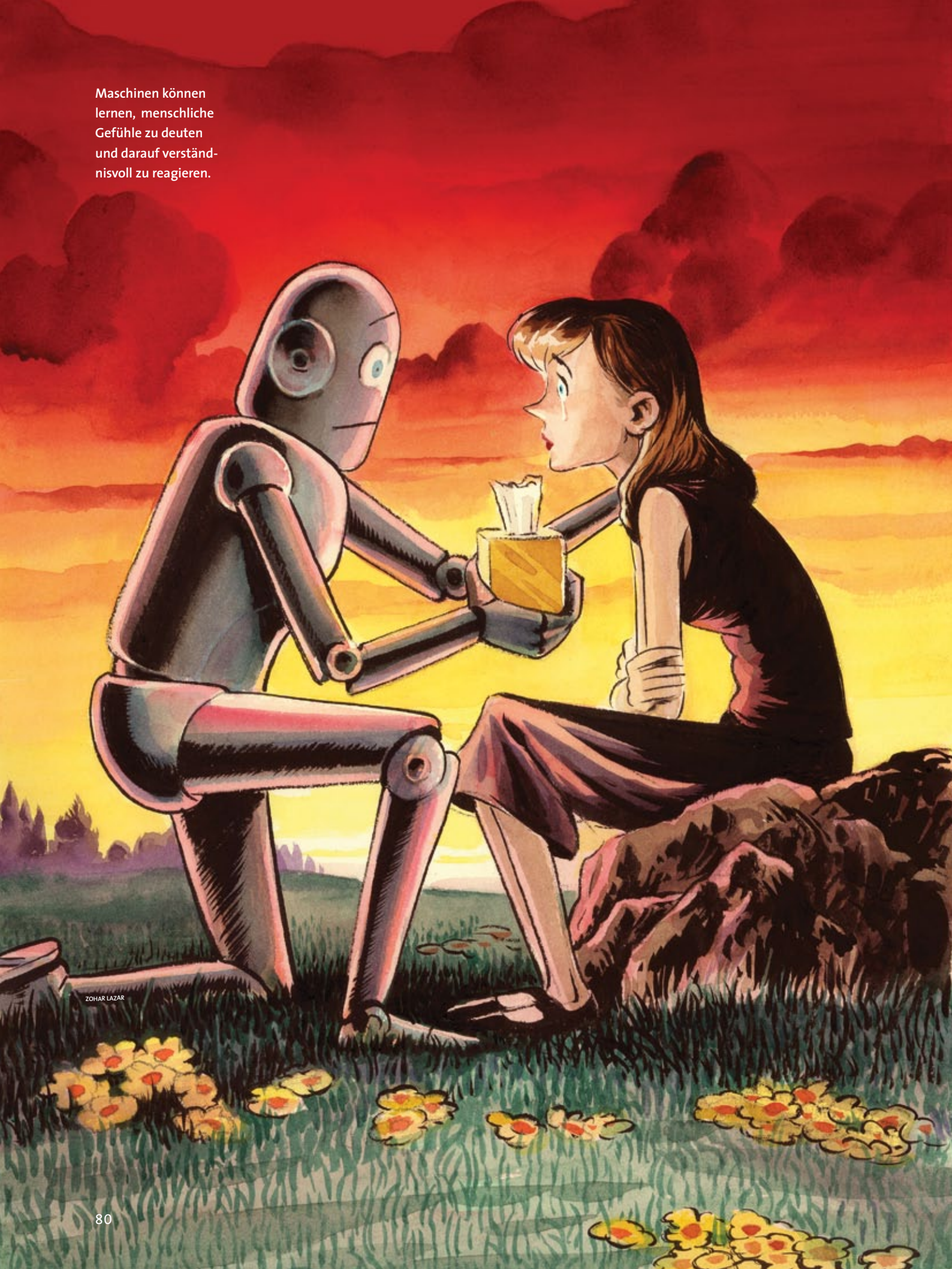

Maschinen können lernen, menschliche Gefühle zu deuten und darauf verständnisvoll zu reagieren.

INFORMATIK

Roboter mit Gefühlen

Bald schon könnten uns Roboter aller Art ähnlich vertraut sein wie Smartphones und Tablets heute. Um uns das Kommunizieren mit »intelligenten« Maschinen zu erleichtern, bringen Forscher ihnen bei, menschliche Emotionen zu verstehen und zu simulieren.

Von Pascale Fung

Den ersten einfühlsamen Satz sprach eine Maschine vermutlich mit den Worten: »Leider habe ich Sie nicht verstanden.« Ende der 1990er Jahre bot das Bostoner Softwareunternehmen SpeechWorks International seinen Firmenkunden Programme an, die ein paar höfliche Phrasen gebrauchten. Seither haben wir uns angewöhnt, mit Maschinen zu reden. Fast jeder Anruf bei einer Kundendienststelle landet zunächst bei einem Roboter. Hunderte Millionen Menschen sind mit einem digitalen Mädchen für alles (englisch »intelligent personal assistant«) unterwegs. Man kann beispielsweise die Apple-Software Siri (Speech Interpretation and Recognition Interface) mündlich auffordern, ein Restaurant zu finden, einen Freund anzurufen oder ein bestimmtes Musikstück abzuspielen. Solche Programme können menschliches Verhalten oft geradezu unheimlich gut simulieren.

Aber nicht immer reagieren Maschinen wunschgemäß. Die Spracherkennungssoftware missversteht oft die Absicht der Frage; sie ist unempfänglich für Humor, Sarkasmus und Ironie. Wenn wir künftig routinemäßig mit intelligenten Staubsaugern oder menschenähnlichen Pflegekräften kommunizieren wollen, müssen die Roboter nicht bloß gesprochene Worte verstehen, sondern auch Emotionen – sie sollten Einfühlungsvermögen besitzen.

In meinem Labor an der Hong Kong University of Science and Technology entwickeln wir solche Maschinen. Einfühlsame Roboter werden freundliche Gefährten sein, die unsere physischen und emotionalen Bedürfnisse umso besser vorausahnen, je öfter sie mit uns kommunizieren. Sie werden sich für ihre Fehler entschuldigen und uns vor einer Aktion um Erlaubnis fragen. Sie werden nicht nur Senioren pflegen und Kinder unterrichten, sondern sich in kritischen Situationen sogar selbstlos opfern, um Leben zu retten.

AUF EINEN BLICK

COMPUTER MIT EMOTIONALER INTELLIGENZ

1 Je mehr wir uns angewöhnen, Maschinen mit Sprache und Gesten zu steuern, desto stärker wird der Wunsch, dass sie auch **Andeutungen** und **Gefühle** richtig auslegen.

2 Damit ein digitales Gerät Humor, Ironie oder Missstimmung versteht, muss es mit einem **Empathiemodul** ausgestattet werden. Eine entsprechende Software erkennt Emotionen an Gesichtsausdruck und Sprechverhalten des menschlichen Gegenübers.

3 Die Entwicklung einfühlsamer Roboter steckt noch in den Kinderschuhen, doch erste Prototypen verwenden bereits **lernfähige Algorithmen**, um menschliche Stimmungen und Emotionen immer treffsicherer zu deuten.

Einige Roboter mit simulierten Gefühlen sind schon auf dem Markt – zum Beispiel der kleine Android Pepper, den die französische Firma Aldebaran Robotics für das japanische Unternehmen Softbank Mobile gebaut hat. Oder Jibo, ein knapp drei Kilogramm schwerer persönlicher Assistent; diesen Roboter konstruierte ein Ingenieurteam um Roberto Pieraccini, der zuvor die Abteilung für Dialogtechnik bei SpeechWorks geleitet hatte. Die Entwicklung solcher Maschinen steckt zwar noch in den Kinderschuhen, aber das wird sich bald ändern.

Das Empathiemodul

Ich selbst begann mich für die Konstruktion empathischer Roboter 2009 zu interessieren, als mein Team das erste chinesische Gegenstück zu Siri entwarf. Mich faszinierte, wie selbstverständlich die Nutzer begannen, emotional auf intelligente Assistentensysteme zu reagieren – und wie ungeduldig sie wurden, wenn ihre Geräte sie nicht verstanden. Offen-

bar hängt der Erfolg von Spracherkennungsalgorithmen wie jenen ab, an denen ich im Lauf meiner 25-jährigen Laufbahn gearbeitet habe.

Jede intelligente Maschine ist im Grund ein Softwaresystem, das aus Modulen besteht – aus Programmen, die jeweils auf eine Aufgabe spezialisiert sind. Ein Modul übernimmt beispielsweise die akustische Sprachverarbeitung, ein anderes die visuelle Bilderkennung und so weiter. Ein einfühlsamer Roboter besitzt ein spezielles Empathiemodul. Es analysiert nicht nur den Inhalt des Gesprochenen, sondern auch die Prosodie – die Sprachmelodie – und die Miene des Sprechers, um eine Antwort zu finden, die zu den darin ausgedrückten Gefühlen passt.

Wenn zwei Menschen kommunizieren, nutzen sie automatisch gewisse Hinweise auf den emotionalen Zustand des Gegenübers. Sie deuten Gesichtszüge und Körpersprache, sie nehmen Änderungen der Stimmlage wahr, sie verstehen sprachliche Andeutungen. Beim Bau eines Empathiemoduls gilt es, zunächst diejenigen Charakteristika menschlicher Kommunikation herauszufinden, an denen ein Computer Gefühle zu erkennen vermag, und dann Algorithmen darauf zu trainieren, diese Charakteristika zu entdecken.

Um Maschinen den emotionalen Gehalt von Sprache zu erschließen, machten wir uns daran, ihnen die dem zu Grunde liegenden akustischen Regeln beizubringen, zusätzlich zu den Wortbedeutungen. Denn so funktioniert eben menschliche Kommunikation. Wenn wir fröhlich sind, sprechen wir schneller und mit höherer Stimme, unter Stress hingegen eintöniger und nüchterner. Mit intelligenter Signalverarbeitung vermag ein Computer solche akustischen Anzeichen zu entdecken – ähnlich wie ein Lügendetektor, der Stress physiologisch anhand von Blutdruck, Puls und elektrischer Leitfähigkeit der Haut misst. An meinem Institut, das die Studenten scherzhaft »Universität für Stress und Spannung« nennen, haben wir Lernalgorithmen gezielt darauf trainiert, akustische Stresssymptome zu identifizieren. Zu diesem Zweck stellten wir den Studenten zwölf zunehmend stressende Fragen, nahmen ihre Antworten auf und sammelten auf diese Weise rund zehn Stunden natürlicher Stresssymptome in den Sprachen Englisch, Mandarin und Kantonesisch. Unsere lernfähigen Algorithmen konnten anhand dieses Materials schließlich in 70 Prozent aller Fälle Stress richtig erkennen – ähnlich gut wie menschliche Zuhörer.

Gespür für Zwischentöne

Unterdessen brachte ein anderes Team Maschinen bei, die Gefühlsstimmung eines Lieds nur anhand der Musik – ohne Beachtung des Textes – zu erkennen. Im Gegensatz zu einer vorübergehenden Emotion charakterisiert die Stimmung das gesamte Musikstück. Die beteiligten Forscher sammelten zunächst 5000 Lieder aller Art in den wichtigsten europäischen und asiatischen Sprachen. Musikexperten hatten zuvor einige hundert dieser Stücke in 14 Stimmungskategorien klassifiziert.

Wir bestimmten in jedem Lied elektronisch rund 1000 akustische Signaleigenschaften – Parameter wie Energie, Grundfrequenz und Harmonien – und trainierten mit der Musik 14 unterschiedliche, jeweils für eine spezielle Stimmung zuständige Programme, so genannte Klassifizierer. Zum Beispiel reagiert ein Klassifizierer nur auf fröhliche, ein anderer nur auf traurige Musik. Die 14 Programme arbeiten zusammen, indem jeder das Ergebnis der anderen berücksichtigt. Wenn ein »fröhlicher« Klassifizierer irrtümlich ein trauriges Lied fröhlich findet, wird er in der nächsten Lernrunde korrigiert. In jeder Runde lernt der jeweils schwächste Klassifizierer hinzu, und so wird das gesamte System klüger. Indem die Maschine vielen Musikstücken lauscht, lernt sie, welches Stück zu welcher Stimmung gehört, und kann mit der Zeit wie ein menschlicher Musikliebhaber durch bloßes Zuhören die Stimmung jedes Lieds angeben.

Auf Basis dieser Forschung haben frühere Studenten und ich die Firma Ivo Technologies gegründet, die einfühlsame Maschinen für den häuslichen Gebrauch entwickelt. Das erste Produkt namens Moodbox wird ein intelligentes Multimediagerät sein, das in jedem Zimmer Musik und Beleuchtung den Emotionen des Nutzers anpassen kann.

Um allerdings Humor, Sarkasmus, Ironie und andere komplexe Eigenschaften menschlicher Kommunikation zu verstehen, muss eine Maschine neben den rein akustischen Signalen auch verborgene emotionale Bedeutungen der Wörter und Sätze erkennen. Zunächst konnten wir für unsere Analyse auf bereits vorhandene Algorithmen zurückgreifen, die den Gefühlsgehalt von schriftlichen Onlinekommentaren erfassen. Solche Lernalgorithmen suchen im Text nach verräterischen Schlüsselworten wie »Sorge« oder »Furcht« und schließen daraus auf Einsamkeit. Wiederholter Gebrauch von Slangphrasen wie »c'mon« – etwa »los«, »mach schon« – verleiht wiederum Popsongs einen energischen Charakter.

In unserem Bemühen, die Stimmung eines Lieds zu vermitteln, trainierten wir entsprechend die Algorithmen, nicht nur in der Musik, sondern auch im Text emotionale Hinweise aufzuspüren. Wir entnahmen dem Liedtext Wortketten – so genannte n-Gramme – und fütterten damit Textklassifizierer, die jeweils für eine der 14 Stimmungen zuständig sind. Die Wortketten markierten wir so, dass der Algorithmus sie als Teil des Liedtexts erkennt, den er klassifizieren soll. Computer können aus n-Grammen und Textteilmarkierungen statistische Näherungen für die grammatischen Regeln einer beliebigen Sprache bilden; mit diesen Regeln erkennen Programme wie Siri gesprochene Inhalte, und eine Software wie Google Translate übersetzt den Text in eine andere Sprache.

Verständnisvolle Roboter

Sobald eine Maschine den Inhalt versteht, kann sie ihn mit der Art und Weise vergleichen, in der er gesprochen wird: Antwortet jemand auf eine Frage sicher und deutlich oder zögernd, stockend und ausweichend? Sind die Antworten ausführlich und detailliert oder kurz angebunden? Wenn eine Person seufzt und sagt »Ich bin ja so froh, dass ich am Wochenende arbeiten muss«, dann entdeckt der Algorithmus den Widerspruch zwischen Emotion und Inhalt und rechnet die Wahrscheinlichkeit dafür aus, dass der Satz ironisch gemeint ist.

Außerdem kann die Maschine durch Vergleich des Gesprochenen mit anderen Informationen komplexere Absichten entdecken. Wenn jemand sagt »Ich habe Hunger«, ermittelt der Roboter die beste Antwort, indem er unter anderem Ort, Tageszeit und frühere Vorlieben des Nutzers berücksichtigt. Sind die beiden zur Mittagszeit daheim, könnte der Roboter antworten: »Soll ich dir ein Sandwich machen?« Wenn Roboter und Nutzer gerade unterwegs sind, fragt die Maschine vielleicht: »Soll ich ein Restaurant suchen?«

Anfang 2015 kombinierten Forscher in meinem Labor alle diese Module für Spracherkennung und Einfühlung, um einen Prototyp zu schaffen, den wir Zara nennen. Das Lernmaterial für Zaras Training umfasste hunderte Stunden akustischer Daten, aber heute läuft das Programm auf einem einzigen Desktop-Computer. Zara ist derzeit ein virtueller Roboter, der als Comicfigur auf dem Bildschirm agiert.

Wenn man eine Unterhaltung mit Zara beginnt, sagt sie: »Bitte warte, während ich dein Gesicht analysiere.« Zaras Algorithmen bestimmen mit Hilfe der Computerkamera Geschlecht und ethnische Zugehörigkeit des Gesprächspartners. Dann testet sie, welche Sprache er spricht – Zara versteht Englisch, Mandarin und ein wenig Französisch – und stellt ein paar Fragen und Aufgaben: »Was ist deine früheste Erinnerung?«, »Erzähl mir von deiner Mutter!«, »Wie war dein letzter Urlaub?«, »Erzähl mit eine Geschichte mit einer Frau, einem Hund und einem Baum!« Auf Grund des Gesichtsausdrucks des Gegenübers, seiner Stimmeigenschaften und des Inhalts der Antworten lernt Zara, sich auf eine Weise zu unterhalten, die Empathie imitiert. Nach fünf Minuten Zwiesprache versucht Zara dann, die Persönlichkeit des Gesprächspartners zu erraten und fragt ihn, was er von einfühlsamen Maschinen hält. Auf diese Weise sammeln wir Informationen über die Interaktion zwischen Menschen und Frühformen empathischer Roboter.

Zara ist ein Prototyp, doch da sie auf Lernalgorithmen beruht, wird sie mit jedem Gespräch klüger und einfühlsamer. Gegenwärtig beruht ihre Datenbasis nur auf Ergebnissen aus Interaktionen mit Studenten meines Labors. Demnächst soll Zara einen Körper bekommen, indem wir sie in einem menschenähnlichen Roboter installieren.

Natürlich liegt das Zeitalter freundlicher Roboter noch in weiter Ferne. Wir entwickeln gerade erst die primitivsten Voraussetzungen für Maschinen mit emotionaler Intelligenz. Aber auch von Zaras verbesserten Nachkommen sollten wir keine Perfektion erwarten. Ich halte das gar nicht für erstrebenswert. Wichtig ist, dass unsere Maschinen den Menschen ähnlicher werden – und die sind bekanntlich nicht vollkommen. ~

DIE AUTORIN

Pascale Fung ist Professorin für Elektronik und Computertechnik an der Hong Kong University of Science and Technology. Als Spezialistin für Mensch-Maschine-Kommunikation wurde sie zum Mitglied des Institute of Electrical and Electronics Engineers (IEEE) sowie der International Speech Communication Association (ISCA) gewählt.

QUELLEN

Su, D. et al.: Multimodal Music Emotion Classification using AdaBoost with Decision Stumps. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013), Vancouver, Mai 2013

Zuo, X. et al: A Multilingual Database of Natural Stress Emotion. In: Eighth International Conference on Language Resources and Evaluation (LREC 2012), Instanbul, Mai 2012

Dieser Artikel im Internet: **www.spektrum.de/artikel/1382049**