World Journal of Psychiatry

Submit a Manuscript: https://www.f6publishing.com

World J Psychiatry 2026 March 19; 16(3): 113646

DOI: 10.5498/wjp.v16.i3.113646

ISSN 2220-3206 (online)

ORIGINAL ARTICLE

Retrospective Study

Voice as the sound of the psyche: Changes in prosodic characteristics in depressive patients in long-term therapy

Christoph Holzweber, Beate Hennenberg, Laurenz Stastka, Malte Kob, Henriette Löffler-Stastka

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's classification

Scientific Quality: Grade B, Grade B

Novelty: Grade B, Grade B

Creativity or Innovation: Grade B, Grade B

Scientific Significance: Grade A, Grade C

P-Reviewer: Deng J, PhD, Lecturer, China; Pan RB, PhD, Principal Investigator, China

Received: September 1, 2025

Revised: October 6, 2025

Accepted: November 24, 2025

Published online: March 19, 2026

Processing time: 181 Days and 24 Hours

Christoph Holzweber, Program for Clinical Academic Psychotherapeutic Propedeutics and Medical Humanities, Postgraduate Unit, Medical University Vienna, Vienna 1090, Austria

Beate Hennenberg, Department of Music Education Research and Practice, University of Music and Performing Arts, Vienna 1040, Austria

Laurenz Stastka, Ludwig-Maximilians University Munich, Munich 81377, Germany

Malte Kob, Erich Thienhaus Institute, Detmold University of Music, Detmold 32756, North Rhine-Westphalia, Germany

Henriette Löffler-Stastka, Department of Psychoanalysis and Psychotherapy, Medical University Vienna, Vienna 1090, Austria

ORCID number: Henriette Löffler-Stastka 0000-0001-8785-0435.

Corresponding author: Henriette Löffler-Stastka, MD, PhD, Dean, Director, Department of Psychoanalysis and Psychotherapy, Medical University Vienna, Währinger Gürtel 18-20, Vienna 1090, Austria. henriette.loeffler-stastka@meduniwien.ac.at

Abstract

BACKGROUND

Voice is individual and therefore several studies consider voice as a biomarker of health. Focusing on the change of voice parameters of depressed patients can provide an additional, objective measurement of the patient's individual progress during psychotherapy. This retrospective study is focused on the analysis of voice parameters, specifically prosodic parameters, gathered from sessions and real-world data out of the Munich psychotherapy study (MPS).

AIM

To investigate the prosodic parameters of patients with depression, their change over a long-term therapy and correlate them with Beck depression inventory (BDI) data. The analysis on the prosodic parameters aims to find, if and to what extent, a change in the patient's voice can be discovered and identified. This data is then correlated with already available BDI data of each individual patient. The hypothesis is that voice expression is impacted by depression and changes with the state of depression.

METHODS
We performed a retrospective investigation of n = 25 patients, who originally participated in the MPS, and their change in prosodic parameters. An analysis of the mean fundamental frequency f_o and its standard deviation was performed on six psychotherapy sessions per patient, divided into three sessions at the beginning of the therapy and three sessions at the end of the therapy. The technical analyses on the voice part have been conducted with Praat, a phonetic analysis software developed at the University of Amsterdam.

RESULTS
The mean fundamental frequency was $f_{o,\text{mean}}$ = 183.1 Hz at the beginning and $f_{o,\text{mean}}$ = 187.8 Hz at the end, in a compounded group of all patients (men and women). The average BDI of all patients at the beginning was BDI_{mean} = 29.96 and BDI_{mean} = 9.6 at the end. A positive influence of psychotherapy can be seen in the majority of patients not only in a reduced BDI rating, but also in an increased baseline frequency.

CONCLUSION
With regard to the clinical picture of depression, its recognition and constant monitoring during therapy, it is important to exhaust all possibilities and examine additional relevant data. Prosodic characteristics can make a contribution here for various reasons. On the one hand, a person acts unconsciously *via* and through their voice and emotional states become detectable. On the other hand, voice can be represented in objective parameters and is therefore suitable for analyzing individual characteristics and changes. Dealing with the topic from the perspective of psychotherapy therefore seems unavoidable, as an additional component is included in the holistic view of the person.

Key Words: Voice; Depression; Fundamental frequency; Emotional state

Core Tip: Voice is a central component in interacting with the world and a major factor in defining the relationship between patient and therapist. The complex system enabling us to produce a vocal tone is influenced by several physiological components and transports information about our feelings and emotions. Especially for patients with depression the voice is often described as muted and monotone. This research initially provides theoretical background and then focuses on the evolution and change of the fundamental vocal frequency of patients with depression during long lasting psychotherapy. Therefore, a method was created to identify the fundamental frequency of sequences of sessions recorded within the Munich psychotherapy study and applied throughout sessions of 25 individual patients. The phonetic analyses were performed using appropriate phonetic software. The output of the analysis is crosschecked with the output parameters [Beck depression inventory (BDI) for depression] and converted into a formula to calculate a ratio between fundamental frequency and BDI rating scores. Results are shown on an individual basis for specific patients and evaluated for the overall scope and in an overall context.

Citation: Holzweber C, Hennenberg B, Stastka L, Kob M, Löffler-Stastka H. Voice as the sound of the psyche: Changes in prosodic characteristics in depressive patients in long-term therapy. *World J Psychiatry* 2026; 16(3): 113646
URL: https://www.wjgnet.com/2220-3206/full/v16/i3/113646.htm
DOI: https://dx.doi.org/10.5498/wjp.v16.i3.113646

INTRODUCTION

According to the World Health Organization definition, depression is a common mental disorder. At the heart of the symptoms are changes in the thymopsyche, *i.e.*, the feelings and expressions of emotion. The term is derived from the Latin word deprimire, which can be translated to mean to depress. The clinical diagnosis of depression describes the disorder in which the elements of dejection, lack of energy and drive, as well as the loss of joy and interests are found and united[1]. Around 3.8% of the population suffer from depression worldwide, that is 280 million people. In Austria, 6.5% of adults are confronted with depressive disorders during their lifetime, with women being slightly more affected (6.8%) than men (6.3%)[2]. Based on these figures, it can be said that depression is and will continue to be one of the major global challenges for healthcare systems[3]. If left untreated, depression can have a massive impact on quality of life at various levels, both professionally and socially[3]. The heterogeneity of symptoms leads to various clinical pictures, resulting in the need for trained and qualified professionals to find the right diagnosis.

Speech and voice could play its role in the early identification and tracking of the progress in depression as individual markers leading to the question “Is Speech the New Blood?”. Voice can be seen as a human expression that is innate and manifests itself in different ways. In contrast to language, which can be acquired, learned and changed, the perceptibility of the voice is a person-specific, unique characteristic, "a vocal fingerprint"[4]. A person's voice is completely individual. There really should be no two people with the same voice. This vocal fingerprint is used in everyday life for authentication in our social environment, but also, for example, as a means of rapid identification in a professional context. Post Finance Bank, for example, uses voice profile recognition in customer contact because they believe it increases security

and saves time[5].

Voice as indicator of therapeutic progress

The human voice is created through the interaction of various structures and muscles that require coordination between the abdomen, chest, neck and head. This shows that the vocal apparatus is an integral part of the body as a whole and is fundamentally connected to it. Voice is a "highly complex process in which the individual organs and functional circuits involved are very finely coordinated and attuned to each other"[6]. Dysfunctions, impairments or disorders of these central structures result in changes in sounds and voice, as it is also the case in depressive disorders[7-9]. The larynx serves as a central and essential organ and is figuratively referred to as the vocal organ[10] of the vocal apparatus. The primary function of the larynx is to keep the lower airways free of foreign bodies in the respiratory process; the secondary function is to produce voice through the glottis[11]. Protecting the airways and the necessary ejection of foreign bodies through reflexively animated coughing movements requires not only sudden closing movements, but also a high level of granularity in the coordination of the control and safety system and the sub-areas that lie above it. The larynx is integrated into the movements of the entire body *via* the muscle chains, head, shoulder girdle, arms, pelvis, spine or legs, and its function is positively or negatively influenced by these.

A distinction is made between phonation, *i.e.*, voice formation, which takes place in the larynx and produces sound events, and articulation, the formation of sounds, which takes place in the mouth and throat. The production of voice occurs primarily through the release of air, initiated by the activity of the respiratory muscles[12] during exhalation. The glottis refers to the voice-forming apparatus consisting of vocal folds. During inhalation, the glottis is open, letting air flow without resistance. Therefore, no sound can be produced. To produce sound, the vocal folds must be set into a state of vibration, which occurs during the process of exhalation. To ensure that sound is not produced with every breath, the brain sends a signal to set the correct state of tension. This is also known as phonation tension[13]. When exhaling, the vocal folds themselves form a resistance, as they contract again due to the activity of the adductor muscles and cause the glottis to close. This condition in succession leads to a build-up and increase in subglottal pressure[12]. During phonation, the resistance of the closed vocal folds is overcome by the pressure of the air flow. The focal folds open, allowing air to flow outwards. The air flows through the glottis opening like a jet. Due to the increasing speed of the outflowing air, the pressure in the glottis decreases, resulting in the glottis closing again and the vocal cords contracting. This is further supported by means of elastic forces of resetting. This is again followed by an increase in pressure, accompanied by the opening of the vocal folds and thus initiating a cycle of opening and closing of the glottis, known as the Bernoulli effect. This continuous cycle of vibration occurs at a certain frequency, which is referred to as the fundamental frequency or fundamental tone of the voice. Physically, the human body acts as a sound source, with the vocal cords producing the primary tone, which is a raw, unfinished sound product, a kind of primal sound. This primary sound still requires amplification for perception. The amplification requires resonance chambers where, similar to a violin, existing hollow bodies resonate. The human voice relies on mouth resonance and nasal resonance to amplify a sound. On the other hand, the sound waves generated by a sound source are propagated in space *via* a medium, in the case of the voice primarily the air, until they reach a receiver[14].

In phonation, frequency is defined as the parameter that determines the rate of vocal fold fluctuations and is perceived as pitch[15]. The pitch resulting from the number of vocal fold vibrations per second is defined as the mean fundamental frequency f_o[16] and expressed in the SI unit Hz. The fundamental frequency f_o, or mean speech pitch, measures the phonic zero point and any persistent shifts from the resting position, which can occur in stress situations[17].

The fundamental frequency is a function of the mass and length of the vocal folds. These are subject to gender differences and therefore result in different bands of fundamental frequencies for men and women. In women, the average fundamental frequency is in the range of 190 Hz to 250 Hz and in men in the range between 100 and 150 Hz[18].

These physical, innate characteristics already define the vocal range or the range of a person's fundamental frequency variations, the use of which is culturally dependent[4]. In addition to taking part in social dialogue and revealing information and emotional states, the voice also provides biological characteristics and markers. Individual information such as gender, age or height is revealed *via* the voice. It can also be used to draw conclusions about national, regional or psychological characteristics (origin, possible group affiliation, competence, personality, emotion, state of mind)[4].

The voice also shapes people on the receiving end, as the neural processing of voices in infants begins in the primary auditory cortex between the fourth and seventh month of life, in a sensitization to emotional prosody that undergoes permanent development and improvement until the maximum performance level of adults is reached[4]. Furthermore there is an evolutionary link between the frequency range of the voice and the auditory system or the sensitivity of the auditory system. Not only is the person producing sound important for the vocal transmission circuit, but also the perception and reaction of the receiver to the voice, which is spontaneous, intuitive and emotional[19].

In the case of linguistic utterances that are emotionally charged, there are now also differences in the glottal pulses depending on the respective emotion, which result in a change in the glottal spectrum. In contrast to utterances that are made neutrally, anger and joy, for example, show longer closed moments of the glottis. In mourning, the glottal excitation is no longer impulse-like as in normal speech but has more of a sinus shape.

Although the voice is unique, it is not completely monotonous and does not always sound the same, but is changed by the state of mind and emotions. A linguistically neutral text is spoken with a hardened use in a strong excitement, with a clear use in a comfortable mood, and with a softened use in a depressive mood. In this respect, the voice also serves as an instrument for adding variations to statements and thus adding further definition and emphasis to the content.

One's own voice can provide intimate insights into one’s emotional state. It can reveal characteristics and emotions such as fears, longings, tension, excitement or sadness. In personality research in the context of the Big Five, it is primarily the two factors neuroticism and extraversion that correlate with the indicators of voice and speech.

The entirety of all acoustically perceptible forms of speech expression is summarized under the term prosody. This includes accentuation, speech rate, rhythm and timbre. The linguistic feature systems include intonation (the acoustic correlate is the fundamental frequency f_o), loudness (intensity) and quantity (duration)[20].

Following these arguments based on literature[7-9,17] our research hypothesis is that vocal parameters change with depression severity.

MATERIALS AND METHODS

Study design and patients

We performed a retrospective single-center observational analysis. The study relies on available audio files from the Munich psychotherapy study (MPS)[21,22], which is a comparative process-outcome study of three therapeutic approaches: Psychoanalytic, psychodynamic, and cognitive-behavioral. The study is a long-term-study that followed the progress of the individual patients up to several years. Although 100 patients were participating in the study, only 93 available sets of audio files from the individual therapy sessions were available and were further screened for its feasibility to be used in the study. The further exclusion criteria were incompleteness of the audio files or insufficient voice quality from the patient in one or more sessions. This investigation was approved by the Institutional Ethics Committee of the Medical University Vienna (number 1746/2023). A subset of 25 patients with complete audio files in the proper quality for voice analysis was selected to start the individual case analysis (Figure 1).

Data collection

All the data related to the patient was retrieved from digital audio files, which has been anonymized. Available data were sex, age and an evaluation of the severity of depression through Beck depression inventory (BDI)[23] evaluation. A pre-evaluation with audio files from actors talking in different mood states and openly available *via* RAVDESS investigating how the emotional state of a person can be comprehended, was made. The specific parameter to be identified was the fundamental frequency. In the main study each of the patients' audio files were screened for speech without interruptions for up to twenty seconds. These uninterrupted speech-fragments occurred every three minutes for up to 15 individual speech-fragments per session. This resulted in data of up to six minutes per patient under normal conditions. Each of the minimum 20 seconds sequences have data points collected every 35 thousandth of a second, resulting in 28 data points per second and up to 280 for a 20 second sequence. Overall, the data collection for each session of was then conducted in 15 data points/15 individual speech-fragments, merging the single 20 second sequences, for further evaluation. The audio files were analyzed from recordings made in the office of psychotherapists. The signal-to-noise ratio (SNR) strongly depended on the distance of the speakers to the recording device and the background noise. For the analysis only recordings were included for which the fundamental frequency could be analyzed for the whole duration of each segment. The SNR for each recording had to be > 30 dB, to ensure an accurate acceptable level[24].

The central analysis tool for the study is a software program used specifically for the main analysis work of the vocal part studies. This is the free software Praat, which is a phonetic analysis software developed at the University of Amsterdam by Paul Boersma and David Weenink and allows spoken material to be broken down into different parameters for further analysis (https://www.fon.hum.uva.nl/praat/).

There are various methods and algorithms for "determining the fundamental frequency from the acoustic signal"[25], the one used by the Praat software is a proprietary development by Paul Boersma. The algorithms from different programs can produce different results for one and the same data set, *i.e.* an audio file, because fundamental frequency algorithms can also be error-prone, which is why the results must always be subjected to a plausibility check[25]. Praat offers the following advantages. It is open access software that is subject to continuous further development and can be used on different operating systems. By creating own scripts, broad-based analyses can be carried out.

Patient's characteristics at baseline

The material of the study included 25 patients with depressive disorder. The median age of the patients at the beginning of the treatment study was 33 years, 20 (80%) were female and 5 (20%) were male and the median therapy time was 720 days. Out of the 25 patients 14 (56%) had severe depression, 7 (28%) medium depression and 4 (16%) mild depression. The average BDI rating at the beginning of the study was 27.56, while at the end of the study it decreased to 9.6. The patients were treated with different psychotherapy methods, in detail 6 (24%) with cognitive behavioral therapy, 11 (44%) with psychoanalytic oriented psychotherapy and 8 (32%) with psychoanalysis. The median time of psychotherapy evaluated between the start and end sessions was 713.12 days. An overall of 2151 audio sequences were analyzed with an average length of 23.72 seconds, which adds up to a total amount of 14.172 hours of data material.

Outcomes and definitions

In music, the intervals are mathematically defined. The semitones, determined by their frequency, are always separated by the factor twelfth root of 2 (factor of 1.059463094).

Each semitone corresponds to a specific frequency and is also represented by a musical instrument digital interface (MIDI) standard tone number. The MIDI standard defines frequencies between 8 Hz and 12.543 Hz, which are determined relative to a reference frequency, usually according to the concert pitch *a'/A4*[26]. The concert pitch a' has a frequency of 440 Hz and the MIDI number 69. Based on this, every fundamental frequency f_o in this range can be converted into a MIDI note number, logarithmically scaled with the following formula[27]:

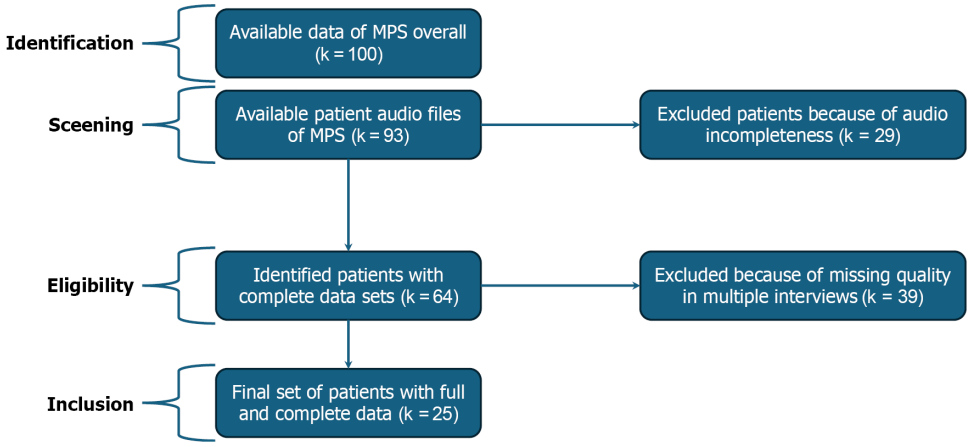

Figure 1 Selection process. MPS: Munich psychotherapy study.

MIDI note number = 12 × [log (f_o/440)/Log (2)] + 69 (1)

The conversion to MIDI notes numbers allows a pooled analysis since the potentially different fundamental frequencies of female and male subjects are not compared directly. Due to the transfer to the logarithmic musical intervals the differences in f_o due to the therapy are not dependent on the original fundamental frequency.

If this conversion is performed for each patient, a MIDI note number is obtained for the end hours and the start hours, the difference between which indicates the relative change. From the individual relevant data of the selected patients, specifically the BDI change and the difference in the MIDI note number, a ratio of the change in relation to the two data can be calculated.

Ratio = (ΔMN/ΔBDI) × (-10) (2)

ΔMN denotes the difference between the calculated MIDI note number of the end interviews and the MIDI note number of the start interviews. The MIDI note number is calculated from the average fundamental frequency using the above formula. It was assumed that the ΔMN value is positive, as the final interviews have higher values than the initial interviews.

ΔBDI is calculated from the value of the BDI at the start of therapy, from which the BDI value at the end of therapy is subtracted. The ΔBDI value is therefore negative. The ratio itself is then calculated by dividing the delta of the baseline frequency by the delta of the BDI and thus obtains the value of Hz per depression point.

For a clear representation and easier interpretation, the ratio is multiplied by -10 in order for positive effects to be represented positively resulting in a meaningful scale. A positive ratio value, resulting from the formula, means an improvement in the vocal parameters. The two-sample *t*-test assuming Unequal Variances was used to examine the significance of relationship between f_o and BDI, *via* the ratio between these two parameters. The respective formula is explained in detail above. Two different groups were compared: Patients with a BDI score > 30 at the beginning of the therapy and patients with a BDI score < 30 at the beginning of the therapy. *P* value < 0.05 is considered as statistically significant. The calculated and identified *P* value was 0.028 for this analysis.

The paired *t*-test was used as method to determine the level of significance of change in total fundamental frequency and BDI over the individual period of the patient's psychotherapy treatment. *P* value of 0.05 is considered as significant. For the fundamental frequency f_o the *P* value resulted in 0.00035, whereas for the BDI the calculated *P* value was *P* = 0.00001.

RESULTS

Of the 25 patients selected, 20 were women (80%) and five men (20%). The women had an average age of 32.8 years at the start of treatment, while men were on average 33.8 years old. The average duration between the first initial interview and the last final interview was 713.12 days, which corresponds to a period of almost two years on average. It should be noted that the recorded interviews do not mark the original start and end of therapy. A few hours took place before and after. The data for the statistical analysis were taken from the time stamps of the audio files and validated against the data given in the MPS study.

A total of 2151 audio sequences were identified in the final interviews and the corresponding voice report data was extracted and collected from the Praat software.

The average duration of an audio sequence is 23.72 seconds, which results in a total data pool of 51021.72 seconds. This corresponds to 850.362 minutes or 14.1727 hours. According to the BDI, 14 patients were diagnosed with severe depression, seven patients with moderate depression and four patients with mild depression. The average BDI of all patients at the beginning was 27.56 and 9.6 at the end (Table 1).

Table 1 Description of the sample, *n* (%)

Variables	Values
Sex	
Female	20 (80)
Male	5 (20)
Age in years (MW, SD)	33 ± 6.33
Suffering from depression	*n* = 25
Depression levels	
Severe	14 (56)
Moderate	7 (28)
Mild	4 (16)
Evaluated sequences	*n* = 2151
Average length	23.72 seconds
Overall data	850.36 minutes, 14.172 hours
Psychotherapy method	
CBT	6 (24)
POP	11 (44)
PA	8 (32)

POP: Psychoanalytic oriented psychotherapy; CBT: Cognitive behavioral therapy; PA: Psychoanalysis.

Statistical data from the evaluation

The mean fundamental frequency was 183.1 Hz at the beginning and 187.8 Hz at the end, in a compounded group of all patients (men and women). For a further linkage of the obtained data of the fundamental frequency change with the data of the BDI and thus the attempt of an interpersonal comparison, additional conversion had to be made, as the changes in the fundamental frequency in men may differ from the change in women. The physiological differences between men and women lie in the different fundamental frequency of the vocal bands. In the analyzed data, men speak between 95 Hz and 123 Hz and women between 160 Hz and 240 Hz. Changes in the fundamental frequency are therefore not directly comparable for men and women, and simply doubling the values for men would distort the results. To integrate the BDI values and maintain consistency in the calculation, it is not possible to work directly with the fundamental frequencies; instead, a conversion had to be carried out using semitones, as defined above.

Combination and correlation of data

Due to the physiognomic difference in the voice area between men and women, a comparison of data simply with fundamental frequency f_o is not leading to the optimal correlation. In the captured data women had a fundamental frequency rate f_o between 160-240 Hz, whereas the one from men was only between 95-123 Hz. Changes in fundamental frequency for women tend to be bigger, because of the higher fundamental frequency overall. Therefore, it is mandatory to find a compensatory element on the voice side, to include women and men in the same correlation with the BDI ratings. In research the semitone is used as a base unit for a logarithmic scale for analysis and display of the pitch of speech.

General changes

The mean fundamental frequency (183.1 Hz at the beginning and 187.8 Hz at the end) showed an increase. The analysis of the fundamental frequency objectified the change in the acoustics. Further analysis points at the data that had been evaluated in the original study with the BDI ratings. The average BDI rating at the beginning of the study was 27.6, while at the end of the study it decreased to 9.6.

A positive influence of psychotherapy can be seen for most of the patients not only in a reduced BDI rating, but also in an increased baseline frequency, as summarized in the following graph (Figure 2).

For most patients, there was an increase in the mean baseline frequency over the course of psychotherapeutic treatment, which was associated with an individual reduction in the depression rating and an increase in the standard deviation. In 64% of patients, this increase in the mean baseline frequency of > 4 Hz could be detected, with 24% showing an increase of up to 3 Hz. Three patients showed a negative development of the fundamental frequency, although this was due to external factors or no progress in the therapeutic process.

Using the formula explained above, an attempt is made to reduce the individual changes to a common denominator and thus present the changes per patient in a graph (Figure 2). It should be noted that this can only be done to a limited

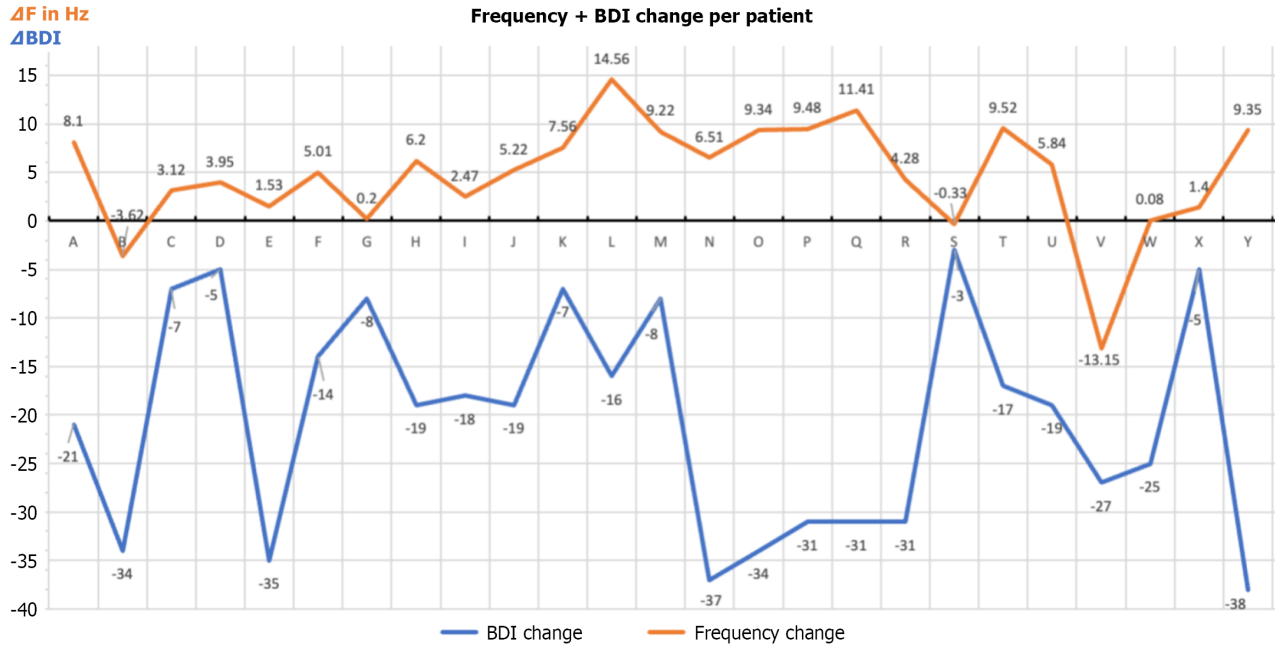

Figure 2 Mean baseline frequency and Beck depression inventory change per patient (*n* = 25; Patient A to Y). BDI: Beck depression inventory.

extent using the available figures, as both voice and progress in psychotherapy are always individual. For example, it is not known what the patients' original average fundamental frequency was before the original depression.

Even if a reduction in the BDI rating simultaneously results in an increase in the average fundamental frequency and thus the MIDI standard note number, this cannot be extrapolated over the entire spectrum. Based *e.g.*, on the value 0.33, this would mean an increase of 20 Hz for the maximum jump from 63 points to 0 points. This does not seem realistic; instead, a kind of saturation must be expected for the change of the mean fundamental frequency. This saturation cannot be verified from the available data. Figure 3 gives an overview of all patients separately.

Nevertheless, on average, the ratio per patient was 0.32 (ΔMN/ΔBDI point without the factor 10). If the calculation data is adjusted upwards and downwards by the two outliers, a ratio of 0.28 (ΔMN/ΔBDI point) remains. Conversely, in figures for the fundamental frequency, this means that a change in the BDI rating by 3 points downwards results in an increase in the fundamental frequency by one Hz. There can be no correlation between an increased frequency change or the ratio, and the therapy method applied.

The results show no increase in the baseline frequency for only three patients. A more detailed exploration of the background shows that for two patients external factors had an influence on the results (therapy in a foreign language and a stroke of fate). In the third patient, there was no change in the depression even after a long period of therapy and therefore no, or only a minimal, change in frequency.

In Table 2, a compact presentation of all specific and relevant data for the respective patients is coded again. This includes both the metadata and the collected and calculated data.

DISCUSSION

The overall assessment shows consistent results regarding the mean basic frequency within the circle of the initial interviews and the final sessions, using the method of analyzing long speech sequences without interruption. To further test the quality of the data, the interviews in the middle phase were also analyzed for three patients with the highest change in mean baseline frequency. These show the expected course and already an increase in the baseline frequency, although they do not yet reach the value of the end interviews. Around 30% of the patients showed a vocal effect in the first recorded therapy session and the extracted data did not match, or only partially matched, the other data from the initial interviews. This could be explained by excitement and nervousness, knowing that their problems were being recorded. Many patients also actively discussed the recording in the first recorded hour. On the other hand, the media effect would also be a model of explanation, in which the voice, at least with non-professional speakers, usually sounds somewhat altered at the beginning. In order to reduce this effect, an attempt was made to slowly adapt the speakers to the new situation and to focus on topics that are relevant[28].

This fact therefore has a partial influence on the results, but also shows that data generated in the laboratory or in experiments must be subjected to close examination and selection. Particular attention is paid to the provision of basic and learning data, which serve as the basis for the algorithms of artificial intelligence (AI) programs.

One strength of the study is the possibility of accessing the audio files of the MPS. They provide a free basis for the conversation in a therapeutic setting, and the patients were given the freedom to speak spontaneously. The data analysis is thus based on natural conversational utterances, the paraverbal attributes of which were formed without any

Table 2 Compact overview of the data per patient

ID	Method	Female	Therapy length in days	BDI start	BDI end	BDI change	Mean for start	Mean for end	Mean for change	MIDI note nr. start	MIDI note nr. end	MIDI note nr. difference	MIDI/BDI
A	CBT	1	170	21	0	-21	183.3	191.4	8.1	53.840	54.589	0.74861	0.36
B	POP	1	225	34	0	-34	239.18	235.56	-3.62	58.447	58.183	-0.26403	-0.08
C	POP	1	1045	22	15	-7	180.76	183.88	3.12	53.599	53.895	0.29627	0.42
D	CBT	1	336	16	11	-5	189.77	193.72	3.95	54.441	54.798	0.35665	0.71
E	POP	1	281	38	3	-35	182.96	184.49	1.53	53.808	53.952	0.14417	0.04
F	POP	1	350	20	6	-14	224.08	229.09	5.01	57.318	57.701	0.38281	0.27
G	PA	1	441	24	16	-8	220.9	221.1	0.2	57.071	57.086	0.01567	0.02
H	CBT	1	589	47	28	-19	178.1	184.3	6.2	53.342	53.935	0.59242	0.31
I	POP	0	181	18	0	-18	115.7	118.17	2.47	45.875	46.240	0.36570	0.20
J	CBT	1	721	32	13	-19	202.15	207.37	5.22	55.535	55.976	0.44137	0.23
K	POP	0	736	28	21	-7	98.2	105.76	7.56	43.035	44.319	1.28399	1.83
L	PA	1	1287	16	0	-16	159.91	174.47	14.56	51.477	52.986	1.50863	0.94
M	PA	1	1397	17	9	-8	207.32	216.54	9.22	55.972	56.726	0.75329	0.94
N	PA	1	972	37	0	-37	192.1	198.61	6.51	54.652	55.229	0.57697	0.16
O	POP	1	1087	36	2	-34	190.6	199.94	9.34	54.517	55.345	0.82823	0.24
P	PA	1	1015	38	7	-31	208.02	217.5	9.48	56.031	56.802	0.77152	0.25
Q	CBT	1	1035	33	2	-31	221.75	233.16	11.41	57.137	58.006	0.86863	0.28
R	PA	1	857	40	9	-31	194.4	198.68	4.28	54.858	55.235	0.37702	0.12
S	PA	0	938	30	27	-3	117.4	117.07	-0.33	46.127	46.078	-0.04873	-0.16
T	POP	1	854	45	28	-17	225.25	234.77	9.52	57.408	58.125	0.71665	0.42
U	PA	1	618	30	11	-19	218.94	224.78	5.84	56.916	57.372	0.45574	0.24
V	CBT	1	749	39	12	-27	218.54	205.39	-13.15	56.885	55.810	-1.07438	-0.40
W	POP	0	406	25	0	-25	95.95	96.03	0.08	42.634	42.649	0.01443	0.01
X	POP	0	397	25	20	-5	122.4	123.8	1.4	46.849	47.046	0.19689	0.39
Y	POP	1	1141	38	0	-38	190.05	199.4	9.35	54.467	55.298	0.83144	0.22

POP: Psychoanalytic oriented psychotherapy; CBT: Cognitive behavioral therapy; PA: Psychoanalysis; BDI: Beck depression inventory; MIDI: Musical instrument digital interface.

specifications.

Limiting factors

As a lot of information on the audio files of the MPS is not available, some limiting factors must be mentioned.

First, some limitations of external validity have to be mentioned in order to prevent false-positive risks. Only 4 cases (16%) represented mild depression, risking a masking of subgroup vocal signatures. Conversely, severe depression constituted 56% of the sample, biasing results toward this subgroup and limiting generalizability across the full depressive spectrum. There are also no records if and when any medication, not only psychopharmacological, was administered during psychotherapy. These could have had psychological, physiological or hormonal influences and effects at any time, which could have also influenced the voice and the fundamental frequency f_o. For the overall f_o elevation alongside treatment effects, the small sample size has to be taken into consideration.

Second, the recording conditions of the audio files and the type of technical equipment used to make the recordings. There is no specific information on the microphone and recording device, however it can be assumed that in most cases cassette recorders were used at the time. No attention was paid to the acoustic influences in the therapy rooms; windows were open and sometimes background noises from drilling machines or passing cars can be heard in the background. Furthermore, there is no information on where the recording device was placed in the room and at what distance the therapist and patient were situated. In fact, no experts were involved in generating this data; instead, the therapists were used as laypersons. It is also unknown, how the data was later converted and transferred onto compact disks. These

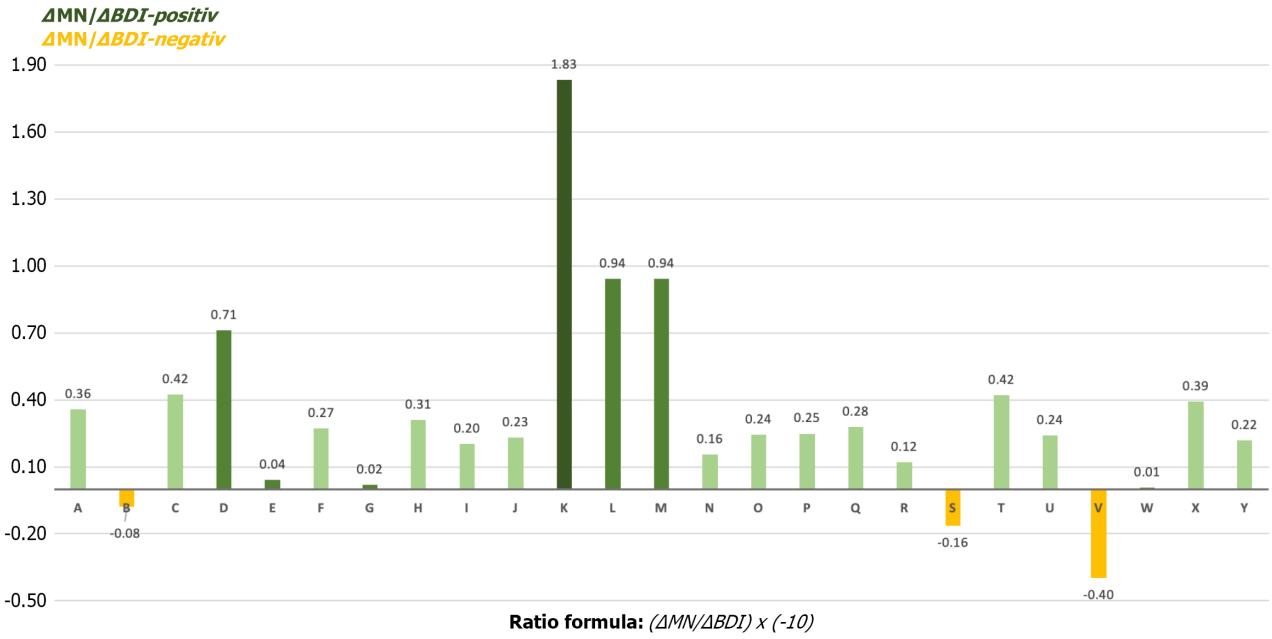

Figure 3 Summary of the calculated ratio per patient. MN: MIDI note number; BDI: Beck depression inventory.

unknown parameters should be mentioned as they may have a potential influence.

Furthermore, potential biases due to the small sample size (n = 25) and gender imbalance (80% female) have to be addressed. The small male cohort compromises statistical power for sex-stratified analyses, preventing validation of sex-specific f_o change patterns. Although the study aims to validate f_0 changes correlating with depression improvement, it fails to explicitly differentiate within-subject changes (pre/post treatment) from between-group differences (treatment modalities). Therefore, the reliability of the conclusion must be examined with caution. There is no doubt that the data material provided is not state of the art, but is characterized by technical and methodological pragmatism, as this was not the focus of the MPS.

It should also be mentioned that the times for the BDI ratings at the beginning and end of therapy do not coincide completely with the start and end interviews but are shifted by a few weeks.

General relevance

The previous sections have shown that voice is characterized by complexity and is subject to different influences. Voice is an individual characteristic of our personality and reveals our innermost self to the outside world, for example when emotions are at work and guide our utterances. People with depression are often characterized by a clouded or subdued mood, which has an influence on the physical and psychological whole and also has a physiological effect. The basic mood is reflected in the voice.

Although a wide range of research is being conducted in the context of prosodic characteristics and mental illness, it is primarily based on experimental data and examines short-term time periods. In addition, they primarily focus on faster and more reliable detection of depression than on vocal changes in therapeutic progress.

This research work can therefore open a new data field and provide findings in the area of real data analysis in relation to the mean fundamental frequency f_o. There is also sufficient data available to carry out a valid analysis. This analysis is positively enhanced by the fact that the same patients are followed over a long period of up to three years and the data from the depression ratings are available at the same time. A significant disadvantage is the quality of the audio recordings.

The research question, whether changes can be detected in voice parameters and especially in the fundamental frequency f_o cannot be clearly answered due to unknown parameters in the original data collection. The data obtained show a tendency that indicates a fundamental increase and stabilization of the fundamental frequency at and a simultaneous reduction in the depression rating. However, relevant information from the original study (medication, hormone status, technical information) is missing for a valid statement.

The discrepancy in the fundamental frequency between the data generated experimentally and the real data from the MPS audio files was striking. The frequency jumps are artificially generated many times higher (joy) and lower (sadness) than was visible in the patients in the MPS. However, it should also be noted here that the comparison cannot be made directly, as the patients in the MPS are assumed to have a subdued mood. Another criterion for generating data appears to be excitement or the media effect, which was evident in almost a third of the patients in the first recorded hour.

Since the basis of the voice and the fundamental frequency f_o is different for all people, cross-comparisons of the increase between patients are only of limited use.

Using prosodic characteristics as a comparative element in depressive patients to have an objective indicator can nevertheless be considered useful. As this study is a preliminary exploration, results should be validated in larger samples and in more rigorously controlled studies. Further studies are recommended and necessary, and the massive

production of valid data material is essential. The data material should be consistent and as close to reality as possible, as well as complying with current methodological and technical standards. Research must also be interdisciplinary and should at least include experts from the fields of psychotherapy, acoustics and computer science ensuring professional standards.

Psychiatry has also become aware of the voice as a feature, and psychiatrist Charles Marmar, among others, notes that the voice is enormously versatile in conveying our emotional signals[5]. Certain research has focused on the detection of mental illnesses such as depression, psychosis, burnout or bipolar disorder using voice analysis, usually achieving detection rates of between 80% and 95% in experiments[5]. A person's speech and voice can also serve as an important indicator for recognizing suicidal intentions[5].

In the context of the current state of research, there are both structural and methodological gaps in this field, that we tried to close within this study. In principle, there is little opportunity to meaningfully compare the different study results with one another. This is primarily because only isolated studies are available due to the different methodological approaches and variants[29]. The difficulties of various studies can now be seen above all in the high technical requirements and the complexity of processing the data. One of the biggest, if not the biggest, problem with using voice analysis to identify physical or mental illness is that the studies and research usually cannot access real data but are based on data sets that were produced in experimental and controlled environments and under specific conditions[5].

A recent study has collected several areas of research on emotional prosody from the last three decades and categorized evolution based on quality/spectral features, dynamic/contours and other features. However, the outcome of that study is also that certain parameters and definitions are missing, to perform comparisons and open the field to a more standardized one.

Therefore, recordings are typically performed in laboratory settings by invited actors or non-actors. It has been shown that the acoustics of play-acted (or posed) recordings differ from those of spontaneous recordings[30].

The steady increase in mental illness is leading to a practical need to invest heavily in this area of research, particularly on the topic the cost-effectiveness of national healthcare systems. Voice-based technologies and recognition open new possibilities for efficiency, quality and reduced costs. Automatic speech recognition, which not only focuses on detecting mental illness, but also to enable easier and faster documentation and thus also gives healthcare staff more time for patient contact[31].

Currently methodological standards for screening and diagnosis of depression, are the Patient Health Questionnaire (PHQ) or the BDI, which are designed as self-evaluation questionnaires. In everyday life people with depression can face hurdles making it difficult for them to seek professional help, as this requires taking an active step that many patients seem to be unable to take. For this reason, a low-threshold and simplified process is being sought that both reduces the initial hurdles and simplifies and shortens the identification process, without compromising the qualitative analysis. Methods that aim to develop an automated screening process for depression are based on complex algorithms that aim to offer simple and practicable solutions, for example *via* telephone calls[32]. Automated screening processes follow two different methods. On the one hand, answers to specific questions are evaluated, such as: Have you had depression in the past? And subjected to weighting. On the other hand, an open approach is chosen, and the original question is not necessarily included, but an attempt is made to identify depression using global and/or time-varying statistics[33].

Another research sector is currently emerging in the field of machine learning and supports the validation process. Here, spoken texts, including those of depressed people and a control group, are recorded and processed into visual spectrograms using specific models (Log-Mel Spectrogram-based Convolutional Neural Network), which form the basis of the learning data sets. These data sets are constantly expanded, refined and improved. The final idea is a quick examination of patients using audio excerpts recorded on a smartphone, for example. One study used defined texts in this way and was able to achieve an accuracy of 78% in the detection of depression using these texts[34]. In the generalization and prediction of PHQ9 data and in the evaluation of spontaneously spoken sentences, no general, validatable statements could be made[34]. Other approaches in the context of machine learning attempt to use the voice to find out at what episode level depression the person is currently experiencing, *i.e.*, whether the depression is mild or severe. The accuracy in initial studies is 60% and thus shows the fundamental potential of the voice as a biometric variable in depression[7]. However, researchers are struggling with fundamental problems because, as already mentioned, voice is a complex and fragile system that is influenced by several different factors when it occurs in the environment. For example, even background noise in a recording reduces the reliability of an analysis tool that detects depression from 94% to 75% [35].

Especially in view of the worldwide prevalence, research in this context is of great importance, not only for early detection, but also to ensure adequate treatment. The different languages and subtleties are problematic and a challenge in these cases, which can also have a vocal impact. There is also the question of data sets and how they can be made available to the machine, increasing the quality of detection.

AI and its algorithms will also play a relevant role in voice analysis in recognizing speech patterns and derive results from them.

In the domain of depression in connection with the characteristics, metrics and profiles of the voice, it is important to separate and differentiate between two approaches. On the one hand, there is the detection of depression based on voice patterns, parameters or analytical findings, which should be quick and uncomplicated, but also accurate and meaningful. Here, a relationship can only be established based on empirically established parameters, which, however, do not take the individuality of callers during initial contact into consideration.

On the other hand, there is the change in voice profiles in long-term treatment, *i.e.*, which parameters change over time and to what extent and can therefore indicate the progress in therapy and thus the stabilization of patients. Behind this is a non-transferable approach in which the specificity of the personal voice is given greater importance.

The software market now offers a wealth of products that are primarily used in the service sector and can recognize callers' emotions using voice profiles. These emotions are displayed to the service employees on the screen in real time and enable them to initiate the appropriate response, as they have learned to do in the relevant training courses[5]. One example product among many is the provider SPITCH, which promotes voice biometrics as the most reliable biometric authentication method and validates over 100 parameters against an original voice profile on a comparative basis. Other software-based products on the market that are based in the healthcare sector and work with specific voice characteristics include Beyond Verbal, Cogito, Corti and Winter Light Labs.

The ever-advancing possibilities of technologies in the domain of AI leads to the voice technology sector being one of the most promising sectors, which will be particularly significant for the healthcare systems of all countries in the world [36]. The aim of using such technologies is not to replace doctors, but to provide them with an additional analysis tool, although this must be accompanied by specific training on how to use and interpret the results. In this context, the voice is also said to have a similar ability to detect diseases as blood currently does.

The latest and most recent development took place in March 2024, when OpenAI (ChatGPT) presented a program called Voice Engine, which can duplicate the voice of a human being from a 15-second original[37]. This is intended to demonstrate the capabilities and possibilities offered by AI, but also to point out how easily and quickly it is technically feasible to duplicate, but also to manipulate, an individual human characteristic. Beyond the technical possibilities, key elements for future applications need to be clarified. This concerns ethical issues as well as the handling of the data produced and how long it is stored.

Outlook for further research

For the field of psychotherapy, further aspects need to be considered for future research and incorporated into the research process. Psychotherapy usually takes place in a setting between two people. They are in a stable relationship with each other, a dyad, which is characterized by "reciprocal and interrelated patterns of action between the partners" [38]. The patient's voice works within this dyad during therapy and must therefore also be considered in this context. The variable of the therapist must be included, or at least taken into consideration, as they can have an influencing effect through the nature of the relationship or the fit. However, voice is also actively perceived by patients as a tool that can be used to regulate the relationship with the therapist. Changes are thus perceived *via* the voice and partial synchronization of prosody also takes place[39].

CONCLUSION

In conclusion, it can be said that voice is an extremely complex phenomenon that we cannot escape in our daily communication. Through our voice the unconscious comes to the surface. In terms of psychotherapy research, the inclusion of the voice as a factor to be investigated is still in its infancy. What effect and influence it has and can have in the therapeutic process and how a positive course of therapy can be traced from other prosodic characteristics that are to be specified for the respective patients.

If the focus is placed on language as a medium for the therapeutic process and the effectiveness of psychotherapy[40], voice and prosodic characteristics cannot be detached from this. The possibility of a linguistic turn in psychotherapy research is conceivable, *i.e.*, a more intensive focus on language and, linked to this, a prosodic turn.

REFERENCES

1 **Beblo T**, Dehn LB. Neuropsychologie der Depression. 2023 [DOI: 10.1026/03188-000] [FullText]

2 **Löffler-Stastka H**, Lehofer M, Rados S. Depressive Erkrankungen. Verlauf und Prognose. In: Nowotny M, Kern D, Breyer E, Bengough T, Griebler R, editors. Depressionsbericht Österreich. Eine interdisziplinäre und multiperspektivische Bestandsaufnahme. Wien: Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (BMASGK), 2019

3 **Löffler-Stastka H**, Kasper S. Erklärungsmodelle. Integrative Ansätze. In: Nowotny M, Kern D, Breyer E, Bengough T, Griebler R, editors. Depressionsbericht Österreich. Eine interdisziplinäre und multiperspektivische Bestandsaufnahme. Wien: Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz (BMASGK), 2019

4 **Kiese-Himmel C**. Phänomenologie der Stimme. Körperinstrument Stimme. 2016 [DOI: 10.1007/978-3-662-49648-0_4] [FullText]

5 **Karaboga M**, Frei N, Ebbers F, Rovelli S, Friedewald M, Runge G. Automatisierte Erkennung von Stimme, Sprache und Gesicht. 2022 [DOI: 10.3218/4141-5] [FullText]

6 **Hofmann J**, Angenehme Stimmen im Radio. Eine Analyse stimmlicher Kriterien. Master's thesis at the University of Regensburg. Center for Language and Communication, 2008. Available from: https://www.uni-regensburg.de/assets/zentrum-sprache-kommunikation/dokumente/mkuse/Masterarbeit_Jessica_Hofmann_-_Angenehme_Stimmen_im_Radio.pdf

7 **Shin D**, Cho WI, Park CHK, Rhee SJ, Kim MJ, Lee H, Kim NS, Ahn YM. Detection of Minor and Major Depression through Voice as a Biomarker Using Machine Learning. *J Clin Med* 2021; **10**: 3046 [*RCA*] [PMID: 34300212 DOI: 10.3390/jcm10143046] [FullText] [Full Text (PDF)]

8 **Silva WJ**, Lopes L, Galdino MKC, Almeida AA. Voice Acoustic Parameters as Predictors of Depression. *J Voice* 2024; **38**: 77-85 [*RCA*] [PMID: 34353686 DOI: 10.1016/j.jvoice.2021.06.018] [FullText]

9 **Menne F**, Dörr F, Schräder J, Tröger J, Habel U, König A, Wagels L. The voice of depression: speech features as biomarkers for major depressive disorder. *BMC Psychiatry* 2024; **24**: 794 [*RCA*] [PMID: 39533239 DOI: 10.1186/s12888-024-06253-6] [FullText]

10 **Habermann G**. Stimme und Sprache. Eine Einführung in die Physiologie und Hygiene für Ärzte, Sänger,Pädagogen und andere Sprechberufe

(4th, unchanged ed.). Stuttgart, New York: Thieme, 2003 [DOI: 10.1055/b-002-44921] [FullText]
11 **Friedrich G**, Bigenzahn W, Zorowka P. Phoniatrie und Pädaudiologie. Einführung in die medizinischen, psychologischen und linguistischen Grundlagen von Stimme, Sprache und Gehör (5th, completely revised ed.). Bern, Göttingen, Toronto, Seattle: Huber, 2012
12 **Helfrich H**. Satzmelodie und Sprachwahrnehmung. Berlin: De Gruyter, 1985 [DOI: 10.1515/9783110865752] [FullText]
13 **Wilde M**. Stimme und Transidentität: über die Bedeutung der Stimme-Stimmangleichung und Stimmtherapie für trans* Menschen. Germany: Schulz-Kirchner Verlag, 2018
14 **Dickreiter M**, Hoeg W. Grundlagen der Akustik. In: Dickreiter M, Dittel V, Hoeg W, Wöhr M, editors. Handbuch der Tonstudiotechnik. Berlin, Boston: De Gruyter Saur, 2023: 1-66 [DOI: 10.1515/9783110759921-001] [FullText]
15 **Kannengieser S**. Sprachentwicklungsstörungen: Grundlagen, Diagnostik und Therapie. Germany: Elsevier, Urban & Fischer, 2012 [DOI: 10.1016/C2021-0-01444-9] [FullText]
16 **Titze IR**, Baken RJ, Bozeman KW, Granqvist S, Henrich N, Herbst CT, Howard DM, Hunter EJ, Kaelin D, Kent RD, Kreiman J, Kob M, Löfqvist A, McCoy S, Miller DG, Noé H, Scherer RC, Smith JR, Story BH, Švec JG, Ternström S, Wolfe J. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J Acoust Soc Am* 2015; **137**: 3005-3007 [*RCA*] [PMID: 25994732 DOI: 10.1121/1.4919349] [FullText]
17 **Stassen HH**. Affekt und Sprache. Monographien aus dem Gesamtgebiete der Psychiatrie. Heidelberg: Springer Berlin, 1995 [DOI: 10.1007/978-3-642-79726-2] [FullText]
18 **Schmiedel A**. Phonetik ironischer Sprechweise: Produktion und Perzeption sarkastisch ironischer und freundlich ironischer Äußerungen. Germany: Frank & Timme Verlag für wissenschaftliche Literatur, 2017
19 **Spang W**. Qualität im Radio: Determinanten der Qualitätsdiskussion im öffentlich-rechtlichen Hörfunk in Deutschland. Germany: Röhrig Universitätsverlag, 2006
20 **Möbius B**. Ein quantitatives Modell der deutschen Intonation. Berlin: De Gruyter, 1993 [DOI: 10.1515/9783111355870] [FullText]
21 **Huber D**, Henrich G, Klug G. Moderators of change in psychoanalytic, psychodynamic, and cognitive-behavioral therapy. *J Am Psychoanal Assoc* 2013; **61**: 585-589 [*RCA*] [PMID: 23720030 DOI: 10.1177/0003065113490638] [FullText]
22 **Huber D**, Klug G. Münchner Psychotherapiestudie. *Psychotherapeut* 2016; **61**: 462-467 [*RCA*] [DOI: 10.1007/s00278-016-0139-7] [FullText]
23 **Beck AT**, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; **4**: 561-571 [*RCA*] [PMID: 13688369 DOI: 10.1001/archpsyc.1961.01710120031004] [FullText]
24 **Maryn Y**, Zarowski A. Calibration of Clinical Audio Recording and Analysis Systems for Sound Intensity Measurement. *Am J Speech Lang Pathol* 2015; **24**: 608-618 [*RCA*] [PMID: 26161588 DOI: 10.1044/2015_AJSLP-14-0082] [FullText]
25 **Mayer J**. Phonetische Analysen mit Praat Ein Handbuch für Ein- und Umsteiger 2022. [cited 29 August 2025]. Available from: https://praatpfanne.lingphon.net/downloads/praat_manual.pdf
26 **Huber S**. Untersuchung verschiedener Verfahren zur Grundfrequenzbestimmung mit Einstellung einer Applikation zur Midi-Konvertierung. Germany: Diplom.de, 2004
27 **Nakra TM**. Constructing Music. 2024 [DOI: 10.1093/oso/9780197669198.001.0001] [FullText]
28 **Eckert H**, Laver J. Menschen und ihre Stimmen. Aspekte der vokalen Kommunikation. Weinheim: Beltz, Psychologie Verl, 1994
29 **Stang P**, Schleider K. Stimmliche Merkmale depressiver Störungen. Aktueller Forschungsstand. DGSS @current 2018. [cited 21 April 2024]. Available from: https://www.dgss.de/fileadmin/user_upload/Download/Service/Bibliothek/dgssaktuell/dgssaktuell_2018-3.pdf
30 **Larrouy-Maestri P**, Poeppel D, Pell MD. The Sound of Emotional Prosody: Nearly 3 Decades of Research and Future Directions. *Perspect Psychol Sci* 2025; **20**: 623-638 [*RCA*] [PMID: 38232303 DOI: 10.1177/17456916231217722] [FullText] [Full Text(PDF)]
31 **Latif S**, Qadir J, Qayyum A, Usama M, Younis S. Speech Technology for Healthcare: Opportunities, Challenges, and State of the Art. *IEEE Rev Biomed Eng* 2021; **14**: 342-356 [*RCA*] [PMID: 32746367 DOI: 10.1109/RBME.2020.3006860] [FullText]
32 **Tasnim M**, Stroulia E. Detecting Depression from Voice. In: Meurs MJ, Rudzicz F, editors. Advances in Artificial Intelligence. Canadian AI 2019. Lecture Notes in Computer Science. Cham: Springer, 2019: 11489 [DOI: 10.1007/978-3-030-18305-9_47] [FullText]
33 **Al Hanai T**, Ghassemi M, Glass J. Detecting Depression with Audio/Text Sequence Modeling of Interviews. Interspeech 2018. 2018: 1716-1720 [DOI: 10.21437/interspeech.2018-2522] [FullText]
34 **Kim AY**, Jang EH, Lee SH, Choi KY, Park JG, Shin HC. Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach. *J Med Internet Res* 2023; **25**: e34474 [*RCA*] [PMID: 36696160 DOI: 10.2196/34474] [FullText]
35 **Anthes E**. Sprachanalyse als Schnelltest In: Gehirn & Geist 5/2021. Germany: Spektrum der Wissenschaft, 2021
36 **Fagherazzi G**, Fischer A, Ismael M, Despotovic V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digit Biomark* 2021; **5**: 78-88 [*RCA*] [PMID: 34056518 DOI: 10.1159/000515346] [FullText]
37 **OpenAI**. Navigating the Challenges and Opportunities of Synthetic Voices. [cited 2 May 2024]. Available from: https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices
38 **Herzberg PY**. [Dyadic assessment and dyadic data analysis]. *Psychother Psychosom Med Psychol* 2011; **61**: 383 [*RCA*] [PMID: 21826631 DOI: 10.1055/s-0031-1276857] [FullText]
39 **Bauer S**. Die Bedeutung von Stimme und Stimmklang im psychotherapeutischen Prozess aus Sicht der Patienten und Patientinnen. *Psychoanalyse Körper* 2010; **9**: 27-50
40 **Gumz A**, Spranz-Fogasy T. Sprachliche Charakteristika einer Psychotherapiesitzungssequenz aus der Perspektive unterschiedlicher Analysemethoden. *Psychotherapeut* 2022; **67**: 150-157 [*RCA*] [DOI: 10.1007/s00278-021-00568-5] [FullText]

FOOTNOTES

Specialty type: Psychiatry

Country of origin: Austria

Author contributions: Holzweber C, Kob M and Löffler-Stastka H participated in the conception and design of the study and were

involved in the acquisition, analysis, or interpretation of data; Hennenberg B was in charge of text structure and analysis; Holzweber C wrote the manuscript; Holzweber C, Löffler-Stastka H and Kob M accessed and verified the study data; Stastka L edited, translated and reviewed the manuscript. All authors critically reviewed and provided final approval of the manuscript; all authors were responsible for the decision to submit the manuscript for publication.

Institutional review board statement: This investigation was approved by the Institutional Ethics Committee of the Medical University Vienna, No. 1746/2023.

Informed consent statement: The need for patient consent was waived due to the retrospective nature of the study.

Conflict-of-interest statement: All authors declared that they have no conflict of interest.

Data sharing statement: Additional data are available on request from the corresponding author.

Corresponding Author's Membership in Professional Societies: OEGPP; IPA.

S-Editor: Qu XL
L-Editor: A
P-Editor: Wang WB