

Control of Fundamental Frequency in Dysphonic Patients During Phonation and Speech

*A. Ziethe, *S. Petermann, †U. Hoppe, *N. Greiner, *M. Brüning, *C. Bohr, and *M. Döllinger, *†Erlangen, Germany

Abstract: Purpose: The pitch-shift reflex (PSR) is the adaptation of the fundamental frequency during phonation and speech and describes the auditory feedback control. Speakers without voice and speech disorders mostly show a compensation of the pitch change in the auditory feedback and adapt their fundamental frequency to the opposite direction. Dysphonic patients often display problems with the auditory perception and control of their voice during therapy. Our study focuses on the auditory and kinesthetic control mechanisms of patients with muscle tension dysphonia (MTD) and speakers without voice and speech problems. Main purpose of the study is the analysis of the functionality of the control mechanisms within phonation and speech between patients with MTD and normal speakers.

Method: Sixty-one healthy subjects (17 male, 44 female) and 22 patients with MTD (7 male, 15 female) participated following two paradigms including a sustained phonation (vowel /a/) and speech ([ˈmama]). Within both paradigms the fundamental frequency of the auditory feedback was increased synthetically. For the analysis of the PSR the electroencephalogram, electroglottography, the voice signal, and the high-speed endoscopy data were recorded simultaneously. The PSR in the electroencephalogram was detected via the N100 and the mismatch negativity. Statistical tests were applied for the detection of the PSR in the physiological response within the electroglottography, voice, and high-speed endoscopy signals. The results were compared between both groups.

Results: No differences were found between the controls and patients with MTD regarding latency and magnitude of the perception of the pitch shift in both paradigms, but for the magnitude of the behavioral response. Differences also could be found for both groups between the “no pitch” and “pitch” condition of the two paradigms regarding vocal fold dynamics and voice quality. Patients with MTD showed more vibrational irregularities during the PSR than the controls, especially regarding the symmetry of vocal fold dynamics.

Conclusion: Patients with MTD seem to have a disturbed interaction between the auditory and kinesthetic feedback inducing the execution of an overriding behavioral response.

Key Words: Auditory feedback–Kinesthetic feedback–Pitch-shift reflex–Muscle tension dysphonia–voice quality.

INTRODUCTION

For undisturbed and accurate phonation fine-tuned movements of the involved muscles of the larynx are needed.¹ The precision of the phonation determines voice quality and leads in case of a dyscoordination by exerting hyper- or hypotension in the larynx muscles to a lower voice quality and an increased strain.¹ Patients with a muscle tension dysphonia (MTD) show an imbalance in the coordination of the larynx muscles during phonation.¹ The phonation of patients with MTD is characterized by breathiness and roughness.¹ Often hypertension can be observed together with a higher voice frequency during speech and a reduced modulation of the voice.

Human verbal communication as a social key competence requires intonation and prosody to convey verbal information.² Therefore, the ability to modulate voice frequency is important. Neural mechanisms are hypothesized to control speech production by feedback (auditory control) and feedforward mechanisms (kinesthetic/somatosensory control) involving auditory and somatosensory processes.³ Both mechanisms are interacting subsystems comparing auditory and somatosensory input with a memorized model of a sound or a word for their verification.⁴ The feedback and feedforward mechanisms control motor speech reactions in order to reach the desired target.⁴ In contrast to the feedback mechanism, the feedforward mechanism begins before the intended sound is produced due to its relation to previous learned models instead of incoming sensory information.⁴ This model of motor speech control is called DIVA model (directions into velocities of articulation).⁴

Effects of a disturbed auditory feedback were in focus in different studies.^{5–8} The physiological reaction to a perturbation of the auditory feedback is called the pitch-shift reflex (PSR).⁹ The function of the auditory feedback can be investigated with the PSR. When subjects perceive a higher or lower auditory feedback than the aimed pitch, they adapt the pitch (following response) or they compensate for the pitch to the opposite direction (opposing reaction).¹⁰ Either following or opposing reactions could be observed with

Accepted for publication July 2, 2018.

This work was supported by Deutsche Forschungsgemeinschaft, grant no. DO 1247/4-1.

Conflict of interests: The authors claim no conflict of interest.

From the *Division of Phoniatics and Pediatric Audiology, ENT Clinic, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany; and the †CICERO – Cochlear Implant Centrum at the Department of Otorhinolaryngology – Head and Neck Surgery, Universitätsklinikum Erlangen, Erlangen, Germany.

Address correspondence and reprint requests to Anke Ziethe, Division of Phoniatics and Pediatric Audiology at the Department of Otorhinolaryngology – Head and Neck Surgery, Universitätsklinikum Erlangen, Raumerstr. 1b, 91054 Erlangen, Germany. Tel.: +49 9131 85 33815; fax: +49 9131 85 39272. E-mail: anke.ziethe@uk-erlangen.de

Journal of Voice, Vol. 33, No. 6, pp. 851–859

0892-1997

© 2019 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jvoice.2018.07.001>

latencies between 100 and 150 ms post stimulus.^{5,10,11} The control mechanisms of the audio–vocal system seems to react flexible depending on the nature of the applied perturbation and sound type.^{11,12} With an increasing stimulus magnitude, varying between 25 and 300 cents, the percentage of opposing reactions decreases¹¹ and with a higher F_0 the response magnitudes were larger and faster in contrast to a lower F_0 .¹²

In previous studies we found that the auditory and kinesthetic control process adjusts pitch and presumably voice quality in normal voices during phonation.^{13,14} In general, patients with MTD often show a reduced kinesthetic control of phonation compared to normal voices and singers¹ that can contribute to a persisting voice disorder. If patients with MTD do have an insufficient working kinesthetic control, the adaptation of the voice pitch and control can be affected. In our study we investigate the mechanism of the auditory and kinesthetic feedback of patients with MTD and control speakers during phonation and speech. We hypothesize, that a PSR can be detected in both groups (normal voices and patients with MTD) and that for the patient group significant differences occur during the PSR due to an insufficient kinesthetic control of phonation during sound production and speech. For the experimental setup we will use the electroencephalogram (EEG) to determine the latency of the pitch perception, the electroglottography (EGG) and high-speed videoendoscopy (HSE) to describe vocal fold dynamics as well as acoustic voice measures [F_0 , Jitter, Shimmer, amplitude perturbation quotient (APQ), amplitude perturbation factor (APF), harmonics-to-noise ratio (HNR), signal-to-noise ratio (SNR), normalized noise energy, cepstral peak prominence (CPP), spectral flatness] for the analysis of voice quality.

METHODS

This study was approved by the local ethics committee of Universitätsklinikum Erlangen (approval number: 4364) and fulfilled all requirements, including those of the Declaration of Helsinki (2013).

Participants

Sixty-one healthy subjects (17 male, 44 female, aged between 20 and 30 years) fulfilling the inclusion criterion of being healthy with no reported voice-, hearing-, speech-, and neurological disorder and 22 patients with diagnosed MTD (7 male, 15 female, aged between 21 and 64 years) without history of neurological deficits, hearing problems, or a speech disorder (exclusion criteria) participated in this study. The patients with MTD were diagnosed by the Department of Phoniatics and Pediatric Audiology in the ENT clinic at the university hospital Erlangen and did not have had any treatment before they participated in the study. The hearing level of all participants was checked by audiometry and was about 20 dB SPL. None of the participants were trained singers.

Study design

All participants had to fulfill two paradigms, the phonation and the speech paradigm. The participants were blinded regarding the purpose of the experiment. Within the phonation paradigm the participants first heard a model female/male voice via headset (Logitech Premium Stereo Headset; USB port, 22.05 kHz) during the first 4 seconds phonating the vowel [a] with a fundamental frequency of 220 Hz or 110 Hz depending on the gender of the subject. The participants had to join in and heard their own recorded voice for 1 second until it was pitched up by 700 cents for 300 ms. After that, the pitch-shifted feedback turned into the normal, unpitched feedback for the last 700 ms. To reduce bone conducted feedback, pink noise with an upper cut-off frequency of 9000 Hz was also applied over the headset. Within the speech paradigm the participants first heard their own recorded articulation of the disyllabic word [‘mama] for also 4 seconds. Then they had to join in and hold on with the same intonation followed by the beginning of the pitched up auditory feedback. To minimize disruptions in the HSE video by articulatory movement (application of a transnasal endoscope) and to allow full analysis, a word with nasals and simple word structure (CVCV) was chosen.

Within both paradigms the feedback was played at 75 dB SPL and the participants hold 75–85 dB SPL during phonation and speech. Both paradigms were repeated 20 times. The recordings were performed in a sound attenuating and electrically shielded chamber. The complete experimental setup was first introduced and validated by Petermann et al.¹³ They applied first the HSE and showed similar sensitivity regarding the detection of the PSR. With the HSE further information can be gained concerning vocal fold dynamics and voice quality. Therefore, it was selected for the presented study.

For the multidimensional analysis of the PSR the EEG, EGG, acoustic voice signal, and the video signal of the HSE were recorded simultaneously (see 8;13;14). The N100 (phonation paradigm) and MMN (speech paradigm) of the EEG were detected as neurophysiological correlates of the PSR. The N100 is a negative peak in the waveform of the EEG occurring 100 ms after presentation of an auditory stimulus and reflects the perception of the pitch shift. The mismatch negativity (MMN) is an auditory event-related brain potential occurring as a response of a series of expected standard stimuli followed by an unexpected deviant stimulus and can be used as indicator for the perception of the PSR.

The EGG, acoustic voice signal, and video signal were acquired for monitoring the behavioral response (pitch change, voice quality).

Data acquisition

EEG

Seven electrodes were attached to the subject's scalp following the 10–20 system.¹⁵ The ground electrode was placed on the forehead, the main electrodes called Fz and Cz were placed on the scalp, in between the reference electrode Fcz, following the

midline to the vertex (for further information see 15). Two additional electrodes were placed each on one of the mastoids. Additionally the EOG electrode (electrookulography) was in the position below the eye to record eye blinking as disturbing evoked potentials in the EEG (see also 13,14). The BrainAmp amplifier (Brain Products GmbH, Germany) connected to a PC via an adapter box and the Brain Vision recorder software 2.0 were used for the EEG recordings. The goal was to detect the perceived PSR as N100 for the phonation paradigm and as MMN for the speech paradigm.¹³ The MMN is the difference waveform between the neurophysiological response of a series of expected standard stimuli and an unexpected deviant stimulus. For the phonation paradigm there was no series of standard stimuli presented and therefore, the N100 was chosen as perception of the deviant stimulus (see also 13,14). The speech paradigm included the repeated presentation of the disyllabic [‘mama] and the pitch shift of the word as deviant stimulus, so that a computation of the MMN was possible. A grand average was computed over all trials of all participating subjects of the important electrodes Fz and Cz. Waveforms with more than 1000 μV and less than 100 μV were rated as non-physiological reactions and were discarded from further analysis (see also 13,14). For statistical analysis an interval of 100 ms before stimulus onset until 500 ms after stimulus onset was selected.

Acoustic voice signal

With a Logitech Premium Stereo Headset (USB port, 22.05 kHz) and a sampling rate of 44.1 kHz the acoustic voice signal was recorded. For the detection of the PSR in the acoustic voice signal the interval of 500 ms before stimulus onset was compared to the interval of 500 ms after stimulus onset (see also 13,14 for the validated setup and other studies^{16,17}). Therefore, each interval of 500 ms was subdivided into 23 intervals and 60 ms with an overlap of 40 ms to the following interval.

Three kinds of behavioral responses were expected to possibly occur: (1) an increase of fundamental frequency (followers), (2) a decrease of fundamental frequency (opposers), and (3) no responses (non-responders).

EGG

The electroglottogram (EGG) was recorded with the Laryngograph Processor (Laryngograph Ltd., England) via a collar with two electrodes and a sampling rate of 44.1 kHz. The analysis of the PSR was computed in the same way than for the acoustic signal, both intervals (pre- and post-stimulus) subdivided into 23 intervals (see 13).

High-speed videoendoscopy (HSE)

For the HSE recordings a flexible laryngoscope (Olympus ENF GP, Olympus GmbH, Germany) was coupled via a 25 mm lens to a Photron SA 1.1 (Photron Ltd., Japan) high-speed camera. The recordings were conducted with 8000 fps and a spatial resolution of 128 \times 128 pixel. During data preparation and analysis the glottal area waveform

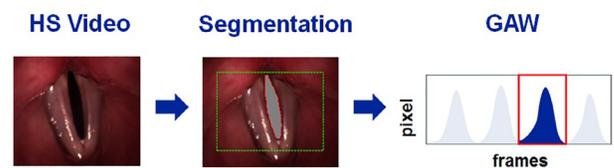


FIGURE 1. Analysis of the GAW extracted from the high-speed video.

(GAW) was extracted using the in-house software *Glottis Analysis Tools*. The GAW is the function of pixels over time detected in the glottis. When the glottis is wide open, we get a high amount of pixels and when the glottis is closed, there are fewer pixels (see Figure 1). The fundamental frequency was extracted from the GAW. The PSR was analyzed within the video signal comparing the acquired parameters 500 ms before and 500 ms after stimulus onset (see 13 for further description and validation).

For the analysis of the PSR of the signals of the acoustic voice, EGG, and high-speed video, different parameters were analyzed. The acquired parameters were commonly used in voice assessment and comprise acoustic perturbation measures for the objective description of voice quality: fundamental frequency, HNR,¹⁸ normalized noise energy,¹⁹ SNR-v1,²⁰ CPP,²¹ spectral flatness,²² shimmer,²³ APQ-3,^{22,24} APF,^{22,24} jitter,^{23,25} and PPQ-3.^{22,24} The perturbation measures jitter and shimmer and their derivatives APQ-3 (shimmer computed over 3 cycles), APF (measure of period perturbation), and PPQ-3 (jitter computed over 3 cycles) were analyzed for the fluctuations in period and amplitude of the vocal fold oscillations. Harmonics-to-noise ratio gives the ratio between the harmonic and the non-harmonic parts of the voice signal. Similar, the SNR displays the portion of noise in the acoustic voice signal. Together, HNR, SNR, spectral flatness, and normalized noise energy estimate the noise level in voiced speech signals and are used within the present study to describe voice quality within normal voices. The periodicity of the voice signal resp. the breathiness can be also determined with the CPP. An aperiodic voice signal results in decreased CPP.

All acoustic perturbation measures were computed and compared between the “no pitch” condition (500 ms before stimulus onset) and “pitch” condition (500 ms post stimulus onset).

Statistical analysis

The statistical testing was computed with SPSS 21 (IBM Corp. 2012). For the statistical detection of the PSR in the acoustic voice signal, EGG, and video signal, we compared the values of the fundamental frequency (dependent variable) interval 500 ms before stimulus onset with the interval 500 ms post stimulus. Furthermore, we compared within groups all parameters of all signals between the “no pitch” and “pitch” – condition for the analysis of the voice quality changes (Jitter, Shimmer HNR, etc – dependent variable) during the PSR (stimulus: independent variable) for both paradigms. To analyze the voice quality between the control group and the group with MTD, we conducted group

TABLE 1.
Subjects with Complete Analysis Showing the Detection of the PSR

	Phonation paradigm		Speech paradigm	
	Controls	MTD	Controls	MTD
# subjects	28	16	31	16
PSR	17	10	22	12
No PSR	11	6	9	4

PSR: pitch-shift reflex; MTD: muscle tension dysphonia.

comparisons for the “no pitch” and “pitch” – condition. For the whole statistical analysis the *t* test or Wilcoxon signed rank test was computed depending on the occurrence of the normal distribution. The level of significance for all statistical testing was set to $\alpha = 0.05$.

RESULTS

Evocation of the PSR in the phonation paradigm

For 39 subjects (33 healthy and 6 patients with MTD) no signal analysis could be conducted due to the subjects' discomfort (1 hour experiment) or due to the insufficient quality of the HSI image segmentation (covered vocal folds, mucus on vocal folds and blurry videos). Hence, the recordings of 28 healthy subjects and of 16 patients with MTD could be analyzed for the phonation paradigm. The PSR was detected for 17 healthy subjects and 10 patients with MTD. No PSR was measured in 11 healthy subjects and in 6 patients with MTD (see Table 1).

Evocation of the PSR in the speech paradigm

The PSR could be detected in 22 subjects of the control group and in 12 patients with MTD (see Table 1). In 13 subjects (9 control group, 4 patients with MTD) no PSR was identified.

Temporal course of the evocation and distinctness of the PSR

Phonation paradigm

The mean overall latency of the N100 of all subjects is 121 ms post stimulus (see Table 2). The magnitude of the N100 is $-4.3 \pm 3.4 \mu\text{V}$ for the controls and thus larger than the magnitude of the patients with MTD with -2.4 ± 2.2

μV . The voice pitch response of the controls occurs 217 ± 136 ms with a magnitude of 10.6 ± 5.8 cents and for the patients with MTD 190 ± 167 ms with a magnitude of 23.8 ± 18.8 cents after stimulus onset (see Table 2). The voice pitch response of the patients with MTD is measured as faster (latency of voice pitch response: 312 ms vs. 338 ms for controls) and with a larger magnitude. The latencies of the voice pitch responses (VP_L) are higher than those of the auditory responses ($\text{N100}_L = \text{A1}$) (see Figure 2). This reflects that the voice pitch change is initiated after the cortical runtime A1 and that the voice pitch response is evoked by the auditory process (see Figure 2). The second and executive part of the auditory process (A2), where the PSR is executed, takes almost twice the time (approx. 200 ms) as the cortical runtime (A1, approx. 121 ms). A2 consists of the kinesthetic laryngeal preresponse preparation time (K1) and the laryngeal implementation time (K2) of the voice pitch change (Figure 2). While K2 can be seen as supporting the fine-tuning for the implementation of the voice pitch change, K1 is understood as the kinesthetic process where the preresponse (prephonatory) settings are conducted (Figure 2). Time intervals of K1 (preparation) are shorter than those of K2 (execution). The kinesthetic component K1 (preparation) showed shortest duration, followed by the cortical runtime (A1) and the kinesthetic process K2 (execution and fine tuning). The auditory process A2 is longer than A1 and per definition as long as the sum of K1 and K2. K1 as the preresponse kinesthetic process seems to be shorter for the MTD-group (34 ± 82 ms) than for the healthy group (61 ± 72 ms).

Comparison of the PSR of the controls vs. patients with MTD (phonation paradigm). The pitch perception perception, response latency and magnitude of both groups are depicted in Figure 3. No significant differences between both groups were found within the analysis with the Wilcoxon signed rank test (A1: $P = 0.941$; A2: $P = 0.749$; K1: $P = 0.443$; K2: $P = 0.749$; VP_L : $P = 0.652$; N100_L : $P = 0.94$). For the comparison of the voice pitch magnitude (VP_M) a significant difference could be found ($P = 0.046$).

Speech paradigm

The mean values of the latency and magnitude of the voice pitch response as well as of the MMN (EEG signal) of the control group and of the patients with MTD are within the

TABLE 2.
Results of Subjects Showing A Voice Pitch Response in the Phonation Paradigm.

Phonation paradigm							N100_M (μV)
	VP_L (ms)	VP_M (cents)	A1 (N100_L) (ms)	A2 (ms)	K1 (ms)	K2 (ms)	
Controls	338 ± 139	10.6 ± 5.8	121 ± 18	217 ± 136	61 ± 72	157 ± 124	-4.3 ± 3.4
MTD	312 ± 169	23.8 ± 18.8	121 ± 20	190 ± 167	34 ± 82	156 ± 109	-2.4 ± 2.2

Voice signal: voice pitch response latency = VP_L ; voice pitch response magnitude = VP_M . **EEG signal:** A1 = N100 latency = N100_L ; N100 magnitude = N100_M ; area under the curve = AUC. **Relations between voice and EEG signal:** begin of the voice pitch response (VP_B) – $\text{N100}_L = \text{K1}$; $\text{VP}_L - \text{VP}_B = \text{K2}$; $\text{VP}_L - \text{N100}_L = \text{A2}$.

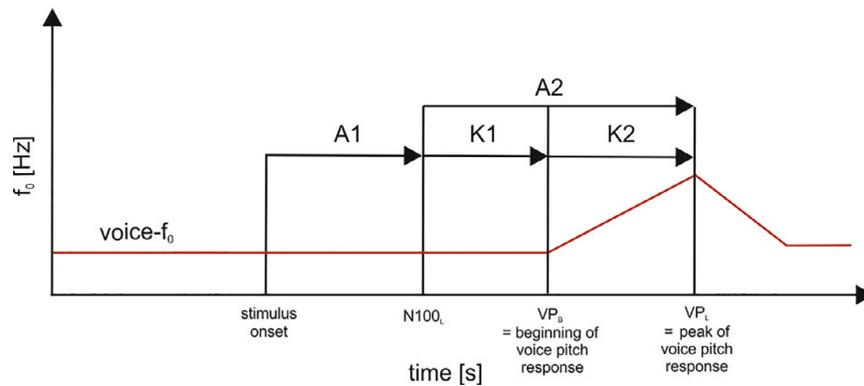


FIGURE 2. Schematic depiction of the auditory and kinesthetic control process of the phonation paradigm.

TABLE 3.
Results of Subjects Showing a Voice Pitch Response in the Speech Paradigm.

Speech paradigm	VP_L (ms)	VP_M (cents)	MMN_L (ms)	MMN_M (μV)	MMN_{OL} (ms)	MMN_D (ms)	$VP_L - MMN_L$ (ms)
Controls	256 ± 109	43.6 ± 24.0	232 ± 77	-9.8 ± 2.1	183 ± 57	61 ± 39	-8 ± 129
MTD	214 ± 104	69.8 ± 33.1	230 ± 29	-10.2 ± 5.0	163 ± 46	72 ± 59	5 ± 121

Voice signal: voice pitch response latency = VP_L ; voice pitch response magnitude = VP_M . **EEG signal:** MMN latency = MMN_L ; MMN magnitude = MMN_M ; MMN onset latency = MMN_{OL} ; MMN duration = MMN_D . **Relations between voice and EEG signal:** $VP_L - MMN_L$.

standard deviation of each other (see Table 3; VP_L and VP_M ; MMN_L and MMN_M). There is a negligible time difference ($VP_L - MMN_L$) between the voice pitch response VP_L and MMN peak occurrence MMN_L (Table 3, last column: -8 ± 129 vs. 5 ± 121). Furthermore, the beginning of the MMN (MMN_{OL}) is located approx. 50–70 ms before the voice pitch response (VP_L).

Comparison of the PSR of the controls vs. patients with MTD (speech paradigm). The pitch perception perception, response latency and magnitude of both groups are depicted in Figure 4. No significant differences between both groups were found within the analysis with the Wilcoxon signed rank test for the perception of the pitch shift (MMN_L : $P = 0.949$) and the latency of the following physiological voice response (VP_L : $P = 0.846$). However, the magnitude of the voice pitch response showed a significant difference between both groups (VP_M : $P = 0.046$).

Comparison between the “no pitch” vs. “pitch” condition regarding all analysis parameters of all recorded signals

Table 4 shows all parameters of the HSE-, audio, and EGG signal of both paradigms which reflect significant changes between the “no pitch” and “pitch” condition. During the “no pitch” condition the subjects’ phonation/speech was not pitched up whereas in the “pitch” condition it was pitched up for 300 ms.

For the control group HNR of the audio signal turned out to be the prominent parameter changing during the PSR. In the speech paradigm the fundamental frequency of the audio, EGG, and video signal as well as the CPP of the EGG signal is the prominent parameter during the PSR (see Table 4). Summarizing the results for the control group, the fundamental frequency and increasing noise changes by the PSR.

For the MTD group mainly the shimmer and derivatives of the amplitude perturbation of the audio signal in the phonation paradigm turned out to be the prominent parameters during the PSR (see Table 4). In the speech paradigm F_0 of the EGG signal proved to be the only parameter reflecting the pitch response. Summarizing the results for the patients with MTD, increasing irregularities of the vocal fold vibration occur by the pitch shift.

Comparison of both groups regarding the most prominent parameters

Comparing both groups with regard to the most prominent parameters during the pitch shift concerning the phonation paradigm reveals in the phonation paradigm no significant difference between both groups. In the speech paradigm the fundamental frequency (F_0) of the video signal and the CPP of the EGG signal show significant differences between the control group and the patients with MTD. Thus, primarily the periodicity of the voice signal during the “no pitch” condition and the frequency during the “pitch” condition lead to significant differences between both groups.

TABLE 4.
Significant Parameters Within Groups Between the Sequences Before and During the Pitch Perturbation of the Signals HSE, Acoustic and EGG for Both Paradigms

Paradigm	Audio	MW ± SD	EGG	MW ± SD	HSE	MW ± SD
Control group						
Speech	F_0	191.4 ± 53.53 (np)	CPP	0.78 ± 0.26 (np)	F_0	191.65 ± 53.71(np)
	$P=0.026$	194.9 ± 51.26 (p)	$P=0.049$	0.97 ± 0.49 (p)	$P=0.022$	194.8 ± 51.52 (p)
	–		F_0	191.78 ± 53.72 (np)	–	
Phonation	HNR	8.88 ± 3.43 (np)	–	195.03 ± 51.69 (p)	–	
	$P=0.050$	7.64 ± 4.22 (p)				
Patients with MTD						
Phonation	Shim %	3.37 ± 2.5 (np)	–			–
	$P=0.023$	2.47 ± 1.83 (p)				
	APQ_3 %	6.0 ± 4.91 (np)	–			–
	$P=0.018$	4.34 ± 4.13 (p)				
	APF %	7.68 ± 4.88 (np)	–			–
Speech		5.74 ± 3.95 (p)				
	–			F_0	184.05 ± 50.21 (np)	–
			$P=0.018$	188.83 ± 49.56 (p)		

Acoustic= acoustic voice signal; EGG= electroglottogram; HSE= high-speed endoscopy; CPP= cepstral peak prominence; HNR= harmonics to noise ratio; F_0 = base frequency; APQ_3%= amplitude perturbation quotient (three cycles); APF % = amplitude perturbation factor; Shim % = Shimmer %; np = “no pitch” condition; p = “pitch condition”.

DISCUSSION

The aim of this study was to investigate whether different auditory and kinesthetic processes can be detected between normal speakers and patients with MTD during phonation and speech. Therefore, we first detected the PSR as a neurophysiological response to a pitch perturbation in the auditory feedback by the analysis of the EEG, the acoustic voice signal and the EGG signal. We further investigated the quality of phonation during the PSR by adding high-speed video endoscopy to visualize vocal fold dynamics during the pitch shift response. The PSR was evoked by pitching up the subjects' own voice by 700 cents for 300 ms during the sustained phonation of the vowel /a/ and during the articulation of the disyllabic word [‘mama].

Phonation paradigm

N100

The N100 was detected for all subjects showing a PSR in the voice signal. The auditory responses of both groups (N100: 121 ± 18 ms CG, 121 ± 20 ms MTD) are comparable to previous studies.^{26,27} Both studies evoked an auditory response of 123 ± 13 ms and 129 ± 4 ms by a pitch-shift of 200 cents for 200 ms. Regarding the cortical runtime (N100_L) no differences between the control group and the patients with MTD could be found ($P=0.94$). There are no differences between both groups in the perception of the pitch shift.

Response latency VP_L

The latencies (VP_L) of the voice pitch responses of both groups (338 ± 139 ms CG, 312 ± 169 ms MTD) show a high inter and

intra subject variability. Diverse studies of Burnett et al. (1998)¹¹ found similar response latencies for healthy subjects ranging from less than 100 ms to more than 400 ms. Shorter response latencies are reported by Liu et al. (2010)¹⁶ and Behroozmand et al. (2011)²⁶ varying between 200 ms and 276 ms post stimulus. As expected, our data reflect the initiation of the voice pitch response by the auditory process (cortical runtime < VP_L). Phonation and speech production is controlled by feedback and feedforward mechanisms involving auditory and somatosensory processes [DIVA model (Directions Into Velocities of Articulators)], Guenther et al. (1998, 2006).^{4,28} In order to correct errors during phonation and speech, we rely on the auditory and kinesthetic feedback to adapt phonation and speech. No differences could be found between both groups regarding response latency ($P=0.652$). The patients with MTD show a normal auditory feedback process (N100_L and VP_L).

Response magnitude VP_M

The response magnitude of both groups differ by their mean value, but lie within the standard deviation of each other (10.6 ± 5.8 cents CG; 23.8 ± 18.8 cents MTD). In this case the patients with MTD reacted with a larger magnitude. Multiple authors also reported varying response magnitudes for healthy subjects ranging from 7.27 cents up to 69 cents.^{2,3,27,29} However, comparing the results of both group statistically showed a significant larger response magnitude of the patients with MTD ($P=0.046$). Larson et al. (2008)³ detected a larger magnitude of the voice pitch response VP_M by eliminating the kinesthetic control. This result leads to the assumption that the patients with MTD do have an

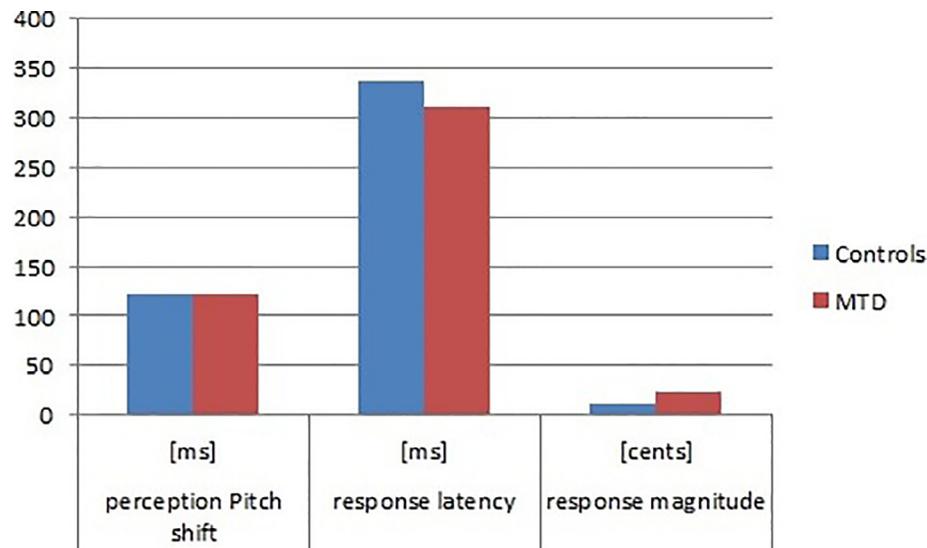


FIGURE 3. Phonation paradigm: latency and magnitude of the pitch perception and voice response in both groups.

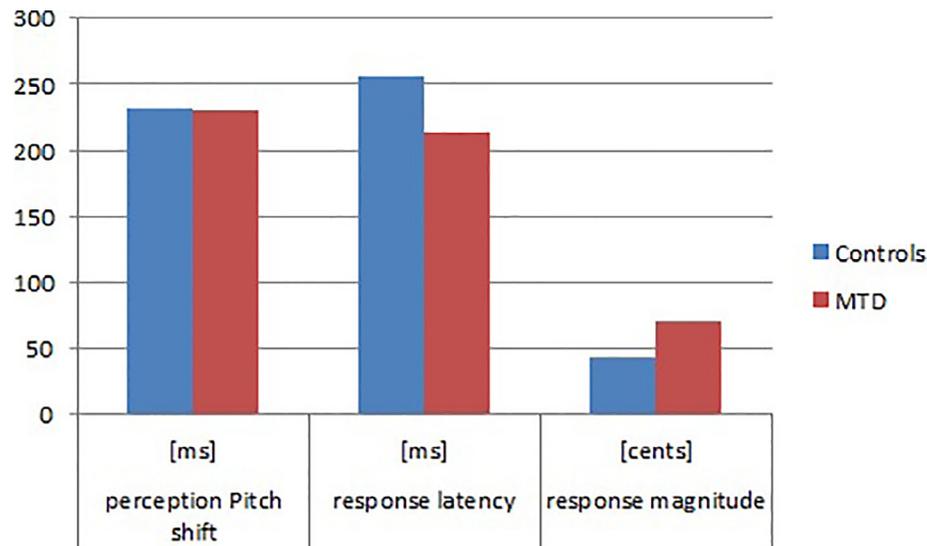


FIGURE 4. Speech paradigm: latency and magnitude of the pitch perception and voice response in both groups.

affected kinesthetic process. Further indication could be found within the analysis of the auditory and kinesthetic feedback. Comparing the auditory (A1 and A2, Figure 1) and kinesthetic process (K1 and K2, Figure 1) of both groups between each other, showed descriptively a shorter kinesthetic process of the patients with MTD (61 ± 72 ms CG, 34 ± 82 ms MTD), but without statistical significance (K1: $P = 0.443$, K2: $P = 0.749$). The high standard deviation might explain the non-significant statistical result.

Comparison between the “no pitch” vs. “pitch” condition regarding all analysis parameters of all recorded signals

Considering within groups changes between the “no pitch” and “pitch” condition, the control group showed an increased noise (HNR) within the phonation during the

pitch shift while the patients with MTD showed less irregularities (shimmer, APQ) in the vocal fold oscillation during the pitch shift. The pitch shift does not reduce voice quality for both groups. Typical values of the HNR for normal voices are about 7–8 dB.³⁰ The control group shows in the “no pitch” condition as well as in the “pitch” condition values between 7 and 8 dB, hence the voice quality remains unremarkable. The patients with MTD improved voice quality and the control group stayed within a normal range, thus statistically significant differences could not be found between both groups regarding voice quality during the PSR in the phonation paradigm.

Summarizing all results of the phonation paradigm, indications for an affected kinesthetic process could be found, revealing a larger magnitude and a tendency for a shorter kinesthetic process.

Speech paradigm

MMN

The PSR could also be detected in both groups in the speech paradigm as MMN with similar values (232 ± 77 ms CG, 230 ± 29 ms MTD). No statistically significant differences could be found between both groups ($P = 0.949$). Thus, also the speech paradigm confirms a normal perception of the pitch shift. In general, the grand average values of our groups lie in a normal range (~ 300 ms³¹).

Response latency VP_L

The latency of the voice pitch response (256 ± 109 ms CG, 214 ± 104 ms MTD) did not show any statistically significant differences between both groups ($P = 0.846$). This result confirms, like in the phonation paradigm, an undisturbed auditory feedback process of the patients with MTD (MMN_L and VP_L). Generally, the values of the response latency of both groups lie in range of the reported values of Donath et al. (2002).³² They describe the voice response in a disyllabic word at maximum in the second syllable (1. syllable: approx. 200 ms, 2. syllable: approx. 300 ms). The averaged voice pitch response in our study varies highly, but seems to occur at maximum in the second syllable.

Between the voice pitch response (VP_L) and MMN peak occurrence (MMN_L) we found a negligible time difference. This indicates that the voice pitch response is initiated quite before the peak of the auditory perception process (MMN_L) for the deviant stimulus is reached. Table 3 shows that the beginning of the MMN (MMN_{OL}) is located approx. 50–70 ms before the voice pitch response (VP_L). Relating this to the phonation paradigm would mean that the beginning of the MMN is located within the execution of the pitch raise (K2, Figure 1). Therefore, it may be concluded that the MMN, which reflects higher cortical change detection processes (eg 33) than the N100, is not crucial for the initiation of the rather simple PSR process. Previous studies concerning the MMN as reflection of higher cognitive processes showed that familiar sounds and meaningful words of a language elicited a larger MMN than unfamiliar sounds or meaningless pseudo-words.³³ Further, lexical and semantic properties of word stems and affixes started cortical activation patterns and led to the assumption that the MMN indicates higher cortical processes in language processing.³³ Our study focused on the manipulation of F_0 and did not require higher cortical processing. We located the MMN within the execution of the pitch raise (K2, Figure 1), thus confirming that no higher cortical processes are needed for the feedback mechanisms.

Response magnitude VP_M

Likewise, in the speech paradigm the response magnitude of both groups (43.6 ± 24.0 cents CG, 69.8 ± 33.1 cents MTD) reflects a high variability within groups and a descriptive difference between the control group and the patients with MTD. Statistical testing verified a significant difference between both groups regarding the response magnitude in the

speech paradigm ($P = 0.046$). The patients with MTD seem to show an overriding reaction that again confirms the assumption of a disturbed kinesthetic feedback process.

Comparing the response magnitudes of both groups with other studies, similar results can be found. Former studies stated a range of 13–97 cents³¹ or values between 15 and 65 cents.³⁴ It is possible that the magnitude of the voice pitch response can be influenced by syllable stress and the magnitude of the stimulus.^{9,17} With increasing magnitude of the stimulus and with syllable stress on the first syllable, the response magnitude is larger.^{9,17} In our study large response magnitudes can have occurred due to stress on the first syllable and a large magnitude of the stimulus (700 cents).

Comparison between the “no pitch” vs. “pitch” condition regarding all analysis parameters of all recorded signals

The analysis within each group between the “no pitch” and “pitch” condition in the speech paradigm revealed the fundamental frequency and the CPP for the control group and the fundamental frequency for the patients with MTD as prominent parameter for the pitch change. Group comparison between the control speakers and the patients with MTD determined primarily the CPP in the “no pitch” condition ($P = 0.028$) (fundamental frequency in “pitch” condition, $P = 0.057$) as differentiating parameters. The CPP measures the degree of harmony within a voice sample and reflects lower values for aperiodic voices due to a lower degree of harmony that can be measured.³⁵ Therefore, the CPP reflects an aperiodic voice (MTD) by decreased values and thus, the significant difference between both groups is expected under the “no pitch” condition and confirms the diagnosis of MTD. Patients with MTD show typically an aperiodic voice¹ and should be differentiated from the control group during the “no pitch” condition.

Summarizing the results of the speech paradigm, likewise a larger response magnitude of the patients with MTD can be found leading to the assumption of an existing disturbed kinesthetic feedback process.

CONCLUSION

Patients with MTD do not show any problems with the auditory feedback process (pitch perception and response latency) within phonation or speech. However, larger response magnitudes and the tendency of a shorter kinesthetic process become obvious and occur during phonation and/or speech suggesting the existence of a disturbed kinesthetic feedback process.

ACKNOWLEDGMENT

This work was supported by Deutsche Forschungsgemeinschaft (DFG), grant no. DO1247/4-1.

LITERATURE

1. Schultz-Coulon HJ. The neuromuscular phonatory control system and vocal function. *Acta Oto-Laryngol.* 1978;86:142–153.
2. Liu P, Chen Z, Jones JA, Huang D, Liu H. Auditory feedback control of vocal pitch during sustained vocalization: a cross-sectional study of adult aging. *PLoS One.* 2011;6:e22791.
3. Larson CR, Altman KW, Liu H, Hain TC. Interactions between auditory and somatosensory feedback for voice F0 control. *Exp. Brain Res.* 2008;187:613–621.
4. Guenther FH. Cortical interactions underlying the production of speech sounds. *J Com Dis.* 2006;39:350–365.
5. Burnett TA, Senner JE, Larson CR. Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. *J. Voice.* 1997;11:202–211.
6. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am.* 2000;108(3 Pt 1):1246–1251.
7. Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift velocity on voice F0 responses. *J Acoust Soc Am.* 2000;107:559–564.
8. Sivasankar M, Bauer JJ, Babu T, Larson CR. Voice responses to changes in pitch of voice or tone auditory feedback. *J Acoust Soc Am.* 2005;117:850–857.
9. Liu H, Larson CR. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *J Acoust Soc Am.* 2007;122:3671–3677.
10. Behroozmand R, Korzyukov O, Sattler L, Larson CR. Opposing and following vocal responses to pitch-shifted auditory feedback. Evidence for different mechanisms of voice pitch control. *J Acoust Soc Am.* 2012;132:2468–2477.
11. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am.* 1998;103:3153–3161.
12. Larson CR, Sun J, Hain TC. Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *J Acoust Soc Am.* 2007;121:2862–2872.
13. Petermann S, Döllinger M, Kniesburges S, Ziethe A. Analysis method for the neurological and physiological processes underlying the pitch-shift reflex. *Acta Acust. united Acust.* 2016;102:284–297.
14. Arbeiter M, Petermann S, Hoppe U, Bohr C, Doellinger M, Ziethe A. Analysis of the auditory feedback and phonation in normal voices. *Ann Otol Rhinol Laryngol.* 2017;127:89–98.
15. Klem GH, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the International Federation. *Clin Neurophysiol Supplement.* 1999;52:3–6.
16. Liu H, Auger J, Larson CR. Voice fundamental frequency modulates vocal response to pitch perturbations during English speech. *J Acoust Soc Am.* 2010;127:1–5.
17. Chen SH, Liu H, Xu Y, Larson CR. Voice F0 responses to pitch-shifted voice feedback during English speech. *J Acoust Soc Am.* 2007;121:1157–1163.
18. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am.* 1982;71:1544–1549.
19. Kasuya H, Ogawa S, Mashima K, Ebihara S. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J Acoust Soc Am.* 1986;80:1329–1334.
20. Klingholz F. The measurement of the signal-to-noise ratio in continuous speech. *Speech Commun.* 1987;6:15–26.
21. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res.* 1994;37:769–778.
22. Lessing J. Entwicklung einer Klassifikationsmethode zur akustischen Analyse fortlaufender Sprache unterschiedlicher Stimmgüte mittels Neuronaler Netze und deren Anwendung [Development of a classification method for the acoustic analysis of continuous speech of different vocal quality by means of neural networks their applications]. Göttingen, Germany: Georg-August-Universität Göttingen, Mathematisch-Naturwissenschaftlich Fakultät, (2007).
23. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res.* 1996;39:126–134.
24. Michaelis D. Das Göttinger Heiserkeits-Diagramm-Entwicklung und Prüfung eines akustischen Verfahrens zur objektiven Stimmgütebeurteilung pathologischer Stimmen [Development and testing of an acoustic method for objective voice quality assessment of pathological voices]. Dissertation zur Erlangung des Doktorgrades, Mathematisch-Naturwissenschaftliche Fakultät, Georg-August-Universität Göttingen, (1999).
25. Inwald EC, Döllinger M, Schuster M, Eysholdt U, Bohr C. Multiparametric analysis of vocal fold vibrations in healthy and disordered voices in high-speed imaging. *J. Voice.* 2011;25:576–590.
26. Behroozmand R, Korzyukov O, Larson CR. Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clin Neurophysiol.* 2011;122:2408–2417.
27. Li W, Chen Z, Liu P, Zhang B, Huang D, Liu H. Neurophysiological evidence of differential mechanisms involved in producing opposing and following responses to altered auditory feedback. *Clin Neurophysiol.* 2013;124:2161–2171.
28. Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* 1998;105:611–633.
29. Hain TC, Burnett TA, Larson CR, Kiran S. Effects of delayed auditory feedback (DAF) on the pitch-shift reflex. *J Acoust Soc Am.* 2001;109:2146–2152.
30. Ferrand CT. Harmonics-to-noise ratio: an index of vocal aging. *J Voice.* 2002;16:480–487.
31. Duncan CC, Barry RJ, Connolly JF, Fischer C, Michie PT, Näätänen R, Polich J, Reinvang I, Van Petten C. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin Neurophysiol.* 2009;120:1883–1908.
32. Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *J Acoust Soc Am.* 2002;111:357–366.
33. Pulvermüller F, Shtyrov Y. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog Neurobiol.* 2006;79:49–71.
34. Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J Speech Lang Hear Res.* 2001;44:577–584.
35. Heman-Ackah YD, Sataloff RT, Laureyns G, et al. Quantifying the cepstral peak prominence, a measure of dysphonia. *J. Voice.* 2014;28:783–788.