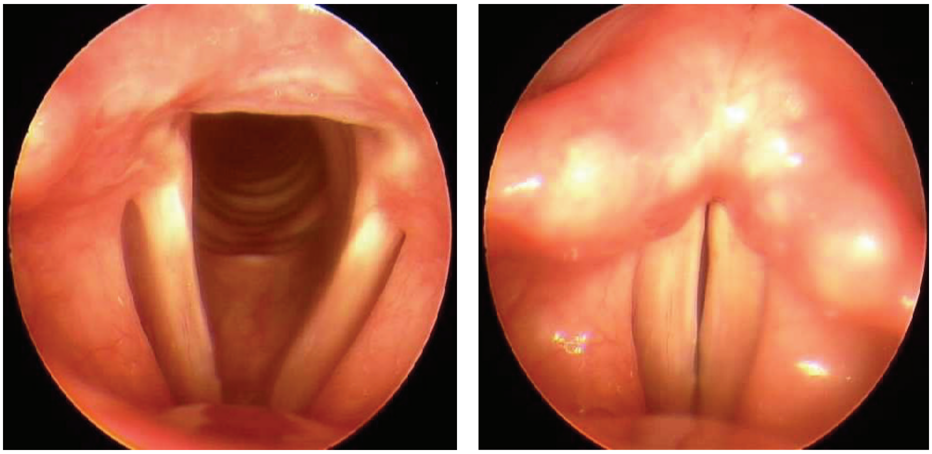


Figure 1: Intraoperative view of pulsed dye laser- (PDL-) assisted enucleation laryngomicrosurgery. (a) After suspension laryngoscopy under general anesthesia, a hemorrhagic vocal polyp is noted on the right vocal fold. (b) The PDL is delivered by a 0.6 mm fiber (arrow), which is held directly over the surface of the hemorrhagic polyp. The treated portion of the vocal fold can be confirmed by the blanching change of the epithelium. (c) After the PDL application, a longitudinal incision is made at the overlying epithelium. (d) The epithelium is easily peeled off from the lesion and opened using a small cotton ball mounted on microforceps. (e) After careful dissection with appropriate microinstruments, the hemorrhagic polyp is easily enucleated out using microcurved alligator forceps. (f) The remaining epithelium is repositioned after removal of the lesion.



(a)



(b)

Figure 2: (a) Preoperative laryngeal stroboscopic images. A typical hemorrhagic vocal polyp with a sessile base is noted on the left vocal fold. (b) Laryngeal stroboscopic images taken two months after the operation. Notice the preservation of the mucosal wave.

intermediate, and deep), and the thyroarytenoid muscle. According to Hirano's cover-body theory of vocal fold vibration, the epithelium and the SLP constitute the "cover," the vocalis muscle acts as the "body," and the intermediate and deep layers of the lamina propria, which make up the vocal ligament, are the "transition" layer [11]. The cover layer is considered important, as it is mainly involved in the normal mucosal wave vibration of the vocal folds. The basement membrane zone (BMZ) is a collection of extracellular matrix (ECM) that attaches and secures the overlying epithelium and the SLP underneath. The BMZ is further divided into two distinct layers, the superficial lamina lucida (LL) and the deep lamina densa (LD). The LL connects the basal epithelium by hemidesmosomes, and the LD is attached to the SLP by anchoring fibers consisting of collagen type VII. The LL and the LD are bound by anchoring filaments made of collagen type IV and fibronectin [12]. Hemorrhagic vocal polyps are known to develop from phonotrauma such as severe voice abuse or misuse, and these benign vocal fold lesions can lead to acute or persistent dysphonia or hoarseness. Its main pathogenesis is microvascular trauma within the BMZ and the SLP. Although some hemorrhagic vocal polyps may resolve spontaneously with minimal or conservative treatment [13], surgery is the standard treatment of choice for persistent polyps after voice therapy and observation.

A PDL beam with a 585 nm wavelength is selectively absorbed by hemoglobin, and the energy from the laser beam penetrates the superficial epithelium without causing damage yet results in intravascular coagulation of subepithelial microvasculature. The laser was initially adopted in dermatologic clinics to treat port-wine stains, telangiectasias, and any other cutaneous vascular lesions. Recently, the PDL has generated substantial interest in the field of phonosurgery for treating various laryngeal lesions. Since the initial attempt of McMillan et al. of applying PDL in treating laryngeal papillomas [14, 15], there have been numerous reports of PDL application in the management of leukoplakia, glottic dysplasia, granuloma, and vascular abnormalities of the vocal folds [16–19]. The utility of the PDL has been further verified with the treatment of glottic hyperkeratosis and sulcus vocalis by Choi et al. [8, 9].

Table 1: Voice quality improvement after the operation.

	Preoperative	Postoperative	P value
Aerodynamic measures, mean (SD)			
MPT (sec)	12.43 (5.73)	14.35 (5.50)	0.084
MFR (L/sec)	0.16 (0.13)	0.11 (0.09)	0.005*
Psub (cmH_20)	6.80 (2.08)	6.23 (2.28)	0.287
EGG analysis, mean (SD)			
CQ (%)	42.42 (5.45)	42.91 (8.06)	0.728
CFx (%)	12.50 (10.89)	9.54 (6.83)	0.197
CAx (%)	6.86 (3.41)	6.35 (4.60)	0.515
Acoustic analysis, mean (SD)			
$F0$ (Hz)	158.62 (38.14)	157.21 (41.61)	0.630
NHR	0.15 (0.03)	0.14 (0.02)	0.121
Jitter (%)	2.33 (1.58)	1.50 (0.87)	0.006*
Shimmer (%)	4.55 (2.92)	3.70 (1.61)	0.134
Patient-perceived satisfaction, mean (SD)			
VHI score	40.82 (28.62)	13.96 (13.50)	<0.001*
Auditory perceptual judgment, mean (SD)			
G	1.32 (0.54)	0.63 (0.46)	<0.001*
R	0.57 (0.40)	0.26 (0.38)	<0.001*
B	0.86 (0.48)	0.37 (0.32)	<0.001*
A	0.03 (0.19)	0 (0)	0.326
S	0.32 (0.36)	0.18 (0.24)	0.004*

IQR: interquartile range; MPT: maximum phonation time; MFR: mean airflow rate; Psub: subglottic pressure; CQ: closed quotient; CFx: irregularity of frequency; CAx: irregularity of amplitude; $F0$: average fundamental frequency; NHR: noise-to-harmonic ratio; VHI: voice handicap index; G: grade; R: roughness; B: breathiness; A: asthenia; S: strain.
* refers to the values that are statistically significant (P value < 0.05).

The photoangiolytic effect of the PDL can greatly facilitate the operation, not only from improved hemostasis, but also due to the induration effect of the lesion. From the results of this study, the hemostasis could control the natural bleeding tendency of the hemorrhagic vocal polyp, thereby avoiding any troublesome obstruction of surgical view with bleeding. Additionally, we observed that the lesion became relatively hardened due to the photocoagulation caused by the PDL, which further aided in the manipulation and dissection of the polyp. Essentially, there are three fundamental effects of laser on living tissue: photoacoustic, photothermal, and photochemical effects. According to the ultrastructural evaluation made by Ayala et al., the photoacoustic and photothermal effects of the PDL create a cleavage plane, specifically between the LL and LD of the BMZ [20]; thus, the PDL treatment would cause the mucosal epithelium to separate and elevate above the lesion within the polyp-containing vocal fold. Therefore, from this effect, not only could the epithelium be easily “peeled off” from the surface of the polyp, achieving enhanced precision of cold instrumental dissection and selective extraction of the polyp, but the all-important cover layer of the vocal fold could also be preserved with appropriate wound healing, allowing the reestablishment of normal mucosal wave vibration and phonation after the surgery. From the results of the laryngeal stroboscopic examinations (Figure 2), favorable postresection mucosal pliability and glottal closure could be verified in all cases, and voice assessment results showed general improvement of voice quality following the surgery (Table 1). There has been an earlier report by Zeitels et al. of treating the hemorrhagic vocal polyp by an angiolytic laser after raising a subepithelial microflap with saline solution infusion [19]. Our specific technique goes one step further, exploiting not only the photocoagulative effect of PDL but also the cleavage plane formation effect created by the laser. The application of PDL as an initial step of the surgery omits the necessity of subepithelial infusion.

Given that PDL is easy to use and is a safe and efficacious tool for laryngeal lesions, it has been incorporated as a minimally invasive technique for use in office-based procedures [17, 21]. The so-called “awake” PDL technique for treating vocal polyps does not require resection; however, it may lead to repeated procedures and a longer postoperative recovery period. It may also present certain limitations due to the constant movements of the larynx in conscious patients. The procedure described in the present study, however, is performed under general anesthesia and can be considered as one particular method of LMS utilizing a PDL as an effective surgical tool. It may be a useful, alternative surgical technique to the conventional surgical resection of hemorrhagic vocal polyps by subepithelial microflap elevation followed by subepithelial infusion of saline and epinephrine as initially reported by Hochman and Zeitels [6]. Although further investigation with a longer period of follow-up and a larger volume of patients may be required to establish the feasibility of the PDL-assisted enucleation LMS technique, this procedure was found to be effective in removing hemorrhagic polyps and was easy to perform. It has also shown promising postoperative results with a decreased period of healing.

5. Conclusion

The PDL-assisted enucleation LMS technique can be easily and effectively performed to treat hemorrhagic vocal polyps indicated for surgery. The mucosal wave of the vocal fold is well preserved as postoperative healing is improved, resulting in good voice outcomes. This surgical technique may also prove to be a valuable treatment for various laryngeal lesions.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] V. Kambic, Z. Radsel, M. Zargi, and M. Acko, “Vocal cord polyps: incidence, histology and pathogenesis,” *Journal of Laryngology & Otology*, vol. 95, no. 6, pp. 609–618, 1981.

[2] M. M. Johns, “Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts,” *Current Opinion in Otolaryngology and Head and Neck Surgery*, vol. 11, no. 6, pp. 456–461, 2003.

[3] O. Kleinsasser, “Pathogenesis of vocal cord polyps,” *Annals of Otology, Rhinology and Laryngology*, vol. 91, pp. 378–381, 1982.

[4] M. S. Courey, M. A. Scott, J. A. Shohet, and R. H. Ossoff, "Immunohistochemical characterization of benign laryngeal lesions," *Annals of Otology, Rhinology and Laryngology*, vol. 105, no. 7, pp. 525–531, 1996.

[5] R. H. G. Martins, J. Defaveri, M. A. C. Domingues, and R. de Albuquerque e Silva, "Vocal polyps: clinical, morphological, and immunohistochemical aspects," *Journal of Voice*, vol. 25, no. 1, pp. 98–106, 2011.

[6] I. I. Hochman and S. M. Zeitels, "Phonomicrosurgical management of vocal fold polyps: the subepithelial microflap resection technique," *Journal of Voice*, vol. 14, no. 1, pp. 112–118, 2000.

[7] S. M. Zeitels, "Laser versus cold instruments for microlaryngoscopic surgery," *Laryngoscope*, vol. 106, no. 5, pp. 545–552, 1996.

[8] C. S. Hwang, H. J. Lee, J. G. Ha et al., "Use of pulsed dye laser in the treatment of sulcus vocalis," *Otolaryngology—Head and Neck Surgery*, vol. 148, no. 5, pp. 804–809, 2013.

[9] Y. M. Park, K. H. Jo, H. J. Hong, and H. S. Choi, "Phonatory outcome of 585 nm/pulsed-dye laser in the management of glottic leukoplakia," *Auris Nasus Larynx*, vol. 41, pp. 459–463, 2014.

[10] H. Kim, "Perceptual, acoustical, and physiological tools in ataxic dysarthria management: a case report," in *Proceedings on the 2nd Conference in the Korean Society of Phonetic Sciences and Speech Technology*, vol. 2, pp. 9–22, 1996.

[11] M. Hirano and Y. Kakita, "Cover-body theory of vocal fold vibration," in *Speech Science*, R. G. Daniloff, Ed., pp. 1–46, College-Hill Press, San Diego, Calif, USA, 1985.

[12] S. D. Gray, S. S. N. Pignatari, and P. Harding, "Morphologic ultrastructure of anchoring fibers in normal vocal fold basement membrane zone," *Journal of Voice*, vol. 8, no. 1, pp. 48–52, 1994.

[13] A. M. Klein, M. Lehmann, E. R. Hapner, and M. M. Johns III, "Spontaneous resolution of hemorrhagic polyps of the true vocal fold," *Journal of Voice*, vol. 23, no. 1, pp. 132–135, 2009.

[14] K. McMillan, S. M. Shapshay, J. A. McGilligan, Z. Wang, and E. E. Rebeiz, "A 585-nanometer pulsed dye laser treatment of laryngeal papillomas: preliminary report," *Laryngoscope*, vol. 108, no. 7, pp. 968–972, 1998.

[15] T. A. Valdez, K. McMillan, and S. M. Shapshay, "A new laser treatment for vocal cord papilloma-585-nm pulsed dye," *Otolaryngology—Head and Neck Surgery*, vol. 124, no. 4, pp. 421–425, 2001.

[16] R. A. Franco Jr., S. M. Zeitels, W. A. Farinelli, W. Faquin, and R. R. Anderson, "585-nm pulsed dye laser treatment of glottal dysplasia," *Annals of Otology, Rhinology & Laryngology*, vol. 112, no. 9 I, pp. 751–758, 2003.

[17] S. M. Zeitels, J. A. Burns, R. A. Franco Jr., R. E. Hillman, S. H. Dailey, and R. R. Anderson, "Office-based treatment of glottal dysplasia and papillomatosis with the 585-nm pulsed dye laser and local anesthesia," *Annals of Otology, Rhinology and Laryngology*, vol. 113, no. 4, pp. 265–276, 2004.

[18] S. B. Clyne, J. A. Koufman, S. L. Halum, and G. N. Postma, "Pulsed dye laser treatment of laryngeal granulomas," *Annals of Otology, Rhinology and Laryngology*, vol. 114, no. 3, pp. 198–201, 2005.

[19] S. M. Zeitels, L. M. Akst, J. A. Burns, R. E. Hillman, M. S. Broadhurst, and R. R. Anderson, "Pulsed angiolytic laser treatment of ectasias and varices in singers," *Annals of Otology, Rhinology and Laryngology*, vol. 115, no. 8, pp. 571–580, 2006.

[20] C. Ayala, M. Selig, W. Faquin, and R. A. Franco Jr., "Ultrastructural evaluation of 585-nm pulsed-dye laser-treated glottal dysplasia," *Journal of Voice*, vol. 21, no. 1, pp. 119–126, 2007.

[21] H.-T. Kim and H.-J. Auo, "Office-based 585 nm pulsed dye laser treatment for vocal polyps," *Acta Oto-Laryngologica*, vol. 128, no. 9, pp. 1043–1047, 2008.

Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 579206, 7 pages
http://dx.doi.org/10.1155/2015/579206

Research Article

Are Auditory Steady-State Responses Useful to Evaluate Severe-to-Profound Hearing Loss in Children?

Signe Schuster Grasel, Edigar Rezende de Almeida, Roberto Miquelino de Oliveira Beck, Maria Valéria Schmidt Goffi-Gomez, Henrique Faria Ramos, Amanda Costa Rossi, Robinson Koji Tsuji, Ricardo Ferreira Bento, and Rubens de Brito

Department of Otolaryngology, University of São Paulo, 255 Dr. Eneas de Carvalho Aguiar Avenue, 05403-900 São Paulo, SP, Brazil

Correspondence should be addressed to Roberto Miquelino de Oliveira Beck; robertomobeck@gmail.com

Received 21 January 2015; Accepted 8 April 2015

Academic Editor: Haldun Oguz

Objective. To evaluate Auditory Steady-State Responses (ASSR) at high intensities in pediatric cochlear implant candidates and to compare the results to behavioral tests responses. *Methods*. This prospective study evaluated 42 children with suspected severe-to-profound hearing loss, aged from 3 to 72 months. All had absent ABR and OAE responses. ASSR were evoked using binaural single frequency stimuli at 110 dB HL with a 10 dB down-seeking procedure. ASSR and behavioral test results were compared. *Results*. Forty-two subjects completed both ASSR and behavioral evaluation. Eleven children (26.2%) had bilateral responses. Four (9.5%) showed unilateral responses in at least two frequencies, all confirmed by behavioral results. Overall 61 ASSR responses were obtained, most (37.7%) in 500 Hz. Mean thresholds were between 101.3 and 104.2 dB HL. Among 27 subjects with absent ASSR, fifteen had no behavioral responses. Seven subjects showed behavioral responses with absent ASSR responses. No spurious ASSR responses were observed at 100 or 110 dB HL. *Conclusion*. ASSR is a valuable tool to detect residual hearing. No false-positive ASSR results were observed among 42 children, but in seven cases with absent ASSR, the test underestimated residual hearing as compared to the behavioral responses.

1. Introduction

As universal newborn hearing screening programs are established in numerous countries, more children will be diagnosed in early childhood with some degree of hearing loss.

Early detection and intervention during the critical period for language and cognitive development can improve individual performance [1].

Children with severe-to-profound bilateral hearing loss are candidates for cochlear implantation (CI) and require specific audiologic evaluation prior to intervention. As early age of indication and presence of residual hearing are important factors for postimplant speech perception and language development, this has resulted in further decrease of minimum age of surgery [2–7].

In these very young children, behavioral audiologic evaluation can be challenging, may not be obtained in children younger than 6 months, and usually does not assess each ear separately. Thus the audiologic evaluation of pediatric cochlear implant candidates relies more and more on electrophysiological measures.

The most widely used electrophysiological procedure for estimating hearing thresholds in young children is click and tone burst auditory brainstem responses (ABR). Due to the transient nature of the stimuli used to evoke ABR, maximum output levels are 95 dB hearing level (HL). In view of that, the possibility of residual hearing at severe-to-profound levels cannot be investigated with ABR [8].

Hearing assessment of children, using the Auditory Steady-State Responses (ASSR), is made by frequency specific continuous modulated tones and allows increased levels of stimulation intensity. Therefore, ASSR can provide ear and frequency specific threshold information at elevated intensity levels up to 120 dB HL and higher, providing better and more reliable investigation of ears with minimal residual hearing [9]. Furthermore, ASSR thresholds may be used for hearing aid fitting prior to cochlear implantation.

For such reasons, ASSR is a unique tool for auditory assessment of young cochlear implant candidates.

Some authors [10, 11] have investigated the use of ASSR to evaluate patients with severe-to-profound hearing loss. They showed that spurious responses might occur during high stimulus intensities, especially in 500 and 1000 Hz.

Solutions have been implemented by the manufacturer to reduce artifacts at high-intensity stimulation [12].

Few papers have been published since 2004. One report evaluated 15 children with severe-to-profound hearing loss by ASSR, but behavioral thresholds were obtained for only one subject [13]. As cochlear implant is the first choice, especially, for the young child with severe-to-profound hearing loss, it is quite important to obtain more data in the pediatric population.

Previously, we performed two studies at the University of São Paulo. One of them evaluated adults with severe-to-profound hearing loss. The responses of pure tone audiometry (PTA) and ASSR were compared. Patients' subjective perception of ASSR stimuli was also evaluated and compared to PTA test results, and no systematic extra-auditory ASSR responses at high intensities were observed [14].

The other study evaluated children with severe-to-profound hearing loss from 6 to 65 months. Most ASSR responses (48%) were found at 500 Hz [15].

The aim of this study was to evaluate Auditory Steady-State Responses (ASSR) at high intensities in pediatric cochlear implant candidates and to compare the results to behavioral test responses.

2. Materials and Methods

2.1. Subjects. This prospective study evaluated 58 children with suspected severe-to-profound hearing loss, aged from 3 to 72 months. All children referred to this institution for pediatric cochlear implant evaluation between January and December 2011 were enrolled.

We *included* only children with normal external and middle ear conditions.

We *excluded* patients with severe neurologic disorders who did not permit behavioral evaluation. We also excluded patients who showed responses on either ABR or OAE or who did not achieve noise ratio under 30 microvolts during ASSR.

Overall 16 patients were excluded: three had severe neurologic disorders and could not complete behavioral evaluation, one showed responses on click ABR, one had bilateral absent click ABR and normal DPOAE (possible auditory neuropathy spectrum disorder), one showed high noise levels during ASSR, and ten did not show up for behavioral evaluation.

2.2. Methods. The procedure was a routine assessment for pediatric cochlear implantation at the Department of Otolaryngology, University of São Paulo School of Medicine. The study was reviewed and approved by the Hospital's Ethic Committee (number 38954) and written informed consent was obtained from all parents.

In a unique session using light general anesthesia with Sevoflurane, all children were examined by otomicroscopy and tympanometry followed by click auditory brainstem responses (click ABR), bone conduction ABR, distortion-product otoacoustic emissions (DPOAE), and ASSR.

All ABR, OAE, and ASSR recordings were obtained in a sound-treated room.

2.3. Stimuli and Recordings

2.3.1. ABR and Otoacoustic Emissions. The tests were run using the Navigator Pro SCOUT and AEP (Natus Bio-Logic Systems Corp., Mundelein, IL) software.

EEG activity was recorded using gold disk electrodes placed on the earlobe and Fpz. The contralateral earlobe was used as ground. Interelectrode impedance was less than 2 KOhm.

Click stimuli (duration: 100 milliseconds) were presented with ER3A insert earphones at the maximum level of 90 dB HL at a rate of 21.1/s with rarefaction and condensation polarity. The responses were considered to be absent when both rarefaction and condensation waves showed no responses at 90 dBHL. Bone conduction was tested with a standard bone vibrator at the maximum intensity of 55 dB HL using alternated click stimuli and contralateral masking of the same intensity.

For DPOAE we applied the diagnostic 750 to 8000 Hz test protocol (Navigator Pro SCOUT, Natus Bio-Logic Systems Corp., Mundelein, IL). Responses greater than 6 dB over background noise (signal-to-noise ratio) at five out of eight frequencies were considered as Pass criteria.

2.3.2. ASSR. The ASSR test was completed in the same session as ABR and OAE. ASSR were evoked using binaural single frequency stimuli at 110 dBHL with a 10 dB down-seeking procedure. Test stimuli were 0.5, 1, 2, and 4 kHz tones modulated in amplitude and frequency. Stimuli were 20% frequency and 100% amplitude modulated at 65 Hz for all tones in the left ear and at 69 Hz for all tones in the right ear according to the default specifications of the ASSR system (Navigator Pro MASTER, Natus Bio-Logic Systems Corp., Mundelein, IL). Modulation rates of 65 Hz or higher were used to ensure acceptable signal-to-noise ratio for response detection. Test stimuli were presented through insert earphones (ER3A) previously calibrated for each frequency as suggested by the manufacturer. Measurements were made with a Brüel & Kjaer sound level meter DB-0138 which conforms to ANSI S3.7-1995.

The ASSR assessments were performed by a dichotic single frequency technique. This implies that a single frequency was offered to both ears simultaneously. Electrode disks were fixed with electrolytic paste to the scalp at Cz (active), midline posterior neck (reference), and Fpz (ground). All electrode impedances were below 5 KOhm, and the interelectrode impedance values were kept below 3 KOhm. A maximum of 10 sweeps containing 16 epochs each were recorded per trial. Each epoch was 1.024 seconds. The electrophysiological recording was converted by means of a fast Fourier transform after each sweep. The response was accepted with an F ratio, comparing the fast Fourier components at the stimulus modulation frequencies to determine whether the difference was significantly different from the background noise ($P < 0.05$). If a sweep contained more than 80 nV of electrophysiological noise, it was rejected. Thresholds were repeated to guarantee reliable results, completing 10 sweeps in each run. The same physicians performed all tests.

2.3.3. Behavioral Evaluation. Two experienced audiologists did the behavioral evaluation in a double-walled sound booth in a free field test condition. Children were tested in one or more sessions. We chose instruments because 3-month-old infants would not be able to perform visual reinforcement audiometry.

The stimuli were uncalibrated sounds, presented at low 55–80 dB SPL, medium 70–88 SPL, and high intensities 80–115 dB SPL, depending on the instrument. Three frequency ranges were evaluated. We used drums for the low frequency spectrum (125–500 Hz), wooden rattle and *agogo* (Brazilian instrument) for mid-low frequencies (1000–3000 Hz), and metallic rattle and bell for high frequencies (over 3000 Hz) [16, 17]. A response was considered positive if the subject localized the stimulus in a lateral, superior, or inferior plane, according to Dworsack-Dodge et al. [18], in one or more frequency ranges. We considered a positive response, if the child changed the suction behavior, increased or decreased facial or body movements, or started crying or showed a startle reflex time-locked to the stimulus [18].

Children had behavioral evaluation within three months of the electrophysiological test battery.

2.3.4. Criteria for Data Evaluation. We analyzed four ASSR frequencies in each ear (500, 1000, 2000, and 4000 Hz). The percentage of present or absent responses at each frequency was determined. We considered present ASSR responses, if responses could be obtained in at least two frequencies in one or both sides.

For the behavioral test, we evaluated if the participant had present or absent responses to the frequency range of low, mid-low, and high frequencies. No ear specific responses were obtained in this test condition.

2.3.5. Comparison between Both Tests. ASSR and behavioral test results were compared, regarding present or absent responses. Therefore, we divided the subjects into four groups.

(1) Present responses in both tests.

(2) Absent responses in both tests.

(3) Present responses in ASSR and no responses in behavioral evaluation.

(4) No responses in ASSR and present responses in behavioral evaluation.

Table 1: Percentage of ASSR responses for each frequency (mean thresholds and standard deviation).

Frequency	Responses (%)	Mean/SD
500 Hz	37.7	101.3/±9.6
1000 Hz	29.5	103.3/±6.8
2000 Hz	21.3	103.0/±7.5
4000 Hz	11.4	104.2/±7.8

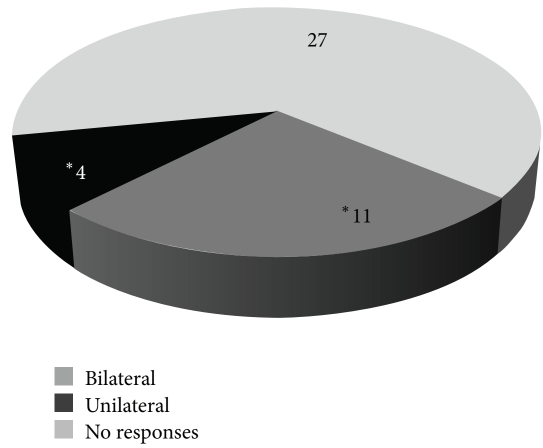


Figure 1: ASSR responses ($n = 42$ subjects). Unilateral and bilateral responses were considered only if they were positive in at least 2 frequencies.

2.4. Statistical Analysis. Association between ASSR and behavioral test results was evaluated by Kendall's rank correlation tau and Cohen's kappa coefficient. The software used was "R statistical computing."

3. Results

From 58 children enrolled in the study, forty-two completed both ASSR and behavioral evaluation.

After all we studied 42 subjects (20 girls and 22 boys) between 6 and 60 months (mean age: 29.3 months, median age: 26.0 months, and SD: 15.6 months).

Fifteen subjects (35.7%) showed ASSR responses in two or more frequencies. Eleven had bilateral and 4 had unilateral responses (Figure 1).

Most responses (37.7%) were obtained at 500 Hz, 29.5% at 1000 Hz, 21.3% at 2000 Hz, and 11.4% at 4000 Hz (Table 1).

Right and left ear responses are presented in Figure 2. Frequency specific thresholds are shown in Figure 3. Twenty-seven subjects (64.3%) had absent ASSR responses (Table 2).

At behavioral evaluation we found 34 responses (48.6%) in the low frequency, 24 (34.3%) in mid-low frequency, and 12 (17.1%) in the high frequency range. Twenty-four children had responses to more than one instrument. Seven subjects of 42 (16.6%) had absent responses to all frequency ranges.

3.1. Comparison between ASSR and Behavioral Test Results. Fifteen subjects had responses in both tests (group 1), fifteen had absent responses in both tests (group 2), no subject had

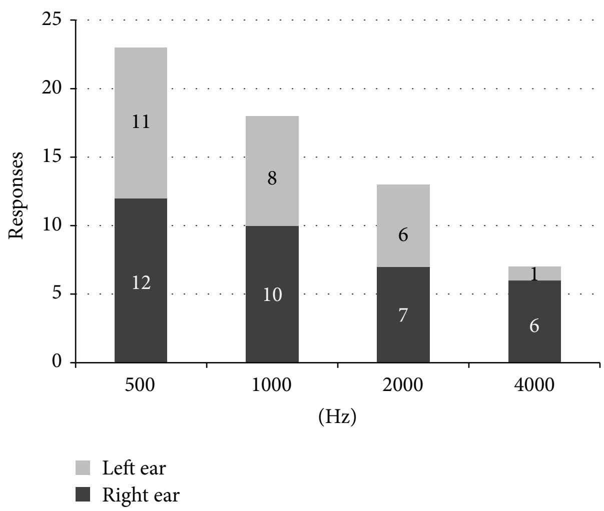


Figure 2: Right and left ear ASSR responses.

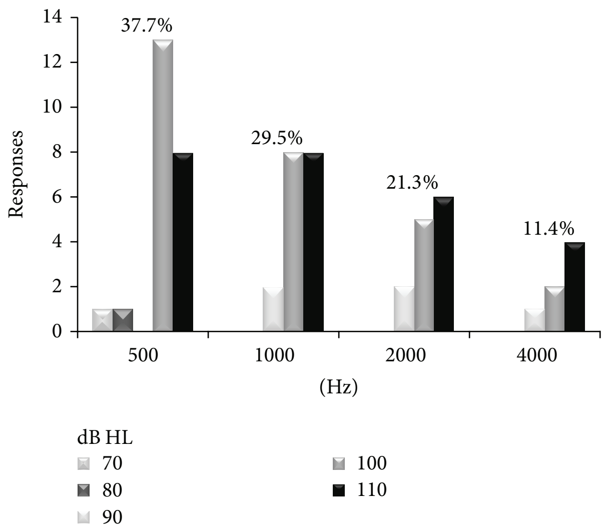


Figure 3: ASSR thresholds for each frequency (n = 61 responses).

Table 2: ASSR responses (n = 42 subjects), thresholds in dB HL.

Patients	Age (months)	500 Hz		1000 Hz		2000 Hz		4000 Hz	
		R	L	R	L	R	L	R	L
1	8	NR	100	NR	NR	NR	NR	NR	NR
2	11	NR	NR	NR	NR	100	NR	NR	NR
3	31	100	100	100	100	100	100	100	NR
4	28	NR	NR	NR	NR	NR	NR	NR	NR
5	60	NR	NR	NR	NR	NR	NR	NR	NR
6	22	NR	100	NR	NR	NR	NR	NR	NR
7	23	NR	NR	NR	NR	NR	NR	NR	NR
8	27	110	NR	NR	NR	NR	NR	NR	NR
9	47	NR	NR	NR	NR	NR	NR	NR	NR
10	24	NR	NR	NR	NR	NR	NR	NR	NR
11	48	NR	NR	NR	NR	NR	NR	NR	NR
12	53	NR	NR	NR	NR	NR	NR	NR	NR
13	29	110	NR	110	NR	110	NR	110	NR
14	36	NR	NR	NR	NR	NR	NR	NR	NR
15	34	NR	NR	NR	NR	NR	NR	NR	NR
16	23	NR	NR	NR	NR	NR	NR	NR	NR
17	30	NR	NR	NR	100	NR	NR	NR	NR
18	21	NR	NR	NR	NR	NR	NR	NR	NR
19	36	NR	NR	NR	NR	NR	NR	NR	NR
20	17	70	NR	90	NR	90	NR	100	NR
21	6	100	NR	NR	110	NR	NR	NR	NR
22	21	NR	NR	NR	100	NR	NR	NR	NR
23	12	NR	NR	100	NR	NR	NR	NR	NR
24	23	80	NR	100	NR	NR	NR	NR	NR
25	9	NR	110	110	110	NR	110	NR	110
26	25	NR	NR	NR	NR	NR	NR	NR	NR
27	53	100	100	NR	NR	NR	NR	NR	NR
28	48	100	100	90	100	100	90	NR	NR
29	12	NR	NR	NR	NR	NR	NR	NR	NR
30	13	NR	NR	NR	NR	NR	NR	NR	NR
31	36	NR	NR	NR	NR	NR	NR	NR	NR
32	57	110	110	110	NR	110	110	110	NR
33	39	NR	NR	NR	NR	NR	110	NR	NR
34	24	NR	100	NR	NR	NR	NR	NR	NR
35	60	110	100	110	110	110	100	NR	NR
36	13	100	100	NR	NR	NR	NR	NR	NR
37	34	NR	NR	NR	100	NR	NR	90	NR
38	36	NR	NR	NR	NR	NR	NR	NR	NR
39	57	NR	NR	NR	NR	NR	NR	NR	NR
40	14	NR	NR	NR	NR	NR	NR	NR	NR
41	9	NR	NR	NR	NR	NR	NR	NR	NR
42	22	110	110	110	NR	NR	NR	110	NR

R = right ear; L = left ear; NR = no response.

present ASSR and absent behavioral responses (group 3), and twelve had absent responses in ASSR and present responses at the behavioral test (group 4) (Figure 4).

In group 4, five subjects had single frequency responses on ASSR (two at 500 Hz, two at 1000 Hz, and one at 2000 Hz) but were considered nonresponders for data analysis.

Overall, in 30 subjects (71.4%), both tests showed consistent results.

4. Discussion

Accurate diagnosis of a severe-to-profound bilateral sensorineural hearing loss remains the primary and most basic requirement for implantation [19]. The ASSR may therefore assist in the decision of cochlear implant candidacy in young infants in whom specific audiologic challenges related to the limitations of the audiometric test battery are encountered [20].

The possibility of an objective evaluation at this age is very important for the correct decision of cochlear implantation and of the side to be implanted, in case of a unilateral CI. Still,

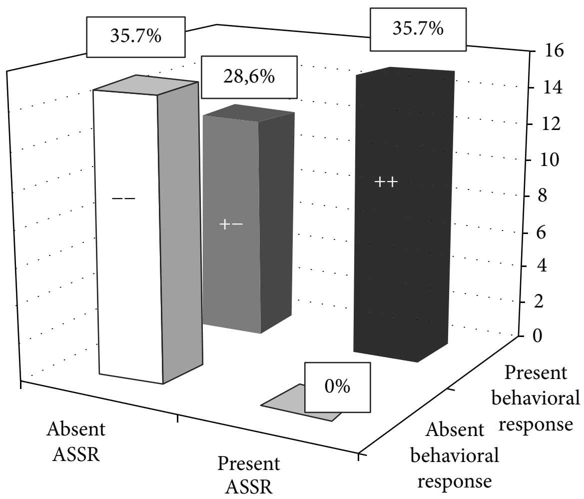


FIGURE 4: ASSR and behavioral tests correlated well in 71.4% of the subjects: 15 children had present responses in both tests and 15 had absent responses in both tests.

behavioral responses should be obtained and correlated with ASSR responses.

Our results demonstrate a strong correlation between ASSR and behavioral test results in 71.4% of the subjects. Moreover, no false-positive ASSR responses were seen in this casuistic, differing from other authors [10, 11].

In this study we excluded patients with external and middle ear disorders, because the focus was on pure severe-to-profound sensorineural hearing loss. Children with suspected auditory neuropathy spectrum disorder (ANSD) or neurologic disabilities were also excluded, as behavioral testing may be very difficult or even unreliable in some cases. Previous reports showed that ASSR thresholds in ANSD may be substantially higher than pure tone thresholds [21], so currently the value of ASSR in evaluating ANSD is still controversial.

In all frequency range both tests showed consistent results, as subjects responded best to low frequencies and less in the high frequency range. Since instruments used for the behavioral test are not frequency specific but encompass a wider frequency range, it is not surprising to observe higher number of responses in this test. This is especially true for the low frequency range, because 250 Hz was not tested during ASSR, but may have contributed to the higher number of behavioral responses in this frequency range (125 to 500 Hz).

Stimulation of the low frequency range in high intensities may evoke responses due to vibration and may not reflect the auditory status [9]. So caution is recommended when considering responses in this frequency range at high intensities.

All subjects with absent behavioral responses had no responses in ASSR. In other words, no subject had spurious or extra-auditory ASSR responses in our cohort, differing from other studies [10, 11]. Few studies have been published comparing high-intensity ASSR results to behavioral tests in the pediatric population [13, 22]. Swanepoel and Hugo [13] evaluated 15 children with severe-to-profound hearing loss by ASSR, but behavioral thresholds were obtained for only one subject. The same group [22] evaluated 10 children (between 10 and 15 years) with severe-to-profound hearing loss. They found no significant difference between ASSR and behavioral thresholds, except at 500 Hz. This small casuistic ($N = 10$) was composed of older children (mean age: 13 years and 4 months), old enough to perform pure tone audiometry. These results are not comparable to the present study with younger patients (mean age: 2 years and 5 months) who were evaluated by an instrument-based behavioral test, since visual reinforcement audiometry is not a test tool for infants aged 3 months. It seems that ASSR has been unpopular for evaluating pediatric cochlear implant candidates. A reason may be that not all equipment permits ASSR stimulation above 100 dB HL, as it is the case for newer equipment, using the Chirp stimulus for ASSR assessment. To our knowledge, there are no published data about ASSR threshold obtained with the Chirp stimulus in high-intensity levels.

Nevertheless, high-intensity ASSR is a promising tool to evaluate residual hearing in children with severe-to-profound hearing loss and absent responses to click ABR. The continuous amplitude and frequency-modulated tones make it possible to determine frequency specific thresholds at high intensities by means of an objective evaluation, with minimal interference of the examiner. Thus, it is a unique electrophysiological technique to get frequency specific thresholds at intensities exceeding 95 dB HL, where tone ABR rarely will detect responses, due to the transient nature of the tone burst stimulus. Moreover, tone ABR depends on visual wave identification by the clinician who should be well trained and cautious.

As ear specific information is available, ASSR thresholds can be used for hearing aid fitting before eventual cochlear implantation. Furthermore, in case of unilateral responses, four of 42 subjects in this study, the results may assist the surgical team to choose the side to be implanted, if unilateral CI is the option.

As in other reports [22, 23], most children exhibited bilateral ASSR responses.

The Navigator Pro MASTER software permits simultaneous stimulation of both ears at one test frequency even at high intensities (>80 dB HL), so overall test time is reduced. Although test time was not the scope of our study, the whole procedure (otomicroscopy/tympanometry and the electrophysiological test battery including click ABR, bone conduction ABR, DPOAE, and ASSR) did not exceed 60 minutes per patient. In our institution these tests are performed by the same physicians for more than four years and are part of routine evaluation before pediatric cochlear implantation, so this may have contributed to the acceptable overall test time.

We considered present ASSR responses, if they could be obtained in at least two frequencies in one or both sides to prevent spurious or extra-auditory responses at single frequencies. However, this decision turned ASSR less sensitive than the behavioral evaluation, where one instrument encompasses a wide acoustic spectrum and stimulates a range of frequencies.

Unlike Swanepoel and Hugo [13] we did not exceed the stimulus intensity of 110 dB HL in any frequency, so we may have missed some ASSR responses at very high intensities. This might have caused twelve subjects in our study to show behavioral responses with absent ASSR, as defined by our criteria. Among these subjects, five had single frequency responses, not considered auditory responses in our study, perhaps underestimating residual thresholds. But seven children had behavioral responses in more than one frequency range with no response at all in ASSR. In these cases, ASSR probably underestimated auditory thresholds. These results differ from other studies [14, 22] obtained in older children and adults. The response amplitude in children is usually smaller, so the signal-to noise ratio is poorer in the very young age group [24]. Responses may be difficult to be detected even at normal hearing levels, as response amplitude decreases toward threshold [25]. So the clinician should be cautious and not overestimate ASSR when testing young children. At least at 110 dB HL, absent ASSR may not predict absent behavioral responses in all children. We do not recommend including higher intensities (120 dB or higher) in routine evaluation of pediatric cochlear implant candidates, because these very high intensities may damage the cochlea and the clinician may be unaware of behavioral test results, as the electrophysiological tests usually precede the behavioral tests.

Therefore, a complete test battery including electrophysiological and behavioral measures is still the best means to correctly evaluate hearing thresholds, even in cases of severe-to-profound hearing loss.

5. Conclusion

ASSR is a valuable tool to detect residual hearing in young children with absent ABR and DPOAE. No false-positive ASSR results were observed among 42 subjects, but in seven cases with absent ASSR, the test underestimated residual hearing as compared to the behavioral responses.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] C. Yoshinaga-Itano, A. L. Sedey, D. K. Coulter, and A. L. Mehl, "Language of early- and later-identified children with hearing loss," *Pediatrics*, vol. 102, no. 5, pp. 1161–1171, 1998.

[2] T. A. Zwolan, S. Zimmerman-Phillips, C. J. Ashbaugh, S. J. Hieber, P. R. Kileny, and S. A. Telian, "Cochlear implantation of children with minimal open-set speech recognition skills," *Ear and Hearing*, vol. 18, no. 3, pp. 240–251, 1997.

[3] T. P. Nikolopoulos, G. M. O'Donoghue, and S. Archbold, "Age at implantation: Its importance in pediatric cochlear implantation," *Laryngoscope*, vol. 109, no. 4, pp. 595–599, 1999.

[4] L. S. Eisenberg, A. S. Martinez, G. Sennaroglu, and M. J. Osberger, "Establishing new criteria in selecting children for a cochlear implant: performance of 'platinum' hearing aid users," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 185, pp. 30–33, 2000.

[5] M. A. Novak, J. B. Firszt, L. A. Rotz, D. Hammes, R. Reeder, and M. Willis, "Cochlear implants in infants and toddlers," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 185, pp. 46–49, 2000.

[6] D. M. Hammes, M. A. Novak, L. A. Rotz, M. Willis, D. M. Edmondson, and J. F. Thomas, "Early identification and cochlear implantation: critical factors for spoken language development," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 189, pp. 74–78, 2002.

[7] K. I. Kirk, R. T. Miyamoto, C. L. Lento, E. Ying, T. O'Neill, and B. Fears, "Effects of age at implantation in young children," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 189, pp. 69–73, 2002.

[8] G. Rance, R. C. Dowell, F. W. Rickards, D. E. Beer, and G. M. Clark, "Steady-state evoked potential and behavioral hearing thresholds in a group of children with absent click-evoked auditory brain stem response," *Ear and Hearing*, vol. 19, no. 1, pp. 48–61, 1998.

[9] O. G. Lins, T. W. Picton, B. L. Boucher et al., "Frequency-specific audiometry using steady-state responses," *Ear and Hearing*, vol. 17, no. 2, pp. 81–96, 1996.

[10] S. A. Small and D. R. Stapells, "Artifactual responses when recording auditory steady-state responses," *Ear and Hearing*, vol. 25, no. 6, pp. 611–623, 2004.

[11] M. P. Gorga, S. T. Neely, B. M. Hoover, D. M. Dierking, K. L. Beauchaine, and C. Manning, "Determining the upper limits of stimulation for auditory steady-state response measurements," *Ear and Hearing*, vol. 25, no. 3, pp. 302–307, 2004.

[12] T. W. Picton and M. S. John, "Avoiding electromagnetic artifacts when recording auditory steady-state responses," *Journal of the American Academy of Audiology*, vol. 15, no. 8, pp. 541–554, 2004.

[13] D. Swanepoel and R. Hugo, "Estimations of auditory sensitivity for young cochlear implant candidates using the ASSR: preliminary results," *International Journal of Audiology*, vol. 43, no. 7, pp. 377–382, 2004.

[14] H. F. Ramos, S. S. Grasel, R. M. Beck et al., "Evaluation of residual hearing in cochlear implants candidates using auditory steady-state response," *Acta Oto-laryngologica*, vol. 135, no. 3, pp. 246–253, 2015.

[15] R. M. O. Beck, H. Ramos, S. S. Grasel, E. R. de Almeida, and R. D. B. Neto, "F079 Auditory steady-state responses (ASSR) in young cochlear implant candidates," *International Journal of Pediatric Otorhinolaryngology*, vol. 75, supplement 1, p. 97, 2011.

[16] A. Poblano, I. Chayo, J. Ibarra, and E. Rueda, "Electrophysiological and behavioral methods in early detection of hearing impairment," *Archives of Medical Research*, vol. 31, no. 1, pp. 75–80, 2000.

[17] D. D. Didoné, L. R. Kunst, T. M. Weich, A. C. Ourique, C. M. Franceschi, and T. Tochetto, "Acompanhamento do desenvolvimento da função auditiva em crianças sem e com indicadores de risco para a surdez," *Distúrbios da Comunicação*, vol. 23, no. 3, pp. 317–323, 2011.

[18] M. M. Dworsack-Dodge, J. Gravel, and A. M. Grimes, *Audiologic Guidelines for the Assessment of Hearing in Infants and Young Children*, 2012, http://audiology-web.s3.amazonaws.com/migrated/201208_AudGuideAssessHear_youth.pdf_5399751b249593.36017703.pdf.

[19] N. L. Cohen, S. B. Waltzman, and S. G. Fisher, "A prospective, randomized study of cochlear implants," *The New England Journal of Medicine*, vol. 328, no. 4, pp. 233–237, 1993.

[20] B. J. Gantz, G. G. Woodworth, J. F. Knutson, P. J. Abbas, and R. S. Tyler, "Multivariate predictors of audiological success with multichannel cochlear implants," *Annals of Otology, Rhinology and Laryngology*, vol. 102, no. 12, pp. 909–916, 1993.

[21] Z. Jafari, S. Malayeri, H. Ashayeri, and M. A. Farahani, "Adults with auditory neuropathy: comparison of auditory steady-state response and pure-tone audiometry," *Journal of the American Academy of Audiology*, vol. 20, no. 10, pp. 621–628, 2009.

[22] D. Swanepoel, R. Hugo, and R. Roode, "Auditory steady-state responses for children with severe to profound hearing loss," *Archives of Otolaryngology—Head and Neck Surgery*, vol. 130, no. 5, pp. 531–535, 2004.

[23] G. Rance and R. J. S. Briggs, "Assessment of hearing in infants with moderate to profound impairment: the Melbourne experience with auditory steady-state evoked potential testing," *The Annals of Otology, Rhinology & Laryngology. Supplement*, vol. 189, pp. 22–28, 2002.

[24] A. I. Tlumak, E. Rubinstein, and J. D. Durrant, "Meta-analysis of variables that affect accuracy of threshold estimation via measurement of the auditory steady-state response (ASSR)," *International Journal of Audiology*, vol. 46, no. 11, pp. 692–710, 2007.

[25] D. R. Stapells, J. S. Gravel, and B. A. Martin, "Thresholds for auditory brain stem responses to tones in notched noise from infants and young children with normal hearing or sensorineural hearing loss," *Ear and Hearing*, vol. 16, no. 4, pp. 361–371, 1995.