

Schumpeter School
of Business and Economics



SCHUMPETER DISCUSSION PAPERS

**Phonetisch-akustische Detektion von Selbstsicherheit -
Entwicklung eines automatisierten Messverfahrens zur
Personalentwicklung**

**Jarek Krajewski
Silke Kessel**

The Schumpeter Discussion Papers are a
publication of the Schumpeter School of
Business and Economics, University of
Wuppertal, Germany

For editorial correspondence please contact
SSBEditor@wiwi.uni-wuppertal.de

SDP 2010-006
ISSN 1867-5352

Impressum
Bergische Universität Wuppertal
Gaußstraße 20
42119 Wuppertal
www.uni-wuppertal.de
© by the author



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Phonetisch-akustische Detektion von Selbstsicherheit - Entwicklung eines automatisierten Messverfahrens zur Personalentwicklung

Silke Kessel, Jarek Krajewski

Schumpeter School of Business and Economics,
Experimentelle Wirtschaftspsychologie, Bergische Universität Wuppertal
{kessel, krajewski}@wiwi.uni-wuppertal.de

Abstract

This paper describes a measurement approach for detecting sympathy and self-confidence based on speech characteristics as investigative personal assessment. The advantages of this automatic real time approach are that obtaining speech data is objective and non obtrusive, and it allows multiple measurement over long periods of time. Different types of acoustic features were computed. In order to identify speech correlates of self-confidence and sympathy, 10 actors were recorded, resulting in 100 segments of speech. 12 raters independently labeled the sympathy and self-confidence impression of the speech segments. Validation strategies reaching recognition rates for 2-class problems of 62.75-76.47 %, in classifying slight from strong sympathy and self-confidence.

Zusammenfassung

Die vorliegende Studie beschreibt ein Validierungsdesign, anhand dessen Sympathie und Selbstsicherheit vorhergesagt werden sollen. Die Methode stellt einen investigativen Ansatz für die Eignungsdiagnostik dar. Die Vorteile der Echtzeitmessung liegen in der objektiven und non-obstrusiven Gewinnung der Daten. Weiterhin ermöglicht sie Mehrfachmessungen im Längsschnittdesign. Um die verschiedenen akustischen Merkmale zu berechnen, wurden 10 Schauspieler aufgezeichnet. Insgesamt ergaben sich auf diesem Wege 100 Sprachsegmente, welche von 12 Beurteilern unabhängig voneinander hinsichtlich Sympathie und Selbstsicherheit beurteilt wurden. Die Klassifikationsgenauigkeit der Sympathievorhersage aufgrund der akustischen Merkmalsgeneration befand sich bei durchschnittlich 62.75 %, die Klassifikationsgenauigkeit der Selbstsicherheitsvorhersage bei durchschnittlich 76.47 %.

1. Einleitung

Die akustische Stimmanalyse soll im Folgenden als innovative Methode der Personalauswahl- und -entwicklung diskutiert werden. Es wird notwendigerweise nur auf solche Bewerbungssituationen Bezug genommen, in denen verbale Anteile dominant sind: Einstellungsinterview und Assessment Center (AC). Beispielhaft für eine eignungsdiagnostische Anforderungsdimension soll Selbstsicherheit herangezogen werden. Selbstsicherheit ist eine Eigenschaft, die sich in vielen Anforderungsdimensionen des Assessment Centers wieder findet (Neubauer, 2001) und auch den im Interview am häufigsten erfragten Dimensionen inhaltlich zuzuordnen ist (Huffcutt, Conway, Roth & Stone, 2001). Weiterhin wurde Sympathie stimmakustisch analysiert. Peters und Terborg (1975) wiesen nach, dass studentische Interviewer von den hypothetischen Bewerbern angezogen wurden, welche ihnen ähnlich waren. Als konsistenter Befund über viele Studien hinweg zeigte sich, dass Kandidaten mit ähnlichem biographischen Hintergrund, Einstellungen und wahrgenommener Persönlichkeit zum Interviewer generell positiver beurteilt werden (Anderson & Shakleton, 1990).

Der Einsatz akustischer Informationen zur Bestimmung von Sprechereigenschaften und Befindlichkeitszuständen ist schon seit vielen Jahren anvisiert (Liebermann & Michaels, 1962; Scherer, 1974). Aber erst mit Hilfe aktueller Prozessor- und Hardwareleistungen ist es möglich, die zentralen Fragestellungen des Forschungsfeldes der akustischen Stimmanalyse anzugehen. Im Folgenden werden die Kernelemente der akustischen Stimmanalyse und die Frage der akustischen Korrelate beschrieben.

Eine anschauliche Beschreibung der Stimme orientiert sich an höreindrucksbezogenen Kategorien. Daher besteht ein Zugang zur Bestimmung akustischer Maße auch darin, sich zunächst am veränderten Höreindruck zu orientieren und in einem zweiten Schritt nach den den Höreindruck quantifizierenden akustischen Kennzahlen zu suchen. Die wichtigsten auditiv-perzeptuellen Kategorien sind: Sprechstimmlage, Intonationsverlauf, Lautstärke, Sprechgeschwindigkeit, Sprechpausen, Segment-Dauer, Rhythmizität, Akzentuierungsstärke, Artikulationspräzision, Nasalität und Klangfülle. Um diverse Stimmphänomene abbilden zu können, werden in dem mit der akustischen Selbstsicherheitsanalyse verwandten Forschungsfeld der sprachbasierten Emotionserkennung (Speech Emotion Recognition; SER) breite Mehrzweck-Kennzahlensets bestimmt. Diese vielfältigen Kennzahlensets setzen sich aus akustischen Kennzahlen zusammen, die sich an den oben genannten auditiven Kategorien von Intensität, Intonation, Rhythmus, und Stimmqualität orientieren (z.B. Batliner, Schuller, Schaeffler & Steidl, 2008; Schuller, Vlasenko, Arsic, Rigoll & Wendemuth, 2008).

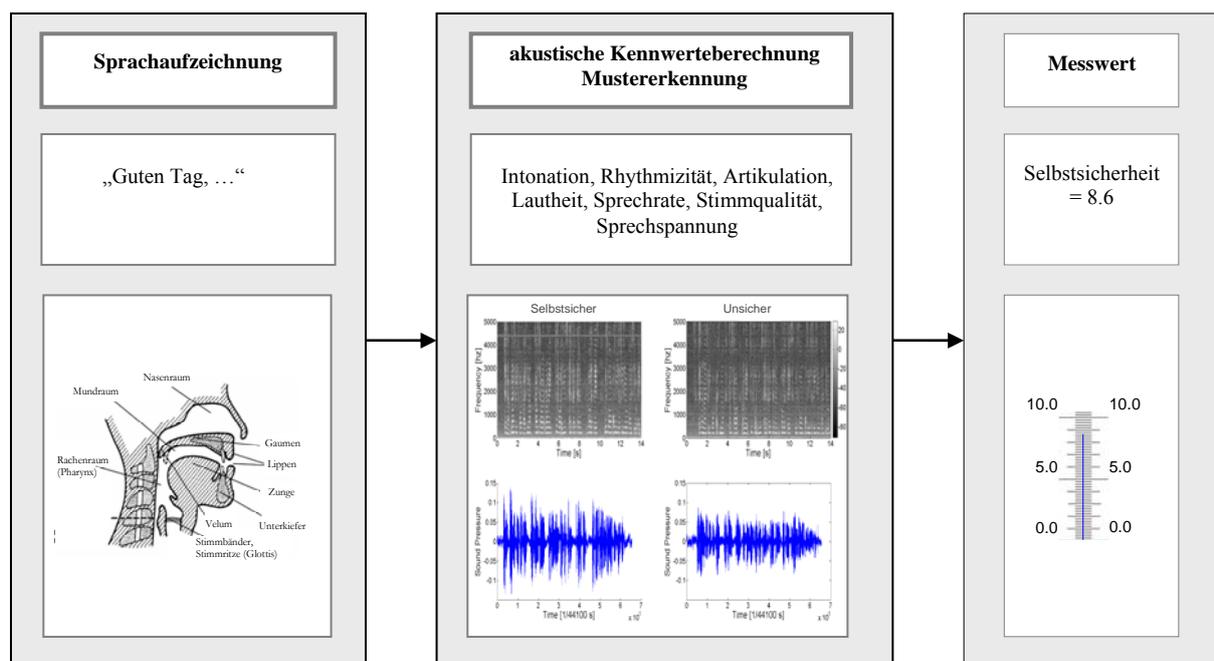


Abbildung 1: Grundprinzipien des akustischen Messansatzes

Die Sprachproduktion kann über folgende mentale und physische Prozessschritte beschrieben werden: Kreieren eines Sprachplans, Auswahl geeigneter linguistischer Einheiten aus dem Gedächtnis, Generieren einer Sequenz artikulatorischer Targets, Aktivierung des Motorprogramms der Targets, Transmission der neuromuskulären Kommandos zu den Muskeln des Respirations- und Phonationssystems (Stimmklappen), Bewegung der Artikulationsorgane, Nutzen des propriozeptiven Feedbacks und Ausstrahlung der akustischen Energie über Mund- und Nasenöffnungen. Auf der Ebene der physiologischen Veränderungen des Sprachproduktionsapparates lassen sich einige begründete Annahmen zu entspannungsinduzierten Wirkungen, wie sie z.B. für selbstsichere Sprecher typisch sind, aufstellen (s. Abbildung 2). Diese hier aufgestellten Annahmen zu möglichen - über parasympathische Aktivierung eines selbstsicheren-souveränen Sprechers vermittelten - Veränderungen des Sprachproduktionsprozesses, bieten einen Referenzrahmen und theoretischen Hintergrund für die Entwicklung des akustischen Messinstrumentes zur Erfassung von Selbstsicherheit und Sympathie.

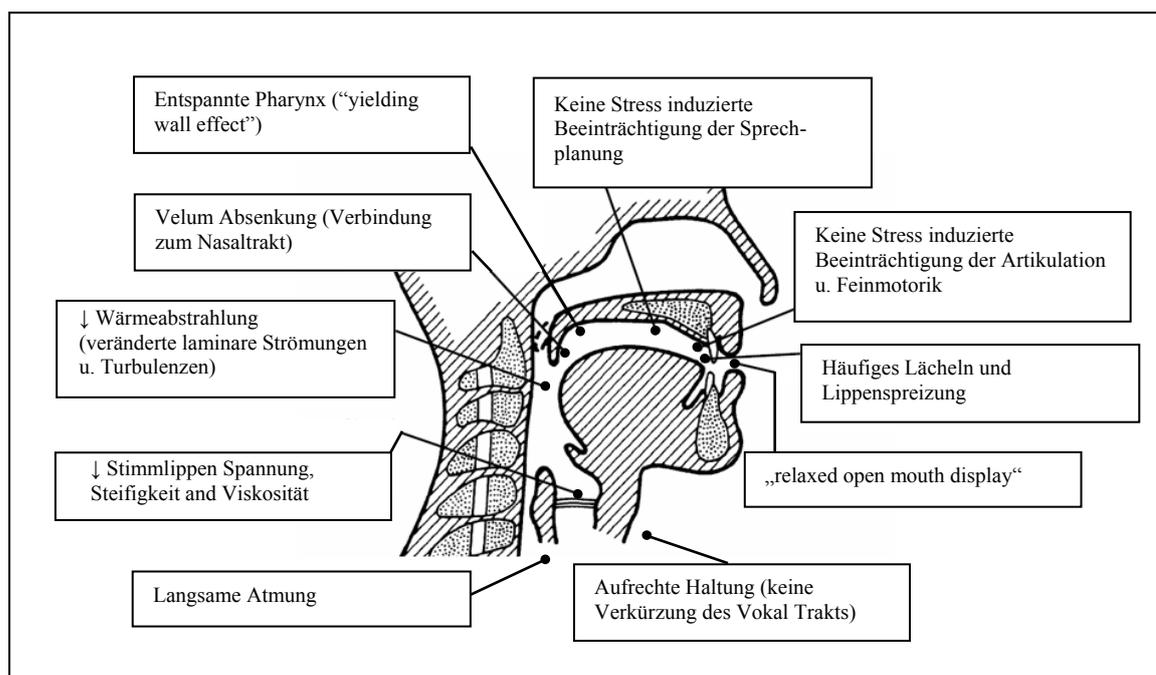


Abbildung 2: Mögliche selbstsicherheitsinduzierte Veränderungen des Sprachproduktionssystems.

Die akustische Stimmanalyse ist eine menpower effiziente, vorurteilsfreie Messung von relevanten Reaktionen, bei der das Gefühl einer starken Beobachtung vermieden werden würde. Neben dem inhaltlichen Mehrwert akustischer Messverfahren (Präzision, Robustheit und Verfälschungssicherheit) liefert die Entwicklung technologiebasierter Forschungsinstrumente einen weiteren wichtigen Reputationsgewinn psychologischer Forschungslinien. Urteile in Personalauswahl und –entwicklung können durch die Implementierung des objektiven Messverfahrens an Akzeptanz gewinnen.

2. Validierungsmethode

Um die akustische Stimmanalyse für die ausgewählten Beobachtungsdimensionen durchführen zu können, musste zunächst ein Sprachkorpora erstellt werden. Zur Gewinnung der Stimmproben wurden die deutschen Synchronstimmen verschiedener Darsteller aus US-amerikanischen Fernsehserien genutzt. Insgesamt wurden auf diese Weise von zehn verschiedenen Personen je zehn Stimmproben aufgezeichnet, so dass insgesamt 100 Samples vorlagen. 12 Personen nahmen die Beurteilung der Stimmproben anhand eines Fragebogens vor. Erfasst wurden so die subjektiven Beurteilungen hinsichtlich Sympathie und Selbstsicherheit.

Anschließend wurde die akustische Stimmanalyse durchgeführt. Sie besteht aus den Phasen: Vorverarbeitung, Merkmalsberechnung, Dimensionalitätsreduktion und Klassifikation. Die Vorverarbeitung fand bereits bei der Gewinnung des Sprachmaterials statt. Im nächsten Schritt wurden mit dem Sprachanalyseprogramm Praat folgende Kennwerte berechnet: Fundamentale Frequenz (F0), F0-Peakprozess, Formanten 1 bis 6 (F1-F6), Bandbreite der Formanten 1-6 (FBW1-6), Intensität, Spektrum, Harmonics-to-Noise Ratio (HNR), Mel-Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding, (LPC) und Linear-Frequency Cepstral Coefficients (LFCCs). Insgesamt wurden 56 Konturen berechnet. Für jede Kontur wurden 394 Kennwerte berechnet, so dass insgesamt 22064 Werte berechnet wurden. Aus den 394 Kennwerten wurden in einer supervised filterbasierten Feature Subset Selection mittels eines Korrelationsfilters die Werte mit der höchsten Relevanz herausselektiert. Auf das gefundene Feature-Set wurden die Klassifikationsverfahren Neural Net, Decision Tree, Naive Bayes-Klassifikation und Nearest Neighbors mit einem, zwei und drei Nachbarn angewendet. Klassifikationsfehler wurden mit einer 10-fachen Kreuzvalidierung bestimmt. Weiterhin wurden signifikanzstatistische Werte berechnet. Um Zusammenhänge zwischen den Ratings der Beurteiler hinsichtlich der empfundenen Sympathie gegenüber einem Sprecher, bzw. der Selbstsicherheit eines Sprechers und den aus den Stimmproben ermittelten Kennzahlen vorzunehmen, wurde zunächst eine Dichotomisierung in sympathische / selbstsichere und unsympathische / unsichere Stimmproben vorgenommen werden. Dies geschah auf Basis der vorliegenden Ratings. Beurteilten mehr als 66 % der Rater eine Stimmprobe als sympathisch / selbstsicher (Skalenwert von 6 oder mehr) oder unsympathisch / unsicher (Skalenwert von 4 oder weniger) wurde die Stimmprobe als sympathisch /selbstsicher bzw. unsympathisch /unsicher klassifiziert. Anschließend wurden Einzelkorrelationen bestimmt. Außerdem wurden t-Tests zur Abschätzung des Zusammenhangs zwischen dem Sprachmaterial und den Ratings gerechnet.

3. Ergebnisse

Drei der ermittelten Einzelkorrelationen zwischen Stimmkennzahlen und der in sympathisch und unsympathisch dichotomisierten Ratings erreichen das Signifikanzniveau: F0, $r = -.48$, $p < .01$, F1-F6, $r = .44$, $p < .01$ und LFCC, $r = -.11$, $p < .01$. Für das selbstsichere Sprachmaterial erreichten alle sieben Einzelkorrelationen das Signifikanzniveau: F0, $r = -.53$, $p < .01$, FBW1-6, $r = -.62$, $p < .01$, F1-F6, $r = -.62$, $p < .01$, HNR, $r = -.62$, $p < .01$, LFCC, $r = -.63$, $p < .01$, LPC, $r = -.52$, $p < .01$, MFCC, $r = -.53$, $p < .01$. Sympathische und unsympathische Sprecher unterschieden sich signifikant voneinander hinsichtlich der

Kennwerte F_0 , $t(151) = 7.47$, $p < .001$, F_1 - F_6 , $t(131) = 5.97$, $p < .001$, und $LFCC$, $t(107) = 4.55$, $p < .001$. Selbstsichere Sprecher unterscheiden sich signifikant von unsicheren Sprechern hinsichtlich der Kennwerte F_0 , $t(22) = 5.14$, $p < .001$, FBW_1 - 6 , $t(95) = 10.52$, $p < .001$, F_1 - F_6 , $t(23) = 2.15$, $p < .001$, HNR , $t(182) = 14.13$, $p < .001$, $LFCC$, $t(70) = 9.09$, $p < .001$, LPC , $t(93) = 10.93$, $p < .001$ und $MFCC$, $t(134) = 9.57$, $p < .001$.

Das Machine-Learning Prozedere ergab, dass für das Sprachmaterial „Sympathisch“ die 2-Nearest-Neighbors Strategie und für das Sprachmaterial „Selbstsicher“ die Strategie der Naiven Bayes die besten mittleren und die stabilsten Vorhersageergebnisse erzielt. Die Klassifikationsgenauigkeit der Sympathie-Vorhersage beträgt 62.75 %. Die vorhergesagten und gerateten Sympathiewerten stimmten nicht signifikant überein, $\chi^2(1, N = 51) = 1.78$, $p > .05$. Die Klassifikationsgenauigkeit der Selbstsicherheits-Vorhersage beträgt 76.47 %. Vorhergesagte und geratete Selbstsicherheitswerte stimmen signifikant überein, $\chi^2(1, N = 66) = 5.76$, $p < .05$.

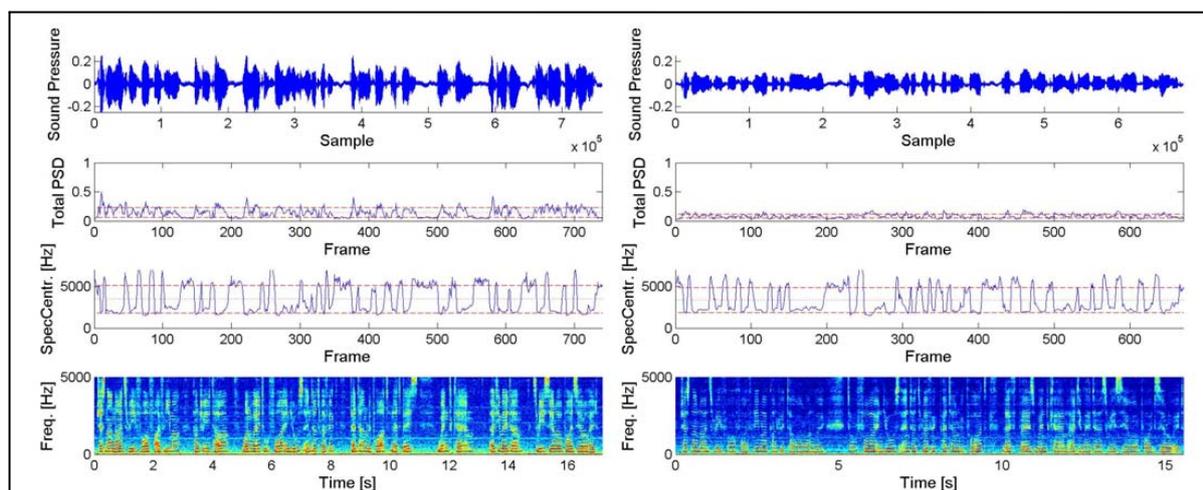


Abbildung 3: Spektrogramm zweier Stimmproben. Links ein selbstsicherer Sprecher und rechts ein unsicherer Sprecher. Dargestellt sind von oben nach unten Sound Pressure, Total Power Spectral Density (TPSD), Spectral Centroid (SpecCentr) und ein Spektrogramm. Hohe spektrale Leistungsdichten sind rot (im Schwarz-Weiß Druck: schwarz), niedrige hingegen blau (im Schwarz-Weiß Druck: grauwertig) kodiert.

4. Diskussion

Für das Sprachmaterial „Sympathisch“ erreichen die linearen Zusammenhänge zwischen drei sprachlichen Einzelmerkmalen (F_0 , F_1 - F_6 , $LFCC$) und gerateten Sympathiewerten das Signifikanzniveau. Die Auswertung der Machine-Learning gestützten Klassifikationsvorhersage ergab nur eine Vorhersagegenauigkeit von 62.75 %. Die Ergebnisse

des Chi²-Tests bestätigten die geringe Assoziation zwischen gerateten und vorhergesagten Werten. Aufgrund der Ergebnisse der t-Tests kann jedoch davon ausgegangen werden, dass sich sympathische von unsympathischen Sprechern hinsichtlich der akustischen Kennwerte F0, F1-F6, LFCC und MFCC signifikant voneinander unterscheiden. Für das Sprachmaterial „Selbstsicher“ erreichen die linearen Zusammenhänge zwischen sieben sprachlichen Einzelmerkmalen (F0, FBW1-6, F1-F6, HNR, LFCC, LPC, MFCC) und gerateten Selbstsicherheitswerten das Signifikanzniveau. Die Machine-Learning gestützten Klassifikationsvorhersage ergab eine Vorhersagegenauigkeit von 76.47 %. Der anschließend durchgeführte Chi²-Test bestätigte den Zusammenhang zwischen gerateten und vorhergesagten Werten. Aufgrund der Ergebnisse der t-Tests kann davon ausgegangen werden, dass sich selbstsichere von unsicheren Sprechern hinsichtlich der akustischen Kennwerte F0, FBW1-6, F1-F6, HNR, LFCC, LPC und MFCC signifikant voneinander unterscheiden.

Die auf der Ebene der physiologischen Veränderungen des Sprachproduktionsapparates begründeten Annahmen zu entspannungsinduzierten Wirkungen, konnten bestätigt werden. Abbildung 4 illustriert die Ergebnisse beispielhaft für selbstsichere Sprecher.

Verbesserungspotential birgt eine ausgeglichene Verteilung der Stimmproben hinsichtlich der Extrema, eine feinere Differenzierung der Ratings und eine größere Stichprobe, sowohl der Sprecher, als auch der Beurteiler. Weiterhin steht die Validierung der verwendeten Skala noch aus. Kritisch zu sehen ist außerdem die Verwendung geschaukelter Sprache.

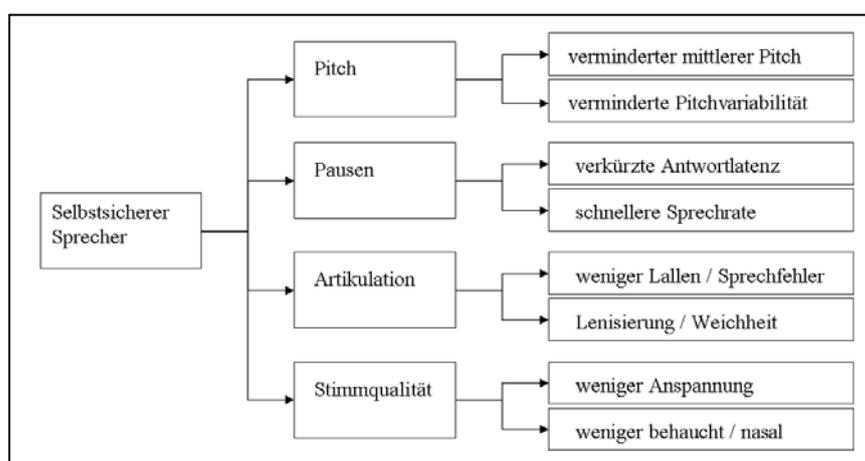


Abbildung 4: Selbstsicherheitsinduzierte auditive Stimmänderungen

Das anwendungsbezogene Transferpotenzial der vorliegenden Ergebnisse ist reichhaltig. Zum einen kann das verwendete Machine-Learning Prozedere auf andere Beurteilungsdimensionen

übertragen werden, zum anderen kann die akustische Stimmanalyse auch über die Grenzen des klassischen Interviews bzw. ACs hinaus eine sinnvolle Ergänzung darstellen. Hier wäre der Einsatz im Rahmen von Telefoninterviews, telefonischem Kontakt vor dem eigentlichen Auswahlprozess, videogeführten Interviews oder Video-Bewerbungen denkbar. Ebenfalls sinnvoll wäre der Einsatz der akustischen Stimmanalyse in Bewerbertrainings. Selbstverständlich ist nicht nur die Anwendung in der Personalauswahl sondern auch in der Personalentwicklung denkbar. Für letzteres spricht die beliebig häufige Wiederholbarkeit der Messung über einen langen Zeitraum, welche es ermöglicht, den Entwicklungserfolg abzubilden.

Auch diverse andere psychologische Anwendungssituationen können von der akustischen Stimmanalyse profitieren. Vorstellbar sind Einsätze in der Intelligenzdiagnostik, vor allem im Bereich der frühkindlichen Hochbegabtenidentifikation, in der Arbeits- und Organisationspsychologie im Kontext der Unfallprävention oder im Rahmen arbeitsgestalterischer Maßnahmen oder auch in der Psychotherapie.

Zusammenfassend eröffnen die zum großen Teil sehr ermutigenden Ergebnisse der akustischen Stimmanalyse in Bezug auf die Eignungsdiagnostik eine attraktive Zukunft. Die reale Anwendung ist ein erreichbares Ziel. Zuvor sollten aber Bemühungen unternommen werden, die Stichprobengröße zu erhöhen, um robustere Vorhersagen zu ermöglichen und einen Transfer auf möglichst viele Personengruppen zu gewährleisten (z.B. Nicht-Muttersprachler und Dialektträger).

Literatur

- Anderson, N. & Shackleton, V. (1990). Decision making in the graduateselection interview: A field study. *Journal of Occupational Psychology*, 63, 63-76.
- Batliner, A., Schuller, B., Schaeffler, S. & Steidl, S. (2008). Mothers, adults, children, pets - Towards the acoustics of intimacy. In ICASSP (Eds.), *International Conference on Acoustics, Speech & Signal Processing 2008* (pp. 4497–4500). Las Vegas, Nevada: University of Nevada.
- Huffcutt, A. I. & Arthur, W. (1994). Hunter and Hunter (1984) revised: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184-190.
- Huffcutt, A. I., Conway, J. M., Roth, P. L. & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897-913.
- Liebermann, P. & Michaels, S. (1962). Some aspects of fundamental frequency and envelope amplitude as related to emotional content of speech. *Journal of the Acoustical Society of America*, 76, 1346-1356.
- Neubauer, R. (2001). *Assessment Center-Studie 2001* [Internet]. Verfügbar unter: <http://www.arbeitskreis-ac.de/index.php/projekte/studien/ac-studie-2001-.html> [20. April 2010].
- Peters, L. H. & Terborg, J. R. (1975). The effects of temporal placement of unfavorable information and of attitude similarity on personnel selection decisions. *Organizational Behavior and Human Performance*, 13, 279-293.
- Scherer, K. R. (1974). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal Communication* (pp. 249-253). New York: Oxford University Press.
- Schuler, H. (2000). *Psychologische Personalauswahl*. Göttingen: Hogrefe.
- Schuller, B., Vlasenko, B., Arsic, D., Rigoll, G. & Wendemuth, A. (2008). Combining speech recognition and acoustic word emotion models for robust text-independent emotion recognition. In ICME (Eds.), *International Conference on Multimedia & Expo 2008* (pp. 2249-2252). Hannover: University of Hannover.