Original Paper

Psychopathology

Psychopathology
DOI: 10.1159/000363247

Received: January 13, 2014
Accepted after revision: April 18, 2014
Published online: September 16, 2014

Affective State and Voice: Cross-Cultural Assessment of Speaking Behavior and Voice Sound Characteristics – a Normative Multicenter Study of 577 + 36 Healthy Subjects

Silke Braun[a] Cristina Botella[d] René Bridler[a] Florian Chmetz[c]
Juan Pablo Delfino[a] Daniela Herzig[c] Viktoria J. Kluckner[a] Christine Mohr[c]
Ines Moragrega[d] Yann Schrag[c] Erich Seifritz[b] Carla Soler[d] Hans H. Stassen[a]

[a]Institute for Response-Genetics, University of Zurich, and [b]Psychiatric University Hospital, Zurich, and
[c]Institute of Psychology, University of Lausanne, Lausanne, Switzerland; [d]Clinical Psychology, University of Jaume I, Castellon, Spain

Key Words
Affect · Affective disorders · Speech analysis · Gender · Age · Education · Native language · Self-assessments

Abstract
Background: Human speech is greatly influenced by the speakers' affective state, such as sadness, happiness, grief, guilt, fear, anger, aggression, faintheartedness, shame, sexual arousal, love, amongst others. Attentive listeners discover a lot about the affective state of their dialog partners with no great effort, and without having to talk about it explicitly during a conversation or on the phone. On the other hand, speech dysfunctions, such as slow, delayed or monotonous speech, are prominent features of affective disorders. ***Methods:*** This project was comprised of four studies with healthy volunteers from Bristol (English: n = 117), Lausanne (French: n = 128), Zurich (German: n = 208), and Valencia (Spanish: n = 124). All samples were stratified according to gender, age, and education. The specific study design with different types of spoken text along with repeated assessments at 14-day intervals allowed us to estimate the 'natural' variation of speech parameters over time, and to analyze the sensitivity of speech parameters with respect to form and content of spoken text. Additionally, our project included a longitudinal self-assessment study with university students from Zurich (n = 18) and unemployed adults from Valencia (n = 18) in order to test the feasibility of the speech analysis method in home environments. ***Results:*** The normative data showed that speaking behavior and voice sound characteristics can be quantified in a reproducible and language-independent way. The high resolution of the method was verified by a computerized assignment of speech parameter patterns to languages at a success rate of 90%, while the correct assignment to texts was 70%. In the longitudinal self-assessment study we calculated individual 'baselines' for each test person along with deviations thereof. The significance of such deviations was assessed through the normative reference data. ***Conclusions:*** Our data provided gender-, age-, and language-specific thresholds that allow one to reliably distin-

KARGER

E-Mail karger@karger.com
www.karger.com/psp

© 2014 S. Karger AG, Basel
0254–4962/14/0000–0000$39.50/0

Hans H. Stassen
Institute for Response-Genetics
Psychiatric University Hospital
CH–8032 Zurich (Switzerland)
E-Mail k454910@bli.uzh.ch

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

guish between 'natural fluctuations' and 'significant changes'. The longitudinal self-assessment study with repeated assessments at 1-day intervals over 14 days demonstrated the feasibility and efficiency of the speech analysis method in home environments, thus clearing the way to a broader range of applications in psychiatry. © 2014 S. Karger AG, Basel

Background

Our interest in speaking behavior and voice sound characteristics has a psychiatric background with focus on affect disturbances, which are constituents of most major psychiatric disorders. Human speech is greatly influenced by the speakers' affective state, such as sadness, happiness, grief, guilt, fear, anger, aggression, shame, sexual arousal, love, amongst others. Attentive listeners discover a lot about the affective state of their dialog partners with no great effort, and without having to talk about it explicitly during a conversation or on the phone. On the other hand, speech dysfunctions, such as slow, delayed or monotonous speech, are prominent features of affective and schizophrenic disorders: 'The patients speak in a low voice, slowly, hesitatingly, monotonously, sometimes stuttering, whispering; try several times before they bring out a word; and become mute in the middle of a sentence. They become silent, monosyllabic, can no longer converse' [1].

Clinicians frequently observe that the speech of depressed patients is uniform and sometimes exhibits a regular repetition of gliding intervals and that the pitch alterations of these patients are narrowed, giving the voice a monotonous quality. In fact, depression significantly reduces the dynamic expressiveness of human voices, thus greatly reducing interindividual differences. As a direct consequence, the patients' voices become more similar to each other ('depressive voice'). During recovery, however, the patients' speaking behavior and voice sound characteristics return to 'normal' values. On the group level this can readily be demonstrated under various experimental settings [e.g., 1–7]. Yet things get much more complex if focus is laid on longitudinal changes in the individual patient over time (single-case analysis), for example, when clinicians monitor speaking behavior and voice sound characteristics among affectively disturbed patients for diagnostic purposes and as indicators of clinical change (e.g., among patients recovering from depression, or among patients at risk of relapse after having recovered from depression).

Speaking behavior and voice sound characteristics encompass a dominating, genetically determined 'static component' superimposed by a much smaller 'dynamic component' that reflects reactions to and interactions with the immediate environment, as well as the speaker's emotional and affective[1] state. Deviations from 'normality' can persist over seconds, minutes, hours, or even days where psychiatric issues may come into play. The complexity of single-case analysis with focus on affective disorders originates from (1) the between-subject heterogeneity of speaking behavior and voice sound characteristics which enables, according to everyday experience, an easy and reliable recognition of persons by their voices: even simple automatic recognition procedures yield rates of uniquely identified speakers after 14 days in the range of 90–93% [10, 11] while the human ear features rates close to 100%; (2) the fact that the 'dynamic component', which transports the speaker's emotional and affective state, is relatively small compared to the 'static component' and accounts for no more than 20% of the total information contained in human speech, and (3) the fact that affective disorders are characterized by heterogeneous symptom patterns which affect speaking behavior and voice sound characteristics in a variety of different ways.

Several methodological problems have to be solved within the scope of single-case analysis. Spoken languages are characterized by rhythm, stress, and intonation of speech ('prosody'). Besides the speaker's emotional state, prosody also reflects the content of an utterance[2] mostly with focus on the comprehension of spoken language (statements, questions, commands, or multiple interpretations of sentences [12]). Languages can be classified according to the distinctive prosodic units that constitute the languages' specific rhythm and sound made up by the

[1] The terms 'affect' and 'emotion' are often used synonymously. We understand affects as elementary, evolutionarily old processes deep inside the human body, such as aggression, fear, anger, sadness, grief, or sexual arousal, which can be triggered by a multitude of endogenous or exogenous events, proceed in a largely uncontrollable way, while being accompanied by distinct bodily reactions, such as sweat, rapidly increased blood pressure or heart rate, dizziness, amongst others. Affects are communicated to the outside world through emotions, which serve as an interface between the organism and the outside world [8, 9]. This interface works in both directions, from inside to outside and from outside to inside, often influenced by cognitive biases. Social skills may even allow one to not communicate certain affective reactions through emotions. In extreme cases the interface between the organism and the outside world can be completely 'blocked': the patient can no longer communicate and cannot be 'reached' by therapists, though the effects of heavy processes deep inside the patient can nonetheless be perceived by experienced clinicians. The term 'mood' describes the quality of feeling at a particular time.

[2] Under 'utterances' we understand those pieces of a sentence that are produced as an entity and separated from each other by pauses.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

mode of pronunciation, pitch or tone, emphasis patterns, and intonation ('accent'). Linguists classify languages as being either 'stress-timed' with highly complex syllables, 'syllable-timed' with less complex syllables, or 'mora-timed' with relatively simple syllables [13]. Stress-timed languages include English and German, and syllable-timed languages French, Italian, and Spanish. Additional modifying factors are gender, age, and educational level. In fact, female speakers display, on average, a mean vocal pitch of 220 Hz, which is exactly 1 octave above the average mean vocal pitch of male speakers (110 Hz). With age, the average man and woman loses muscle mass, mucous membranes become dryer, while some of the fine coordination is lost. Such changes can occur in the larynx as well, thus altering voice sound characteristics in terms of pitch, volume, and tremor in a characteristic way [14]. Even elephants can determine ethnicity, gender, and age from acoustic cues in human voices [15]. Education can influence speech flow and intonation (a low educational level can mean deficiency in speaking skills; reduced fluency; speakers need more time to present a given text) but is a minor issue in Europe.

Given these methodological difficulties, single-case approaches to quantifying a speaker's affective state through analysis of prosodic elements must necessarily rely on a standardized speech production procedure. Language-, gender-, age-, and education-specific criteria are required in order to distinguish between 'natural fluctuations' and 'significant changes' of clinical relevance (for methodological reasons, we cannot detect effects related to affect disturbances which are smaller than 'natural' fluctuations). Based on these prerequisites clinical studies demonstrated the efficiency of the speech analysis method as to assessing the time course of improvement under antidepressants in an objective and reproducible way [e.g., 16, 17]. In 65% of patients, single-case analyses revealed a close correlation over time ($r \approx 0.8$) between the Hamilton Depression (HAM-D) score and speech parameters. Patients not showing such close correlations displayed either no clinical change or an irregular pattern of nonimprovement.

So far, the speech analysis method has been carried out mainly in acoustically shielded high-tech speech laboratories with the aim of monitoring the transition from 'affectively disturbed' to 'normal' among psychiatric patients under treatment. By contrast, in this study we aimed at developing a low-cost, universally usable self-assessment procedure that can detect the transition from 'normal' to 'affectively disturbed' among subjects of the general population. Given the high lifetime prevalence of major depressive disorders in the range of 4–8% along with the immense burden caused by this illness, subjects in the very beginning of developing affective disorders might benefit from early intervention before psychiatric symptoms develop and may reach clinically relevant thresholds. Specifically, our study addressed the following questions: (1) can the prosodic elements 'distribution of pauses and utterances', 'stress', and 'intonation patterns' be quantified in a language-independent way and at a sufficiently high resolution; (2) the extent to which the factors 'gender', 'age', and 'education' modify prosodic elements; (3) the extent to which the thresholds between 'natural fluctuations' and 'significant changes' depend on language, and (4) can self-assessments be realized in typical home environments.

Methods

Speech production is the result of a joint effort of mind and body. It involves a cascade of steps from utterance planning to final sound production with hundreds of degrees of freedom. Rhythm, stress, and intonation ('prosody') greatly influence the verbal and nonverbal content of the transmitted speech. Despite this complexity, speech characteristics can be roughly described by a few major features. Speaking behavior can be modeled in terms of 'speech flow', 'loudness', and 'intonation', while voice sound characteristics relate to the distribution and intensity of 'overtones' that make up the speakers' individual voice 'timbres'. Speech flow describes the speed at which utterances are produced as well as the number and duration of temporary breaks in speaking. Loudness reflects the amount of energy associated with the articulation of utterances and, when regarded as a time-varying quantity, the speaker's dynamic expressiveness. Intonation is the manner of producing utterances with respect to rise and fall in pitch, and leads to tonal shifts in either direction of the speaker's mean vocal pitch. Overtones are the higher tones which faintly accompany a fundamental tone, thus being responsible for the tonal diversity of sounds. These overtone patterns display large interindividual differences and enable a computerized identification of persons through their voices. On the other hand, affect disturbances modify a subject's overtone pattern in a characteristic way.

Our approach to quantifying speaking behavior and voice sound characteristics relies on 'standard texts' specifically selected for grammatical simplicity: (1) 'automatic speech' (counting out loud); (2) 'emotionally neutral speech' (reading out loud a 2-min emotionally neutral passage of a children's book), and (3) 'emotionally stimulated speech' (reading out loud a 2-min emotionally stimulating passage from a famous novel)[3]. For the population-based normative studies we used all 3 texts for a 5-min recording. The 3 types of spoken text along with repeated assess-

[3] The standard texts are available in five languages fromhttp://www.bli.uzh.ch/vox15o.php.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

ments at 14-day intervals allowed us to estimate the 'natural' variation of speech parameters over time ('normative data') and to analyze the sensitivity of speech parameters as to resolving subtle differences between the spoken texts. It is important to note that emphasis lies on resolving subtle text differences rather than on quantifying the 'emotionality' of texts through one of the psychological or psycholinguistic models found in the literature because insufficient resolution of all speech parameters and combinations thereof would greatly limit the applicability of the method.

Once normative data on natural fluctuations and the sensitivity of speech parameters are available, longitudinal studies can rely on a single 2-min text only. In the longitudinal studies involving single-case analyses, we therefore used 'automatic speech' and 'emotionally neutral speech' for a 3-min recording, where each person served as his/her own reference ('baseline') when assessing changes in speaking behavior and voice sound characteristics over time. The 'automatic speech' part in the very beginning of each assessment of our longitudinal studies served as 'warming up' phase and let the speakers 'relax' when starting to speak into the microphone.

A critically important prerequisite for speech parameter extraction is the reliable subdivision of speech recordings into pauses and utterances ('segmentation'). In a first step, our segmentation algorithm screens each individual speech recording for a certain number of intervals without signal. These intervals are then used to determine the thresholds for background noise under consideration of a certain 'guard' zone. Secondly, nonlinear amplifiers in combination with limiters are applied to the speech recording (absolute amplitudes) in such a way that low amplitudes are reduced while amplitudes above a certain threshold are amplified/clipped. The resulting curves are then smoothed by a 16-fold moving average function to yield the segmentation 'basis'. Finally, segmentation 'basis' and background noise thresholds allow one to subdivide time series into pauses and utterances.

'Energy' (loudness) is calculated by summing up the squared amplitudes within utterances. We distinguish between 'energy per utterance' and 'energy per second' (zero energy within pauses by definition). The variation of 'energy per second' throughout a speech entity is used as a measure of the speaker's dynamic expressiveness ('dynamics', variation of loudness).

Once segmentation has been completed, 'spectra' can be calculated by means of discrete Fourier transformation (DFT) of 'pure' utterances with pauses being skipped. We rely on a tonal approach with a quartertone resolution covering 7 octaves in the frequency range of 64–8,192 Hz, so that spectra are comprised of 168 equally spaced quartertones. The tonal approach was chosen because pitch (perceptual quantity) depends logarithmically on frequency (physical quantity). Due to this approach, quartertones (and octaves) are equally spaced on the x-axis, so that rise and fall in pitch can be modeled as linear shifts along this axis.

Voice sound characteristics are determined from a sequence of consecutive spectra: due to intonation, the frequency associated with the maximum spectral line in the first octave of the speaker's spectral distribution displays some variation around a 'rest' position 'F0' (mean vocal pitch). 'F0 variation' is the variation of F0 throughout a speech entity (intonation). Additional aspects of intonation are quantified through (1) the interval defined as symmetrical points around F0 where the amplitudes of the F0 distribution drop by 6 dB ('F0 6-dB bandwidth') and (2) the second-degree polynomial approximation of the F0 distribution in the spectral domain ('F0 contour'). 'F0 amplitude' is computed from consecutive spectra as mean length of the spectral lines at frequency F0 along with the amplitudes of the higher harmonics, which make up the speaker's individual voice (timbre).

Materials

This project was comprised of four studies following the same experimental design and carried out with healthy volunteers in Bristol (English: n = 117), Lausanne (French: n = 128), Zurich (German: n = 208), and Valencia (Spanish: n = 124). The chosen study sites with two stress-timed languages (English, German) and two syllable-timed languages (French, Spanish) allowed for comparisons within and between stress-timed and syllable-timed languages. Samples were stratified according to gender, age (four age classes: 18–30, 31–40, 41–50, and 51–65 years), and education (four categories: remedial, junior high, high, and college). The test persons were asked to fill out the 63-item Zurich Health Questionnaire, which assesses 'regular exercise', 'consumption behavior', 'impaired physical health', 'psychosomatic disturbances', and 'impaired mental health' (available in five languages from http://www.bli.uzh.ch/Left07b.php). Thus, subjects with mental health problems could be excluded from our normative studies.

Test persons were invited to present the three types of text twice at 14-day intervals and at a fixed time in the morning according to the following scheme: (1) counting out loud from 1 to 40; (2) reading out loud the emotionally neutral text; (3) reading out loud the emotionally stimulating text, and (4) counting out loud again from 1 to 40. The entire recording procedure took 10 min including individual volume calibration. All speech recordings were carried out in acoustically shielded rooms, using high-end Sennheiser MKH40 P48 microphones specifically selected for linear frequency response, along with A/D converters featuring 0.1 dB linearity.

In addition to the normative studies, we carried out a longitudinal self-assessment study in order to test the feasibility of the speech analysis method in home environments. The two study groups were comprised of test persons under chronic stress (students with tight schedules and frequent exams; unemployed adults) among whom short-term affective reactions whatsoever were more likely than among test persons randomly selected from the general population. Specifically, test persons were recruited from university students in Zurich (n = 18, German: stress-timed language), and from unemployed adults in Valencia (n = 18, Spanish: syllable-timed language). The test persons were asked to fill out the 28-item Coping Strategies Inventory (COPE) along with the 63-item Zurich Health Questionnaire prior to enrolling in the study (questionnaires available in five languages from http://www.bli.uzh.ch/Left07b.php). Thus, we were able to exclude subjects with mental health problems, and to assess each test person's coping behavior under chronic stress.

The test persons received a low-cost microphone with linear frequency response characteristics (Behringer Studio Condenser C-1) in combination with a netbook, were instructed of how to use the equipment, and were asked to perform voice recordings every afternoon in their home environment over a period of 14 days. The speech recordings were done according to the following

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

scheme: (1) counting out loud from 1 to 40; (2) reading out loud the emotionally neutral text, and (3) counting out loud again from 1 to 40. The entire recording procedure took less than 5 min per day.

All speech signals were recorded in digital mode with a sampling rate of 96 kHz and at a 16-bit resolution. Subsequent data analyses were carried out by means of the program package Master.Vox[4] in combination with the statistics package SAS 9.3 on a 64-bit Windows system. In detail, we determined (1) the distribution of speech parameters in the general population; (2) the intraindividual stability of speech parameters over 14 days in order to draw a line between 'natural' fluctuations and 'significant' changes; (3) the differences between affect-neutral and affect-charged speech; (4) the amount of variance explained by the factors gender, age, education, and spoken language, and (5) longitudinal profiles from 10 to 14 repeated self-assessments at 1-day intervals. The studies were approved by the local Ethics Committees.

Table 1. Normative Speech Study Zurich (n = 577)

	Bristol (English) (n = 117)	Zurich (German) (n = 208)	Lausanne (French) (n = 128)	Valencia (Spanish) (n = 124)
Gender				
Male	45 (38.5%)	90 (44.3%)	39 (33.1%)	59 (47.6%)
Female	72 (61.5%)	113 (55.7%)	79 (66.9%)	65 (52.4%)
Age, years				
18–35	64 (54.7%)	107 (52.7%)	82 (69.5%)	55 (44.4%)
36–64	53 (45.3%)	96 (47.3%)	36 (30.5%)	69 (55.6%)
Education				
Basic	96 (82.1%)	167 (82.3%)	57 (48.3%)	90 (72.6%)
College	21 (17.9%)	36 (17.7%)	61 (51.7%)	34 (27.4%)

Sample composition of the four-center Normative Speech Study Zurich (encompassing the stress-timed languages English and German, along with the syllable-timed languages French and Spanish) with respect to gender, age, and education.

Results

At the participating study sites, recruitment of test persons was initiated through a short advertising campaign in the local media and by posting invitations at public places. During enrollment, the study administrators favored certain test persons over others in order to meet the study's stratification goals as to the sample composition with respect to gender, age, and education. Unexpectedly, females were more likely to participate in the study so that some imbalance in this respect could not be avoided. Also, we got a preponderance of test persons with college degree in one sample (table 1).

The digitized speech signals of one single assessment of our study typically required 120 MB storage space, adding up to no more than 120 GB of raw data for the entire study, which could easily be stored in a databank. Visual inspection of speech signals was carried out in order to mark time series with an artifact code where necessary (fig. 1a). As language is a critical factor in automatic segmentation, our standard algorithm was specifically 'trained' in order to iteratively optimize its free parameters in a language-specific way: (1) we randomly selected 40 recordings (20 males, 20 females) from each language; (2) these recordings were segmented manually and served as reference during iterative optimization; (3) iterative optimization aimed at minimizing the sum of the squared deviations between manual and automatic segmentation marks under consideration of the background noise level in each individual case (fig. 1b). This kind of optimization worked surprisingly well with deviations between manual and automatic segmentation marks becoming virtually zero, even for relatively noisy recordings (<5% across all signals including intervals with artifactual segments).

DFTs were used to quantify the tonal components that constitute the sound of a speech signal. Extensive tests with the empirical data suggested an optimal DFT epoch length of 1 s. The resulting 'spectra' show the frequency range of 64–8,192 Hz at a quartertone resolution along the x-axis[5], with the intensities of these 168 quartertones being displayed along the y-axis on log-proportional scales (fig. 2a, b). Interestingly, the mean vocal pitch of female speakers (220 Hz) lies, on average, approximately 1 octave above that of male speakers (110 Hz). This octave shift of the spectrum as a whole entity is illustrated by figures 2a, b.

Population-Based Normative Studies

We used the five parameters 'pause duration', 'utterance duration', 'energy per utterance' (loudness), 'dynamics' (variation of loudness), and 'energy per second' for the assessment of speaking behavior, and another five parameters, 'mean vocal pitch F0', 'F0 amplitude',

[4] The program package 'Master.Vox' used for voice analysis (27 programs) runs under Solaris, Linux, and Microsoft Windows (Android version in preparation). The respective manual (English) can be found on our website http://www.ifrg.uzh.ch/vox16o.php. The package itself is available to the nonprofit research community for free.

[5] 7 octaves; 24 quartertones per octave ('octave' means doubling frequency).

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

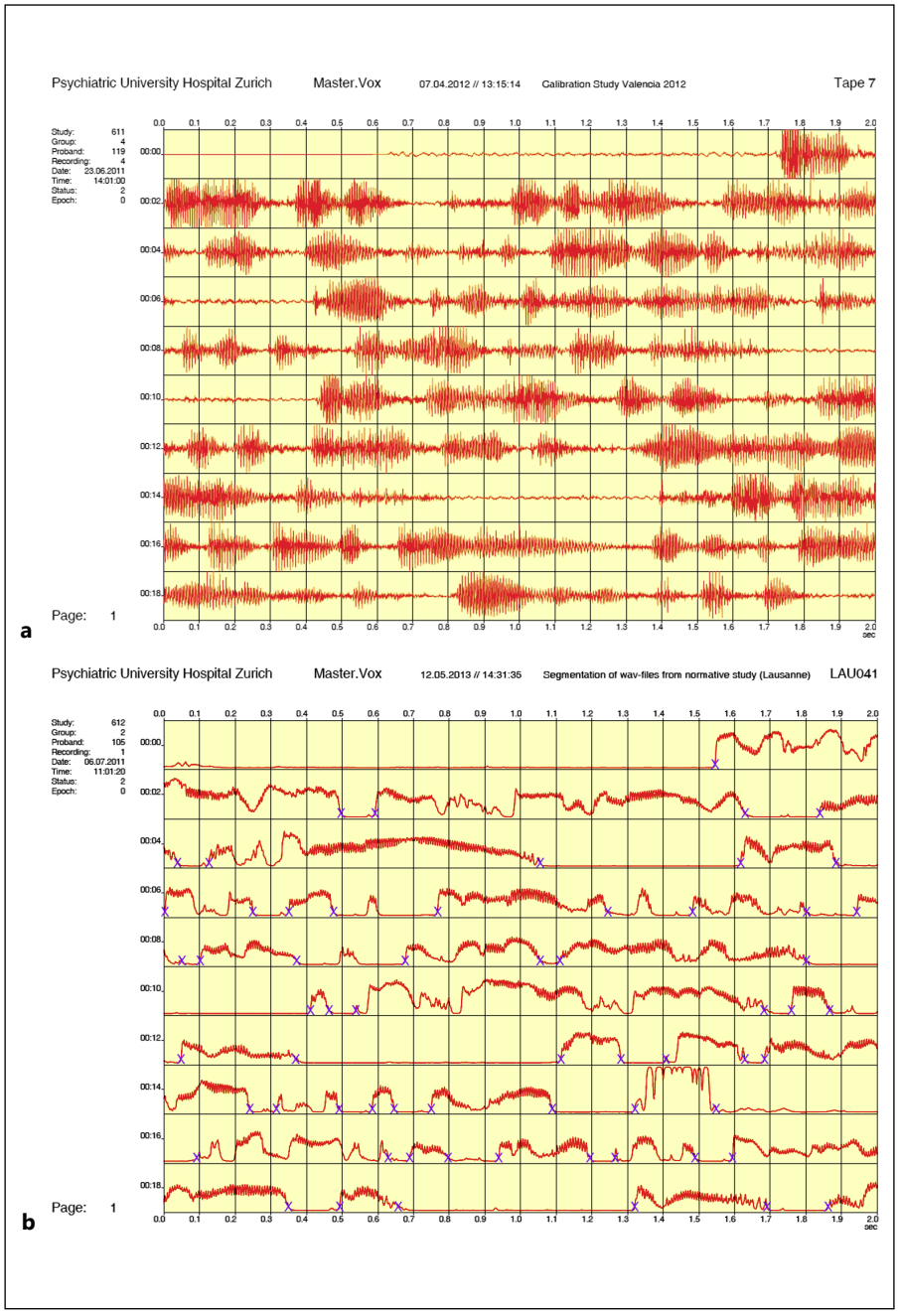
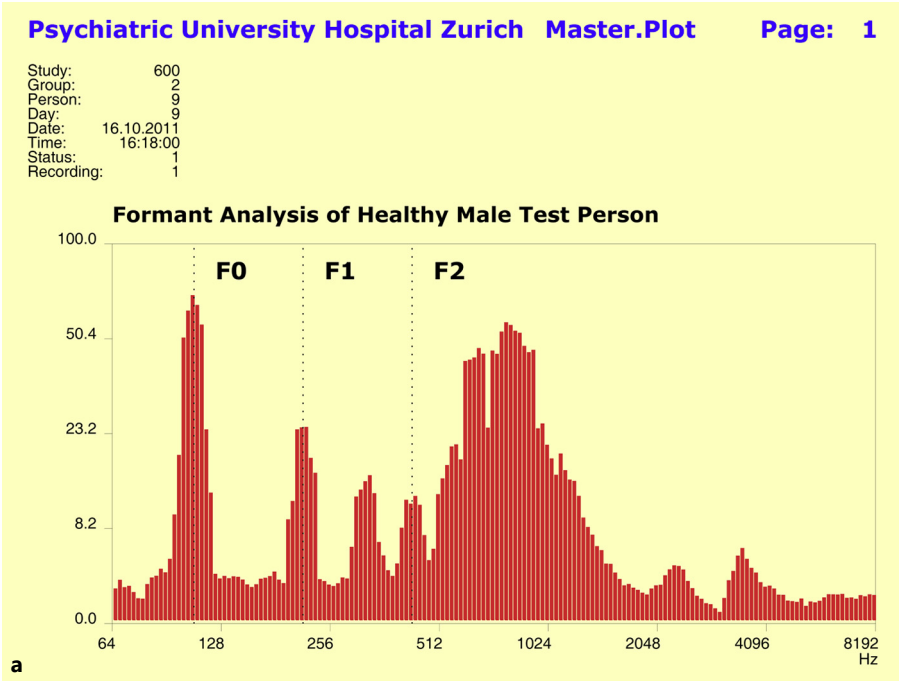


Fig. 1. a All speech signals are recorded as time series with a sampling rate of 96 kHz at a 16-bit resolution. This plot visualizes the characteristics of a spoken text (verbal content, rhythm, stress, and intonation) along with the speaker's affective state at the time point of speech production. From the technical perspective, the plot also gives an impression of the quality of the speech signal, for example, in terms of the signal-to-noise ratio. **b** Speech signals are subdivided into pauses and utterances by means of a 'segmentation' algorithm. This segmentation is carried out under consideration of background noise, using language-specific thresholds. Reliable segmentation is critically important when quantifying speaking behavior and voice sound characteristics because it directly affects the resolution of the speech analysis method with respect to assessing the speaker's affective state.

'F0 variation', 'F0 6-dB bandwidth', and 'F0 contour', for the assessment of voice sound characteristics[6]. The last three parameters describe a speaker's intonation in terms of rise and fall in pitch around his/her rest position F0. Detailed analyses revealed distinct, highly characteristic differences between all four languages under investigation, irrespective of language family (stress-timed, syllable-timed). For example, German test persons produced the longest utterances of all four populations, while French test persons spoke in a much louder voice compared to the Spanish. These distinctive characteristics are remarkably stable over time as demon-

[6] As there is no standard approach to measuring emotion in human speech and a variety of parameter sets is discussed in the literature [e.g., 18–22], we relied on a combination of the most reliable quantities.

Color version available online
Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM



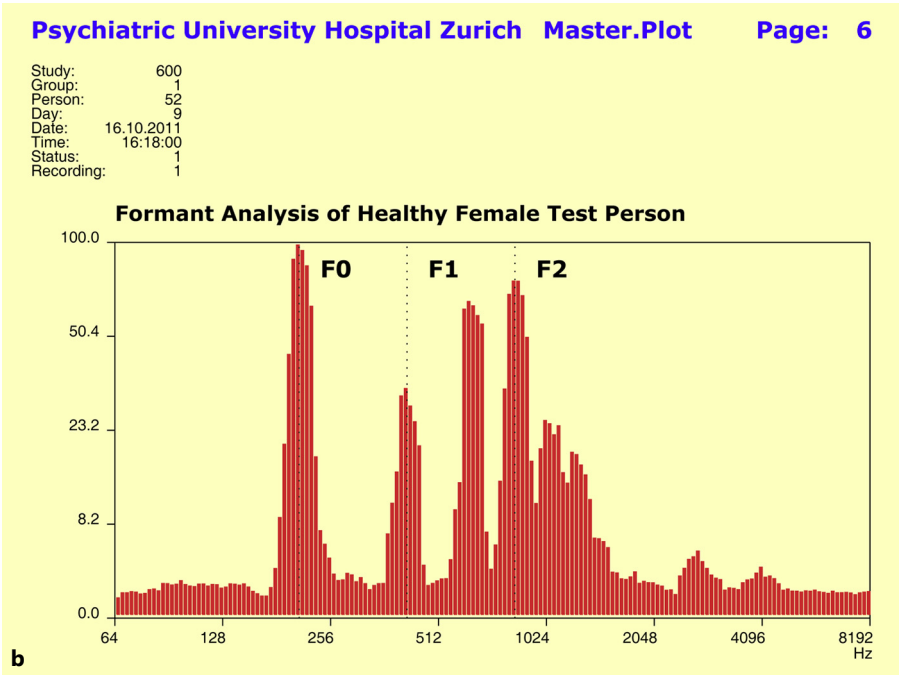


Fig. 2. a Voice sound characteristics (timbre) of a male speaker as quantified through spectral analysis. Spectral intensities are plotted along the y-axis on log-proportional scales and as a function of frequency (x-axis: 7 octaves covering the frequency range of 64–8,192 Hz). Mean vocal pitch F0 is 110 Hz. F1 and F2 indicate the overtones 1 octave (220 Hz) and 2 octaves (440 Hz) above mean vocal pitch F0. The overtone between F1 and F2 is a fifth higher than F1 and a fourth below F2. The distribution and intensity of overtones exhibit characteristic patterns with large interindividual variations while being remarkably stable over time. This enables a computerized recognition of persons by spectral voice patterns. **b** Voice sound characteristics (timbre) of a female speaker as quantified through spectral analyses. Spectral intensities are plotted along the y-axis on log-proportional scales and as a function of frequency (x-axis: 7 octaves covering the frequency range of 64–8,192 Hz). Mean vocal pitch F0 is 220 Hz. F1 and F2 indicate the overtones 1 octave (440 Hz) and 2 octaves (880 Hz) above mean vocal pitch F0. The overtone between F1 and F2 is a fifth higher than F1 and a fourth below F2. The mean vocal pitch of female speakers lies, on average, approximately 1 octave above that of male speakers.

strated through repeated assessments at 14-day intervals (tables 2, 3).

Taken together, the 10 speech parameters enabled direct verification of the linguistic prosody theorem stating that 'prosody is a distinctive feature for all languages'. In fact, linear discriminant analyses yield rates of 85.1% (counting out loud), 90.5% (reading out loud emotionally neutral text), and 90.0% (reading out loud emotionally stimulating text) correctly assigned speakers to their native languages. Stability and reproducibility of speech parameters over time can also be demonstrated by means of scatter plots in which each subject's measurement derived from the first recording is plotted against the second measurement 14 days later.

Color version available online

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

Table 2. Stress-timed languages

Speech parameters	English		German	
	baseline	14 days later	baseline	14 days later
Pause duration	165.7±33.7	164.0±33.7	177.9±30.3	174.8±30.9
Utterance duration	211.0±53.7	214.3±53.7	368.7±62.8	369.3±56.8
Energy per utterance	208.6±61.9	206.8±61.9	180.2±44.7	171.7±38.9
Dynamics	73.8±24.0	74.7±24.0	63.5±15.1	61.4±13.6
Energy per second	143.3±44.7	143.7±44.7	136.8±32.6	131.5±29.3
Vocal pitch F0 (QT)	27.4±9.0	27.9±9.0	32.1±11.3	32.0±11.2
F0 variation (QT)	6.4±3.1	5.9±3.1	4.2±1.7	4.0±1.7
F0 amplitude	80.0±16.1	78.8±16.1	87.3±28.5	87.7±28.4
F0 6-dB bandwidth	10.1±0.9	10.2±0.9	10.9±1.0	10.9±1.0
F0 contour	8.7±2.0	8.4±2.0	8.8±3.3	8.9±3.4

Cross-comparison between the stress-timed languages English and German. Speaking behavior and voice sound characteristics are quantified through a set of 10 speech parameters which are stable over time while revealing highly significant differences between the spoken languages. The table lists mean values ± standard deviations. The quantitative speech parameter 'patterns' act as 'fingerprints', enabling automatic assignment of speakers to native languages at an error rate of approximately 10%.

Table 3. Syllable-timed languages

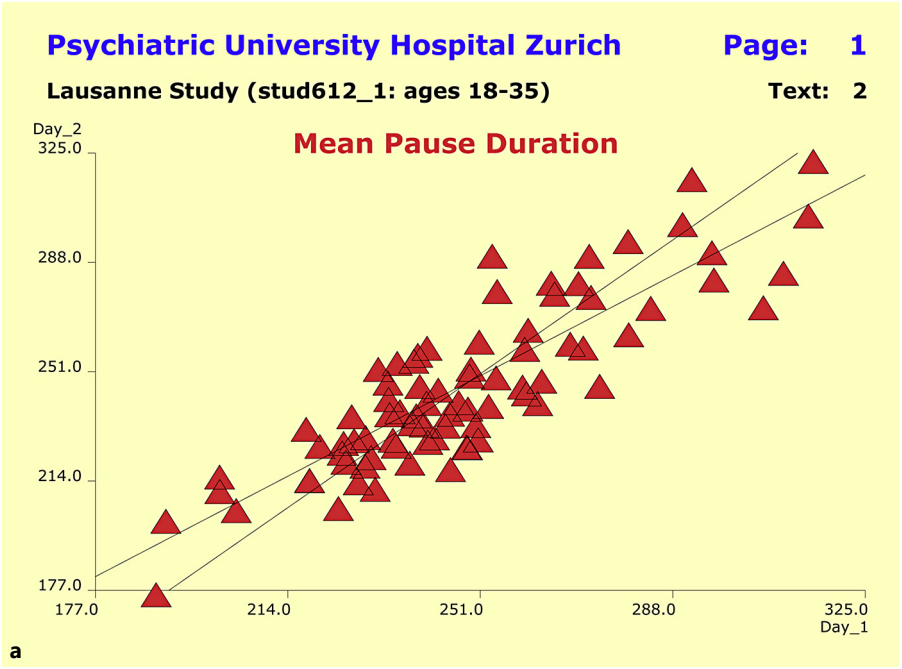
Speech parameters	French		Spanish	
	baseline	14 days later	baseline	14 days later
Pause duration	191.6±22.4	190.6±22.0	123.1±21.4	119.6±16.3
Utterance duration	221.8±25.2	221.8±24.2	223.7±82.1	203.5±77.8
Energy per utterance	242.9±76.3	243.2±82.9	88.3±19.1	79.4±21.1
Dynamics	81.1±25.7	80.4±26.8	30.3±7.9	25.9±6.1
Energy per second	156.8±48.8	157.4±52.7	49.4±14.6	45.3±16.0
Vocal pitch F0 (QT)	34.9±10.0	35.3±9.5	50.1±12.4	48.8±13.0
F0 variation (QT)	5.1±2.6	4.9±2.3	7.4±7.1	8.8±7.5
F0 amplitude	103.7±29.5	104.7±27.8	63.1±18.2	60.4±17.6
F0 6-dB bandwidth	11.3±1.2	11.3±1.1	11.8±1.4	11.7±1.7
F0 contour	10.2±3.6	10.3±3.5	5.8±1.9	5.6±1.7

Cross-comparison between the syllable-timed languages French and Spanish. Speaking behavior and voice sound characteristics are quantified through a set of 10 speech parameters which are stable over time while revealing highly significant differences between the spoken languages. The table lists mean values ± standard deviations. The quantitative speech parameter 'patterns' act as 'fingerprints', enabling automatic assignment of speakers to native languages at an error rate of approximately 10%.

Scatter plots show the interindividual variation[7] of a speech parameter along with the intraindividual stability of the parameter over time. The angle between the two regression lines $y = a_1x + b_1$ and $x = a_2y + b_2$ measures stability: the smaller the angle, the higher the intraindividual stability of the underlying parameter over the 14-day interval. Our results suggested almost perfect reproducibility of the parameter 'vocal pitch' with correlations in the range of 0.9 or higher (fig. 3b) while also showing large interindividual variation. The bimodal distribution reflects the fact that mean vocal pitch in fe-

[7] Variation implies information: the larger a parameter's interindividual variation the better its resolution of subtle between-subject differences.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM



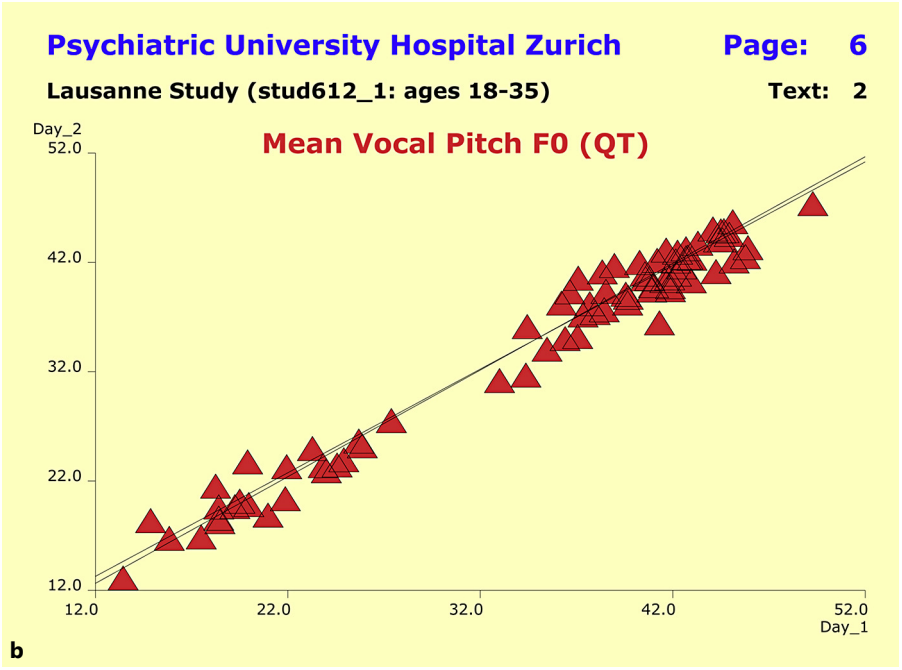


Fig. 3. a Stability of speech parameters 'mean pause duration' as a function of time in healthy volunteers: for each test person the first assessment is plotted along the x-axis and the second assessment 14 days later along the y-axis (red triangles). The angle between the two regression lines $y = a_1x + b_1$ and $x = a_2y + b_2$ is a quantitative measure of stability: the smaller the angle, the higher the intraindividual stability over the 14-day interval. The experimental condition is 'reading out loud emotionally neutral text'. **b** Stability of speech parameters 'mean vocal pitch' (in quartertones) as a function of time in healthy volunteers: for each test person the first assessment is plotted along the x-axis and the second assessment 14 days later along the y-axis (red triangles). As mean vocal pitch in females lies 1 octave above that of male speakers the scatter plot yields a bimodal distribution. The angle between the two regression lines $y = a_1x + b_1$ and $x = a_2y + b_2$ is a quantitative measure of stability: the smaller the angle, the higher the intraindividual stability over the 14-day interval. The experimental condition is 'reading out loud emotionally neutral text'.

males lies, on average, 1 octave above that of males (fig. 3b). The speech parameter 'mean pause duration' showed similar between-subject variation, but at a somewhat reduced stability over time (fig. 3a).

The stability of speech parameters over time can also be quantified through correlation coefficients. Table 4 gives an overview of the extent to which the speech parameters are stable over a 14-day interval and, conversely, are sensitive to the speaker's immediate environment and affective state. Additionally, our analyses yielded estimates of the parameters' 'natural fluctuations'.

Analysis of variance carried out separately for all four languages revealed highly significant influences of the text types on speech parameters. Even an automatized assign-

Color version available online
Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

Table 4. Correlation between recordings at 14-day intervals

Speech parameters	English		German		French		Spanish	
	18–35 years	36–64 years	18–35 years	36–64 years	18–35 years	36–64 years	18–35 years	36–64 years
Pause duration	0.8909	0.8318	0.8337	0.8849	0.8663	0.8619	0.6651	0.8156
Utterance duration	0.7180	0.6973	0.8486	0.8945	0.8576	0.8949	0.9468	0.8265
Energy per utterance	0.6030	0.4203	0.6832	0.5503	0.5670	0.5966	0.1512	0.2478
Dynamics	0.5851	0.3032	0.6411	0.4719	0.6215	0.5991	0.0876	0.1805
Energy per second	0.6836	0.2924	0.6715	0.5777	0.5888	0.5983	0.6629	0.5172
Vocal pitch F0 (QT)	0.9369	0.9244	0.9613	0.9677	0.9857	0.9639	0.7462	0.8959
F0 variation (QT)	0.4900	0.4018	0.3243	0.4329	0.2060	0.2065	0.4993	0.5652
F0 amplitude	0.7620	0.7289	0.9002	0.8651	0.9121	0.8820	0.7068	0.8379
F0 6-dB bandwidth	0.5802	0.7164	0.5570	0.5496	0.7641	0.7009	0.7374	0.7335
F0 contour	0.7662	0.7781	0.8952	0.8458	0.8603	0.8536	0.7308	0.8048

Stability of speech parameters over time in terms of correlations between two recordings at 14-day intervals, separately for the stress-timed languages English and German, the syllable-timed languages French and Spanish, and the age-classes 18–35 and 36–64 years. The higher the correlation coefficient, the lower the impact of environmental factors on this particular aspect of speaking behavior and voice sound characteristics.

ment of speech parameter patterns to text types by means of linear discriminant analysis[8] gave a rate of 69.9% correctly assigned samples. All this underlined the necessity of standardized experimental settings with fixed texts. As one may have expected, gender and age explained as much as 15–35% of the observed variance of the parameters that assess the speakers' voice sound characteristics, compared to only 1–3% for speaking behavior. The interaction gender × age reached significance as well. The effect of 'educational level' on all speech parameters was much smaller in the range of 1–3%, yet did not always reach statistical significance.

Assessing the four dimensions 'speech flow', 'loudness', 'vocal pitch', and 'intonation' through 10 speech parameters necessarily involves redundancy, though each parameter contributes a certain amount of unique information to the quantitative model of speaking behavior and voice sound characteristics.

The parameter 'utterance duration' (speech flow) does not correlate with all other parameters except for 'pause duration' ($r = 0.2586$); the parameters used to quantify loudness ('energy per second', 'energy per utterance', 'variation of energy per utterance') are highly intercorrelated with each other (up to $r = 0.9397$); the parameter 'mean vocal pitch F0' does not correlate with all other parameters, while 'F0 amplitude' shows intercorrelations with most other parameters in the range of $-0.3452 \leq r \leq 0.5601$; the parameters used to quantify intonation ('F0 variation', F0 6-dB bandwidth', 'F0 contour') are intercorrelated with each other showing correlations in the range of $-0.5614 \leq r \leq 0.0053$ (all the above correlation coefficients are for the experimental condition 'reading out emotionally neutral text', the correlations for the two other texts were of the same order of magnitude).

Longitudinal Self-Assessment Study

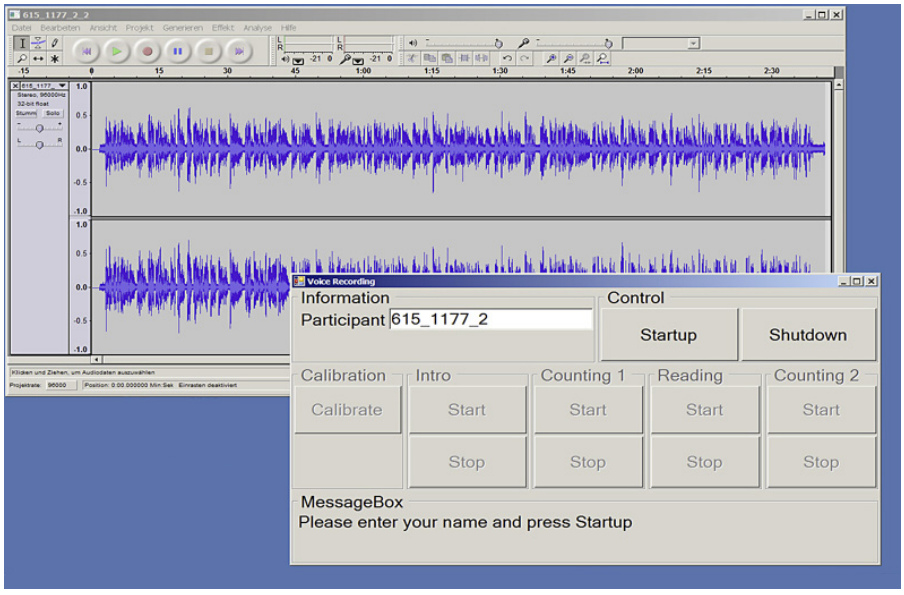
For the purpose of this study, we adapted our high-tech speech laboratory system Master.Vox to run on consumer 64-bit platforms like Windows, Linux, Solaris, or Android, that is, to run on standard netbooks, tablets, or smartphones[9]. Additionally, a self-explaining Graphical User Interface was developed to enable easy-to-use self-assessments (fig. 4).

Based on this newly developed system, our longitudinal study with repeated assessments at 1-day intervals over 14 days was carried out with 18 university students (Zurich: German) and 18 unemployed adults (Valencia:

[8] We used PROC STEPDISC and PROC DISCRIM (SAS 9.3) with the parametric method based on a multivariate normal distribution within each class, along with subsequent cross-validation. All 10 speech parameters under investigation were included in the analysis and reached statistical significance ($p \leq 0.0015$). R^2 ranged between 0.3505 ('mean vocal pitch F0') and 0.6563 ('energy per second'), except for 'utterance duration' (0.0193), 'F0 variation' (0.0921), and 'F0 6-dB bandwidth' (0.2267).

[9] See website http://www.ifrg.uzh.ch/vox16o.php.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM



Color version available online

Fig. 4. Voice recordings can be carried out as self-assessments in the test person's home environment by means of a second-generation microphone with integrated A/D converter that plugs into any standard USB port of a netbook. A tablet/smartphone solution will be available soon as Android application.

Spanish) in their home environments. The voice recording equipment turned out to work surprisingly well in the majority of cases (72.2%), and the speech parameters' stability over time and 'natural fluctuations' were comparable with those of the normative studies. Insufficient data were mainly due to missing measurements on one or several days where test persons 'forgot' the speech recordings for various reasons. Missing data were noncritical as ong as they were isolated and did not compromise baseline estimation. A crucial point, however, was background noise: almost 10% of the recordings could not be analyzed due to insufficient signal-to-noise ratios. Technical issues such as starting and closing a speech recording played a minor role.

Most test persons exhibited relatively 'flat' time series over the 14-day observation period though at interindividually different baseline levels. That is, speaking behavior was virtually unchanged over time except for some 'natural' fluctuations (fig. 5a, b). Significant deviations were relatively rare (fig. 5c), while some habituation effects (e.g., continuous decrease in pause duration and systematic increase in loudness as shown in fig. 5d) were observed in nearly 20% of test persons.

For the single-case analyses, we calculated the individual 'baselines' for each test person along with deviations thereof (taking nonlinear trends into consideration). The clinical relevance of the observed deviations was assessed by comparison with the language-specific and gender- and age-corrected reference data from our population-based normative studies. Although some of the observed deviations from baseline reached statistical significance (fig. 5c), we did not find longer persisting deviations from 'normality' over several days – which was quite unlikely given the design of this pilot study where test persons under chronic stress could be expected to show some short-term affective reactions, but were quite unlikely to develop longer persisting deviations from 'normality'. Despite this somewhat 'negative' selection of test persons we did not observe signs of pronounced dullness or amusement.

In summary, our normative data have shown that the chosen speech parameters can pick up the differences between texts across subjects and languages despite some habituation effects that might exist for repeated assessments at 14-day intervals. Our longitudinal self-assessment data from repeated assessments at 1-day intervals over 14 days have demonstrated: (1) the feasibility and efficiency of the speech analysis method in typical home environments and (2) the sensitivity of the speech analysis method as to detecting deviations from baseline in self-assessments.

Discussion

Within the scope of our EU-funded project 'Early Prediction and Prevention of Depression' (cf. http://www.ifrg.uzh.ch/optimi.php), we carried out a population-based study on speaking behavior and voice sound characteristics involving four languages (English, French, German, Spanish; n = 577), along with a longitudinal self-assess-

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

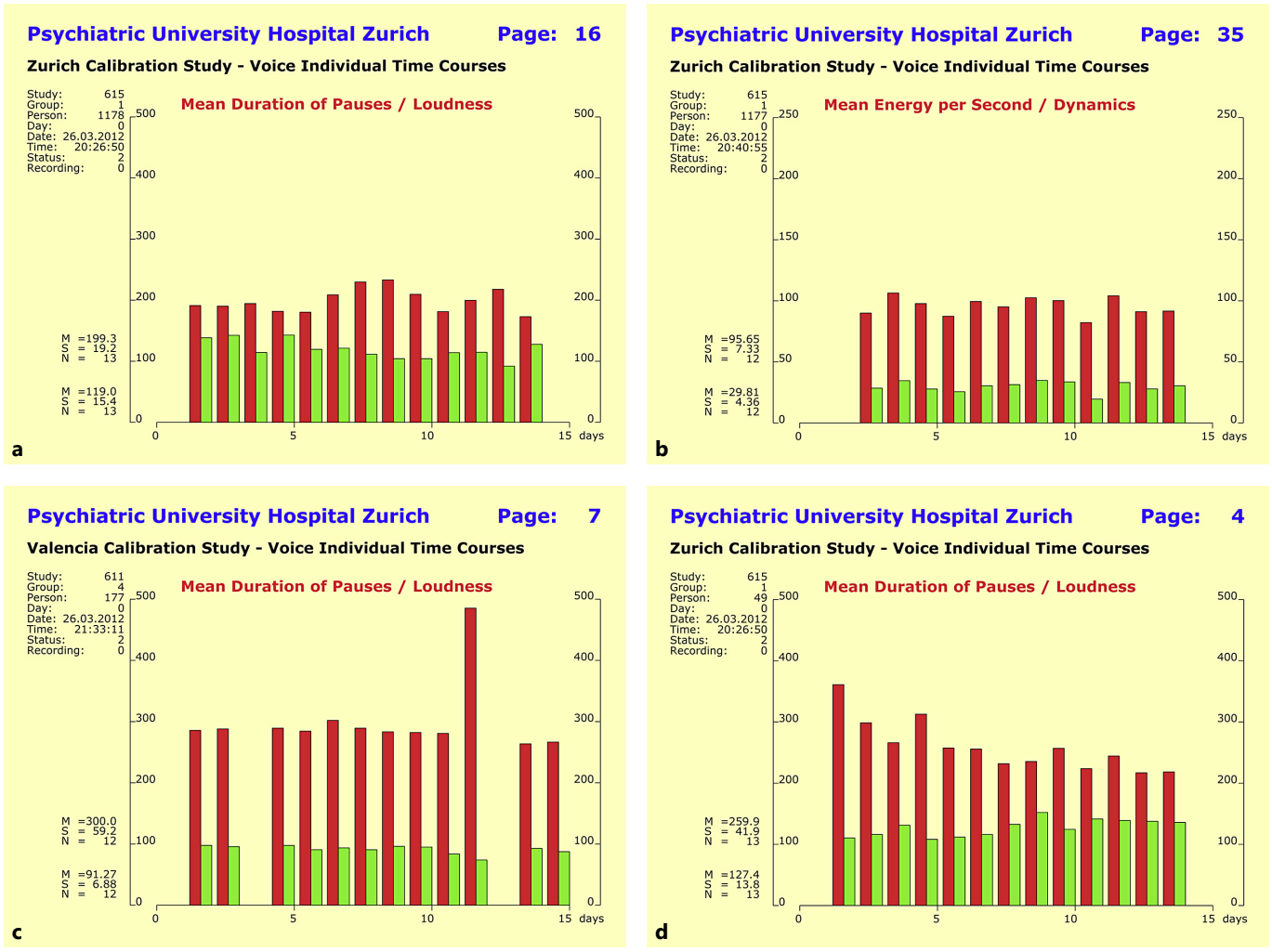
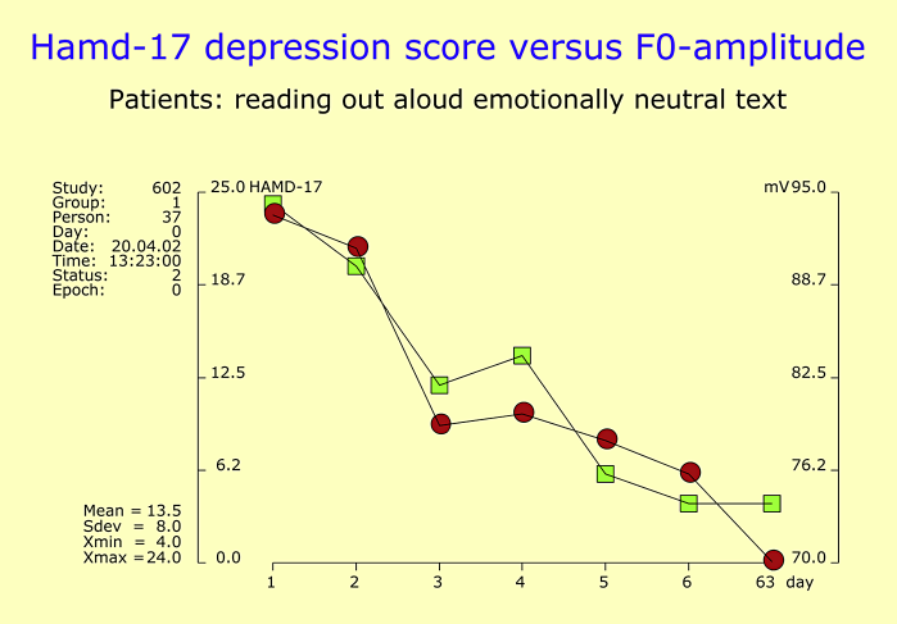


Fig. 5. a Time course of pause duration (red bars) and loudness (green bars) over an observation period of 14 days. Speaking behavior is virtually unchanged over time except for some 'natural' fluctuations. **b** Time course of energy (red bars) and dynamic expressiveness (green bars) over an observation period of 14 days. Speaking behavior is virtually unchanged over time except for some 'natural' fluctuations. **c** Time course of pause duration (red bars) and loudness (green bars) over an observation period of 14 days. Speaking behavior is virtually unchanged over time except for day 12 displaying longer pauses and a low voice (the subject may be tired). No recordings were done on days 4 and 13. **d** Time course of pause duration (red bars) and loudness (green bars) over an observation period of 14 days. Speaking behavior shows a systematic trend towards shorter pauses and greater loudness as a function of time, a fact that may indicate a habituation effect.

ment study involving two languages (German, Spanish; n = 36). The studies addressed the following questions: (1) Can the prosodic elements 'distribution of pauses and utterances', 'stress', and 'intonation patterns' be quantified in a language-independent way and at a sufficiently high resolution? (2) To what extent do the factors 'gender', 'age', and 'education' modify prosodic elements? (3) To what extent do the thresholds between 'natural fluctuations' and 'significant changes' depend on language? (4) Can self-assessments be realized in typical home environments?

Results from our normative data with two speech recordings at 14-day intervals and three different types of spoken text yielded convincing evidence that the central elements of speaking behavior and voice sound characteristics ('prosody') can be quantified in a reproducible and language-independent way through a set of 10 speech parameters. As to measuring subtle between-language and between-text differences, the resolving power of this set of parameters was indeed remarkable. Even a computerized assignment of speech parameter patterns to languages

Color version available online
Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM



Color version available online

Fig. 6. Time course of a patient's recovery from depression as reflected by 17-item HAM-D scores (green square points) assessed at 2-day intervals over an observation period of 2 weeks, plus a final assessment at the time of discharge from hospital (day 63). The corresponding change over time of the speech parameter 'F0 amplitude' is also shown (red circle points) in order to demonstrate the close relationship between the two courses of development.

gave a rate of 90% correctly classified patterns, while the correct assignment to the texts was with 70% noteworthy too. All this indicated that approaches to monitoring the affective state of a speaker through analysis of speaking behavior and voice sound characteristics can successfully rely on the proposed set of speech parameters but have to integrate language- and text-specific criteria. Similarly, gender- and age-specific effects have to be taken into account [cf., 14] when deciding about the significance of an observed variation in speaking behavior and voice sound characteristics[10]. The speaker's educational level, by contrast, was found to be of minor relevance.

Over the past decades, the speech analysis method has been carried out primarily in acoustically shielded speech laboratories with the aim of monitoring the transition from 'affectively disturbed' to 'normal' among psychiatric patients under treatment. The evolution of multiprocessor netbooks, tablets, and smartphones, along with the open-source Android operating system, has changed the situation completely, thus clearing the way to much broader applications, in particular, self-assessments in the test persons' home environment. The results of our longitudinal self-assessment study underlined the feasibility and efficiency of the method in home environments. A crucial point, however, is the necessity of regular assessments, which requires self-discipline and motivation, particularly in times of negative emotions.

Undoubtedly, our longitudinal self-assessment study has cleared the way for a prospective study of English/German patients who recently recovered from depression and are at risk of relapse. This study is designed to detail the transition from 'normal' to 'affectively disturbed' and will complement our patient data (transition from 'affectively disturbed' to 'normal') where single-case analyses revealed in 65% of cases a close correlation over time (at 2-day intervals) between the HAM-D score on the one hand and speech parameters on the other (fig. 6). All this gives rise to optimistic expectations regarding routine applications of the speech analysis method as to (1) prospectively identifying subjects with longer-persisting affective disturbances in the general population; (2) monitoring patients who recovered from depression but are at risk of relapse, and (3) 'objectively' assessing the time course of recovery from depression and the time point of onset of action of psychopharmacological treatment [cf., 23].

On the technical side, background noise may be a limiting factor, and this method of approach may not be equally suited for everyone as suggested by some 3% of test persons with reading problems in our sample. The recording scheme 'counting-reading-counting' was found to be appropriate for self-assessments as well as for professional monitoring studies: it helped the test person to relax, to feel comfortable, and to get 'the job done'. Even though 'objective' assessments of a speaker's affective state

[10] Tentatively defined as 1.5 standard deviations from baseline and stratified with respect to language, gender, and age.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM

through speech analysis play a more prominent role in other fields, such as forensic medicine [e.g., 24, 25], economics [e.g., 26], or psychology [e.g., 22], the availability of low-cost, easy-to-use speech analysis equipment will lead to a broader range of applications in psychiatry as well. Particularly interesting are the prospects of self-assessments among subjects at an elevated risk for developing affective disorders, for example, among subjects with insufficient coping behavior under chronic stress [27, 28].

Conclusions

Results from our normative data from 577 healthy subjects of four different native languages showed that speaking behavior and voice sound characteristics can be quantified in a reproducible and language-independent way. Additionally, these data provided gender-, age-, and language-specific thresholds that allow one to reliably distinguish between 'natural fluctuations' and 'significant changes'. The longitudinal self-assessment study with 36 test persons and repeated assessments at 1-day intervals over 14 days demonstrated the feasibility and efficiency of the speech analysis method in typical home environments, thus clearing the way to a broader range of applications in psychiatry.

Acknowledgment

This project was funded in part through the 7th EU Framework Programme for Research and Technological Development (grant 248544; OPTIMI: http://www.ifrg.uzh.ch/optimi.php).

References

1 Kraepelin E: Manic-Depressive Insanity and Paranoia (translation by M Barclay from Kraepelin E: Psychiatrie Bd 1). Edinburgh, Livingstone, 1921, p 415.
2 Kuny S, Stassen HH: Speaking behavior and voice sound characteristics in depressive patients during recovery. J Psychiatr Res 1993; 27:289–307.
3 Stassen HH, Albers M, Püschel J, Scharfetter C, Tewesmeier M, Woggon B: Speaking behavior and voice sound characteristics associated with negative schizophrenia. J Psychiatr Res 1995;29:277–296.
4 Püschel J, Stassen HH, Bomben G, Scharfetter C, Hell D: Speaking behavior and voice sound characteristics in acute schizophrenia. J Psychiatr Res 1998;32:89–97.
5 Lott PR, Guggenbühl S, Schneeberger A, Pulver AE, Stassen HH: Linguistic analysis of the speech output of schizophrenic, bipolar, and depressive patients. Psychopathology 2002; 35:220–227.
6 Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS: Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguistics 2007;20:50–64.
7 Mundt JC, Vogel AP, Feltner DE, Lenderking WR: Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry 2012;72:580–587.
8 Scherer KR: On the nature and function of emotion: a component process approach; in Scherer KR, Ekman P (eds): Approaches to Emotion. Hillsdale, Erlbaum, 1984, p 293.
9 Stassen HH: Affekt und Sprache. Monographien aus dem Gesamtgebiet der Psychiatrie. Berlin, Springer, 1995.
10 Rosenberg AE, Shipley KL: Speaker identification and verification with speaker-independent word recognition. IEEE Conf ASSP, Atlanta, 1981, pp 184–187.
11 Stassen HH: Affective state and voice – the specific properties of overtone distributions. Methods Inf Med 1991;30:44–52.
12 Cutler A, Oahan D, van Donselaar W: Prosody in the comprehension of spoken language: a literature review. Lang Speech 1997;40:141–201.
13 Arvaniti A: Rhythm, timing and the timing of rhythm. Phonetica 2009;66:46–63.
14 Stathopoulos ET, Huber JE, Sussman JE: Changes in acoustic characteristics of the voice across the life span: measures from individuals 4–93 years of age. J Speech Lang Hear Res 2011;54:1011–1021.
15 McComb K, Shannon G, Sayialel KN, Moss C: Elephants can determine ethnicity, gender, and age from acoustic cues in human voices. Proc Natl Acad Sci USA 2014;111:5433–5438
16 Stassen HH, Kuny S, Hell D: The speech analysis approach to determining onset of improvement under antidepressants. Eur Neuropsychopharmacol 1998;8:303–310.
17 Stassen HH, Angst J, Hell D, Scharfetter C, Szegedi A: Is there a common resilience mechanism underlying antidepressant drug response? Evidence from 2,848 patients. J Clin Psychiatry 2007;68:1195–1205.
18 Scherer KR: Vocal communication of emotion: a review of research paradigms. Speech Commun 2003;40:227–256.
19 Juslin PN, Scherer KR: Vocal expression of affect; in Harrigan JA, Rosenthal R, Scherer KR (eds): The New Handbook of Methods in Nonverbal Behavior Research. Oxford, Oxford Press, 2005, pp 65–135.
20 Ververdis D, Kotropoulos C: Emotional speech recognition: resources, features and methods. Speech Commun 2006;48:1162–1181.
21 Schuller B: On the acoustics of emotion in speech: desperately seeking a standard. J Acoust Soc Am 2010;127:1995.
22 Weninger F, Eyben F, Schuller B, Mortillaro M, Scherer KR: On the acoustics of emotion in audio: what speech, music, and sound have in common. Front Psychol 2013;4:292.
23 Stassen HH, Anghelescu IG, Angst J, Böker H, Lötscher K, Rujescu D, Szegedi A, Scharfetter C: Predicting response to psychopharmacological treatment. Survey of recent results. Pharmacopsychiatry 2011;44:263–272.
24 Sommers MS: Evaluating voice-based measures for detecting deception. J Credibility Assess Witness Psychol 2006;7:99–107.
25 Harnsberger JD, Hollien H, Martin CA, Hollien KA: Stress and deception in speech: evaluating layered voice analysis. J Forensic Sci 2009;54:642–650.
26 Mayew WJ, Venkatachalam M: The power of voice: managerial affective states and future firm performance. J Finance 2012;67:1–43.
27 Mohr C, Braun S, Bridler R, Chmetz F, Delfino JP, Kluckner VJ, Lott P, Schrag Y, Seifritz E, Stassen HH: Insufficient coping behavior under chronic stress and vulnerability to psychiatric disorders. Psychopathology 2014;47:235–243.
28 Delfino JP, Barragán E, Botella C, Braun S, Camussi E, Chafrat V, Mohr C, Bridler R, Lott P, Moragrega I, Papagno C, Sanchez S, Soler C, Seifritz E, Stassen HH: Quantifying insufficient coping behavior under chronic stress. A cross-cultural study of 1,303 students from Italy, Spain, and Argentina. BMC Psychiatry 2013, submitted for publication.

Downloaded by:
Universität Zürich, Zentralbibliothek Zürich
130.60.170.2 - 9/16/2014 11:24:01 AM