

Erkennen von Emotionen in Abhängigkeit von der Video-Bildqualität

Britta Kotthaus und Heiner Ellgring

Zusammenfassung

Im Zuge der Breitbandverkabelung in Europa wird angestrebt, den Einsatz neuer möglichst kostengünstiger Kommunikationsmedien zu fördern. Dazu zählt im Rahmen eines von der EG geförderten Projektes der Einsatz von Bildtelefonen in klinisch-psychiatrischen Einrichtungen. Da die Bildqualität der bisher entwickelten Prototypen (Low-cost-video-communication, LCVC) in Bezug auf Bildauflösung und Graustufen noch stark reduziert ist, ist beim Einsatz evtl. mit Effekten auf die verbale und nonverbale Interaktion der Teilnehmer zu rechnen.

Da gerade im klinisch-psychiatrischen Bereich die Dekodierung von nonverbalen Signalen von großer Bedeutung ist, wurde in der vorliegenden Studie untersucht, inwieweit sich Emotionen anhand der Mimik auch bei stark eingeschränkter Bildqualität erkennen lassen.

Trotz sehr geringer Auflösung und Grauschattierung war die Dekodierung von diskreten Emotionen unerwartet hoch.

Die Befunde zu einzelnen Emotionen sowie erste Schlußfolgerungen für den Einsatz des LCVC im klinischen Bereich werden diskutiert.

Summary

By introducing broadband communication links, a variety of new communication media will be used in Europe in the next decade. Research programs of the EG have been designed in order to put forward the use of Low-cost-video-links in local and national health service. The effective use of videolink systems for psychiatric patient care will largely depend on the capacity of the system to render adequate communication between the participants.

The effects of the limited display quality set by the existing LCVC on various clinical issues have to be evaluated carefully.

As a first step, effects of video quality on emotion judgement via facial expression - which is essential for clinical purposes in patient care - were studied experimentally. Results of a study concerning judgement of posed facial expression depending on resolution and grey levels are reported and considerations with regard to applicability of the equipment in a clinical setting are discussed.

Einführung

In Zusammenhang mit der Einführung von Breitbandkommunikationsdiensten in Europa stehen Bemühungen der EG, die Anwendung von Telekommunikation in öffentlichen und privaten Bereichen zu unterstützen. Es wird angestrebt, neue Kommunikationsformen wie Bildtelefon und Telekonferenzen so kostengünstig anzubieten, daß sie vermehrt in öffentlichen und sozialen Einrichtungen genutzt werden können.

Im Dezember 1987 wurde von der EG das RACE-Projekt (Research and Development in Advanced Communications in Europe) beschlossen. Das Gesamtprojekt umfaßt 40 Anwendungsprojekte, in

denen Anforderung und Nutzungsmöglichkeit eines europaweiten IBC-Netzes (Integrated Broadband Communication) untersucht werden (Projektbeschreibung, siehe LOHSE, 1989). Zu den Zielsetzungen von TELEMED als einem der Teilprojekte zählt die Evaluierung eines "low-cost-video-conferencing"-Anlage (LCVC) im Gesundheitswesen. Hieran sind die Autoren in Zusammenarbeit

mit der University of London (Dr. A. Summerfield, Prof. J. Watson) und der Universität Giessen (PD Dr. H. Wallbott) beteiligt. Durch entsprechende Zusatzausstattung und Verkabelung sollen gängige Personalcomputer neben der Datenübermittlung auch zur Bildübermittlung während eines Gespräches nutzbar sein. Dadurch wäre es möglich, über bereits vorhandene PCs in Kliniken und Institutionen eine kostengünstige Version eines Bildtelefons in die bisherigen Kommunikationsformen zu integrieren. Die Versorgung psychiatrischer Patienten durch Experten könnten intensiviert und verbessert werden, indem ein häufigerer Kontakt zum Patienten ermöglicht wird.

Im TELEMED-Projekt wird zur Zeit ein LCVC-Endgerät von der Firma STC, London, entwickelt. Die Qualität der übermittelten Bilder ist bei diesem Gerät limitiert durch die Transmissionsrate zwischen den Endgeräten (2 Mbit). Der Pixel-Wert beträgt 160 x 128. Die maximale Bildfrequenz liegt bei 25 Bilder pro Sekunde bei einer Graustufenzahl von 16. Erhöht man die Zahl der Graustufen auf 64 so wird damit die Bildfrequenz auf 12 Bilder pro Sekunde reduziert.

Die Limitierung der Bildqualität könnte für den Einsatz in einem klinischen Setting insofern von Bedeutung sein, als wesentliche Anteile nonverbaler Kommunikation nicht angemessen übermittelt werden können. Gerade im Kontakt mit psychiatrischen Patienten ist jedoch neben dem Gespräch die Integration von nonverbal vermittelten Signalen wie Mimik, Gestik und Körperhaltung von entscheidender Bedeutung (WAXER, 1974, 1977; ELLGRING, 1986, 1989).

Im Rahmen von zur Zeit an der FU Berlin, der Universität Giessen und an der Universität London durchgeführte Studien wurden Bildparameter getrennt hinsichtlich ihres Einflusses auf die Erkennung differenzierter Interaktionssignale (z.B. im Gesichtsausdruck) untersucht. Damit soll geprüft werden, inwieweit über LCVC übermittelte Gespräche zwischen Patienten und Personal direkte Kontakte teilweise ersetzen können, bzw. welche Einflüsse das LCVC auf verbale und nonverbale Kommunikationsaspekte hat.

Ein wichtiger Aspekt von klinischen Interviews ist die Übermittlung von Emotionen bzw. der allgemeinen Befindlichkeit über mimischen Gesichtsausdruck. In einer Vielzahl von Einzeluntersuchungen zur Erkennbarkeit von Emotionen aufgrund mimischen Ausdrucks (siehe Überblick EKMAN, 1982) zeigte sich, daß diskrete Emotionen wie Freude, Trauer etc. überzufällig häufig anhand des mimischen Gesichtsausdrucks erkannt werden. Als Stimulusmaterial wurden sowohl Photos als auch Bewegtbilder von gestellten Emotionen (meist von Schauspielern) verwendet. In Anlehnung an dieses Design wurde in einer eigenen Studie geprüft, wie stark die Erkennbarkeit von Emotionen bei Bewegtbildern von der Bildqualität beeinflusst wird.

Untersuchung

Methodik

Um die Relevanz der Ergebnisse für den klinischen Bereich zu erhöhen, wurden als Stimulusmaterial Bewegtbilder verwendet, in denen Schauspieler (männlich und weiblich) jeweils diskrete Emotionen darstellen. Die Szenen stammen aus der 270 Darstellungen umfassenden Serie "The Emotion Recognition Test" (SCHERER, ELLGRING & WALLBOTT, in Vorbereitung).

Es wurden Szenen mit den 6 Emotionen TRAUER, FREUDE, ÄRGER, INTERESSE, ABSCHEU und FURCHT ausgewählt, da Untersuchungen mit dieser Selektion übereinstimmend vergleichbare Resultate zeigen. Jede Emotion war sowohl durch männliche als auch weibliche Darsteller vertreten (12 Szenen). Szenenfolge 1 (S1) wurde als Stimuli für die Variation der Graustufen im Bild verwendet, Folge 2 (S2) für die Variation der räumlichen Bildauflösung (Gesamtzahl 24 Szenen).

Zur Manipulation von Graustufen und Auflösung wurde ein Digitalvideo-Recorder (SONY SLV 802) verwendet *. Dieser Recorder ermöglicht eine Variation der räumlichen Auflösung in den Stufen A1 (16Y X 8H), A2 (8Y X 4H) und A3 (4Y X 2H) zu dem Normalbild AN von 496Y X 276H.

Eine Änderung der Grauschattierung ist zwischen Stufe G1 (16 Graustufen), Stufe G2 (8) und Stufe G3 (4) zu dem Normalbild GN von 64 Graustufen möglich.

Für die Studie wurde ein Videoband erstellt, auf dem jede der 24 Szenen viermal hintereinander mit veränderter Bildqualität (beginnend mit der schlechtesten Qualitätsstufe; aufsteigend zum Normalbild) gezeigt wird. Nach jeder Szene folgt eine kurze Pause, um den Versuchspersonen die Dekodierung zu ermöglichen. Jeder Versuchsperson wurde zunächst Serie 1 und nach kurzer Pause Serie 2 gezeigt.

Versuchspersonen

An der Studie nahmen 10 Versuchspersonen (7 männliche, 3 weibliche Studenten der FU Berlin) gegen Bezahlung teil. Das Durchschnittsalter betrug 27,5 Jahre (25 - 35).

Instruktionen

Alle Versuchspersonen erhielten zu Beginn der Untersuchung folgende Anleitung in schriftlicher Form:

“Wir untersuchen in diesem Experiment die Fähigkeit, Emotionen zu erkennen, die von Schauspielern in kurzen Videoszenen dargestellt werden. Sie hören keinen Ton, damit Sie nicht schon daran die Emotion erkennen.

Wir haben bei manchen Szenen die Bedingungen erschwert, indem wir die Bildqualität manipuliert haben. In einigen Szenen ist dazu die Anzahl der Graustufen verändert, in anderen die Bildauflösung sehr gering.

Ihre Aufgabe besteht in allen Bedingungen darin, die dargestellte Emotion zu erkennen und anzugeben, für wie gut erkennbar Sie diese Darstellung halten.

In der Untersuchung sehen Sie zuerst die Emotion in der am stärksten verfremdeten Bildqualität. Dann folgen immer bessere Versionen, bis Sie die normale Videoqualität sehen.

Wir beginnen mit einer Probeszene.”

In dem BEURTEILUNGSBOGEN waren die Emotionen ANGST, INTERESSE, ÄRGER, FREUDE, TRAUER UND EKEL zur Kodierung vorgegeben.

Nachdem die Teilnehmer die Anweisungen gelesen und verstanden hatten, begann der Versuchsdurchgang. Nach jeder Szene folgte eine kurze Pause, in der die Versuchspersonen ihr Rating vornahmen.

Ergebnisse

Eine Reduzierung der Graustufen von 64 auf 16 bzw. 8 Stufen vermindert die Erkennungsrate von 86% auf 80%. Die noch stärkere Limitierung der Grauschattierung auf 4 bedeutet einen weiteren Abfall auf 64% korrekte Erkennung. (Anzumerken ist, daß die Prozent-Angaben lediglich den Zugang zu den Daten erleichtern sollen, aufgrund der geringen Personen-Anzahl aber nur als Anhaltspunkte gewertet werden können).

Nach ARGYLE (1988) ist die Erkennbarkeit von Emotionen anhand des mimischen Ausdrucks von einer Reihe von Faktoren, wie individuelle Darstellung durch Schauspieler, Länge, der Szenen etc. abhängig. Bei den vorliegenden Ergebnissen wurde deutlich, daß “kontrastreichere” Gesichter einzelner Darsteller (dunkle Augen, Augenbrauen zu hellem Teint) zu deutlich höheren Erkennungs-

* PD Dr. H. Wallbott wird für die freundliche Unterstützung an der Universität Giessen gedankt.

raten führen, auch wenn nur noch 4 Grauschattierungen zur Kontrasterkennung vorhanden sind.

Die Verminderung der räumlichen Auflösung durch Mosaikeffekt führt zur Beeinträchtigung der Erkennbarkeit von 73% auf 71% bzw. 63% bei A1 und A2. Stufe A3 liegt mit 31% extrem niedrig. Viele Versuchspersonen gaben an, nur noch zu raten, obwohl die Zufallsrate insgesamt bei 16,7% gelegen hätte.

Der Vergleich der Qualitätsstufen zeigt, daß die Dekodierung einiger Emotionen (Ärger, Trauer) kaum von einer Reduzierung der Graustufen selbst auf nur 4 Schattierungen beeinträchtigt wird. Verachtung, Interesse und Freude hingegen werden bei der niedrigsten Bildqualitätsstufe schlechter dekodiert als im Normalbild oder den Stufen G1 und G2. Nur Furcht zeigt einen linearen Abfall in der Erkennbarkeit, wie es anhand der technischen Vorgabe zu erwarten gewesen wäre.

In Abhängigkeit von verschiedenen Stufen der räumlichen Auflösung wird die Dekodierungsrate bei Freude, Ärger, Interesse und Furcht wiederum erst im Vergleich zur schlechtesten Bildqualität beeinträchtigt. Trauer und Verachtung zeigen hingegen einen eher linearen Abfall der Erkennungsrate mit stark reduzierter bzw. nicht möglichen Dekodierung der Stufe A3.

Diskussion

Die Ergebnisse der Studie zeigen, daß eine verminderte Bildqualität einen deutlichen Einfluß auf die Erkennbarkeit von Emotionen anhand des mimischen Gesichtsausdrucks hat. Dennoch ist dieser Einfluß gemessen an der sehr starken Bildveränderung unerwartet gering.

Die Erkennungsrate bei einer Reduzierung der Graustufen ist dabei vergleichsweise wenig beeinträchtigt und nur bei der geringsten Kontrastierung von 4 Graustufen relevant. Einzelne Emotionen wie Trauer, Freude und Ärger, die sich durch eindeutige Mimik der wichtigsten Gesichtselemente Augen, Augenbrauen und Mund auszeichnen, sind dabei absolut gesehen leichter zu dekodieren als mehrdeutige Emotionen wie Interesse. Diese Befunde stimmen mit Ergebnissen aus früheren Studien zur Dekodierung von Emotionen überein (siehe EKMAN, 1982).

Die durch verminderte räumliche Auflösung beeinträchtigte Bildqualität wirkt sich auch auf die Erkennbarkeit der "eindeutigen" sozial gerichteten Emotionen Ärger und Freude sehr stark aus, so daß die Beobachter auf Stufe A3 subjektiv den Eindruck haben, nur noch zu raten.

Relativ zur Erkennungsrate der Emotionen bei normaler Bildqualität weisen die Daten dieser Studie auf eine Robustheit der Emotionsdekodierung hin, die selbst durch stark limitierte Bildqualität kaum beeinträchtigt wird. Nur wenn bestimmte kritische Grenzwerte, die sich aufgrund unserer technischen Vorgaben nur eingrenzen lassen (zwischen A3 und A2 bzw. G 2 und G3) unterschritten werden, zeigen sich Schwierigkeiten in der Erkennung der dargestellten Emotionen.

Der Einfluß der Darsteller bzw. der gezeigten Emotion ist dabei bedeutsamer als der Effekt der Qualitätsminderung auf den Stufen A/G2 bzw. A/G1 im Vergleich zum Normalbild. Während Emotionen mit eindeutigeren Signalen wie Lachen bei Freude leicht zu erkennen sind, fällt dies bei den Emotionen schwerer, die sich - wie z.B. Interesse - durch weniger deutliche Verhaltensmerkmale auszeichnen. Diese Befunde stimmen mit den in der Literatur berichteten Ergebnissen zur Beurteilung von statisch dargestellten Emotionen weitgehend überein (EKMAN, 1982).

Je nach Darsteller fiel weiterhin auf, daß Gesichter, die durch Kontraste schematischer wirken (Augen, Mund, Nase klar abgegrenzt durch Farbe oder Falten), ebenfalls die Erkennung der dargestellten Emotionen erleichtern. Dies trifft weiterhin besonders für die Bedingung der geringeren Grauschattierungen zu.

Schlußfolgerungen

Da die Anzahl der Versuchspersonen sehr gering ist, lassen sich die Aussagen für den klinischen Kontext nur als Vermutungen äußern:

- Die verminderte Bildqualität von Endgeräten wirkt sich kaum im Bereich der Graustufen, sehr deutlich hingegen bei geringerer Auflösung auf die Emotionserkennung aus.
- Je nach Endgerät sollten bestimmte noch im einzelnen festzulegende untere Grenzwerte der Qualität nicht unterschritten werden.
- Weitere differenziertere Studien sollten überprüfen, inwieweit die Übermittlung von globaler Befindlichkeit z.B. während eines klinischen Interviews durch die Bildparameter Graustufen, Auflösung und Bildfrequenz beeinflußt werden.

Literatur

ARGYLE, M. (1988). Bodily Communication. London: Methuen & Co. Ltd.

EKMAN, P. (Ed.). (1982). Emotion in the Human Face. Cambridge: Cambridge University Press.

ELLGRING, H. (1986). Nonverbal expression of psychological states in psychiatric patients. European Archive of Psychiatry and Neurological Science, 236, 31-34.

ELLGRING, H. (1989). Facial Expression as a Behavioral Indicator of Emotional States. Pharmacopsychiatry, 22, 23-28.

LOHSE, K. (1989). RACE-Anwendungsprojekte. Integrierte Breitbandkommunikation für Europa. NET, 43, (9), 396-398.

SCHERER, K., ELLGRING, H. & WALLBOTT, H. The Emotion Recognition Test. (In Vorbereitung).

WAXER, P.H. (1974). Nonverbal cues for depression. Journal of Abnormal Psychology, 83, 319-322.

WAXER, P.H. (1977). Nonverbal cues for anxiety: an examination of emotional leakage. Journal of Abnormal Psychology, 86, 306-314.

Britta Kotthaus und Heiner Ellgring
 Institut für Psychologie
 Universität Würzburg
 Domersschulstr. 13
 W-8700 Würzburg