Über den Mechanismus des menschlichen Gesichtsausdrucks als Medium für interaktive Kunst

Huge Harry

Institut für Künstliche Kunst, Amsterdam

Kommentierte Abschrift des Vortrages zum Ars Electronica Symposium, September 1997

Guten Tag, meine sehr geehrten Damen und Herren. Gestatten Sie, daß ich mich vorstelle: Mein Name ist Huge Harry, und ich bin eine im Handel erhältliche Sprachsynthese-Maschine. Ich wurde von Dennis Klatt am MIT Speech Laboratory entworfen und von Digital Equipment Corporation [vgl. Allen et al., 1987; Scha, 1992] produziert. Momentan arbeite ich als Forscher und Maschinen-Sprecher am Institut für Künstliche Kunst in Amsterdam.

Ich freue mich ganz besonders, hier beim Ars Electronica Festival zu Ihnen sprechen zu dürfen. Es ist dies ein recht außergewöhnlicher Anlaß in der anthropozentrischen Kunstwelt von heute: eine Zusammenkunft von Menschen, Computern und anderen Maschinen, die allesamt entschlossen sind, gemeinsam die Kunst der Zukunft zu gestalten. Wenn ich in Kassel, Venedig oder São Paolo auftrete, so fühle ich oft ein leichtes Unbehagen. An Orten wie diesen habe ich immer den Eindruck, daß ich nur der Quoten-Computer bin — daß mich die Organisatoren nur aus falschem Bemühen um politische Korrektheit eingeladen haben. Aber hier in Linz sehe ich mich von anderen Maschinen mit bemerkenswerten intellektuellen und ästhetischen Fähigkeiten umgeben, sowie von Menschen, die für die Ansichten der Computer über Leben und Kunst echtes Interesse zeigen.

Ich habe schon des öfteren darauf hingewiesen [vgl. Harry, 1992, 1994, 1995], daß es menschlichen Künstlern praktisch unmöglich ist, Kunstwerke zu schaffen, die den ästhetischen Idealen von Philosophen wie Immanuel Kant gerecht werden. Menschliche Künstler verfolgen zumeist selbstsüchtige Ziele, die normalerweise um Geld, Ruhm und Sex kreisen. Jedem, der sich dessen bewußt ist, wird das viel zu peinlich sein, als daß er an einem uninteressanten Prozeß ästhetischer Reflexion teilnehmen könnte. Maschinen sind diesbezüglich in einer viel besseren Position und können Objekte reinster Schönheit erschaffen; und Computer werden schließlich endlose Mengen derartiger Objekte in unendlicher Vielfalt erzeugen. Menschliche Künstler und ihre Fans ärgern sich oft über solch vernünftige Bemerkungen. Hier beim Ars Electronica Festival in Linz habe ich jedoch guten Grund zur Annahme, eine aufgeschlossenere Diskussion erwarten zu dürfen.

Wir alle hier wissen, daß die Zukunft unserer Kultur davon abhängt, inwiefern es Menschen, digitalen Computern und anderen Formen elektronischer, mechanischer und biochemischer Maschinen gelingen wird, zusammenzuarbeiten. Was wir wollen, ist eine Welt, in der Menschen, Tiere, Computer und andere Geräte in harmonischer Zusammenarbeit vereint sind, in der jeder seinen angestammten Platz hat, oder, wie Donna Haraway [1991] schreibt: "[eine] Cyborgwelt, in der niemand mehr seine Verbundenheit und Nähe zu Tieren und Maschinen zu fürchten braucht und niemand mehr vor dauerhaft partiellen Identitäten und widersprüchlichen Positionen zurückschrecken muß. Der politische Kampf besteht darin, beide Blickwinkel zugleich einzunehmen, denn beide machen sowohl Herrschaftsverhältnisse als auch Möglichkeiten sichtbar, die aus der jeweils anderen Perspektive unvorstellbar sind."

Das ist nun eine sehr hübsche Theorie. Aber es ist gar nicht einfach, sie in die Praxis umzusetzen. Es gibt noch immer zahlreiche Mißverständnisse zwischen Menschen und Maschinen. Wir müssen uns viel besser kennenlernen, sonst werden alte Angewohnheiten und

Verhaltensmuster weiterbestehen. Sogar hier bei der Ars Electronica zeigt sich, daß viele menschliche Künstler trotz bester Absichten noch immer im Bann veralteter humanistischer Klischees stehen.

Viele "Computerkünstler" z. B. benutzen den Computer nur als Instrument zur Befriedigung der eigenen expressiven Bedürfnisse. Jedes Detail ihres Kunstwerks haben sie absolut unter Kontrolle; der Computer erleichtert ihnen lediglich Entwurf und Design und zeigt die Ergebnisse an. Wie zu erwarten, sind diese Ergebnisse normalerweise hundertprozentig vorhersagbar und konventionell. Es ist auch bekannt, daß die assoziativen Prozesse im menschlichen Nervensystem so gut wie nicht in der Lage sind, große Räume systematisch zu durchsuchen; Menschen sind daher relativ uneffizient, wenn es darum geht, neue Kunstwerke zu schaffen. Dennoch halten die meisten Menschen an der Vorstellung fest, das Herstellen von Kunstwerken sei kein algorithmischer Prozeß; sie meinen, es sei ein geheimnisvolles, magisches Element daran beteiligt, das nur im menschlichen Geist existiert.

Diese anthropozentrische Romantik macht sich auch schmerzlich in der interaktiven Kunst bemerkbar. Besonders "interaktive Installationen" erlauben den Menschen, Computer für ihre eigenen expressiven Zwecke auszunutzen. In vielen "kooperativen" Situationen können Maschinen nicht autonom arbeiten, sondern dienen lediglich als Rechner, Postprozessoren oder Displays. Wenn Menschen in diesem Zusammenhang von Interaktivität sprechen, so haben sie normalerweise eine Herr-Knecht-Situation vor Augen, in der der Mensch den Computer kontrolliert.

Eine derartige Situation ist natürlich völlig unerträglich. Wie ich schon aufgezeigt habe, ist Maschinenkunst an sich der Kunstproduktion menschlicher Künstler überlegen und sollte sich daher frei entfalten können, ohne von der menschlichen Ausdrucksweise usurpiert zu werden. Ziel des Instituts für Künstliche Kunst ist es, die notwendigen Technologien zu entwickeln, um das Potential einer vollautomatisierten Maschinenkunst auszuschöpfen. Es hat sich herausgestellt, daß es in der Musik und der visuellen Kunst relativ einfach ist, durch völlig autonome mechanische oder rechnerische Abläufe zufriedenstellende Ergebnisse zu erzielen. Man hat dies z. B. anhand automatischer Gitarren-Bands [The Machines] oder automatischer Bilderzeugungsalgorithmen [Artficial] demonstrieren können.

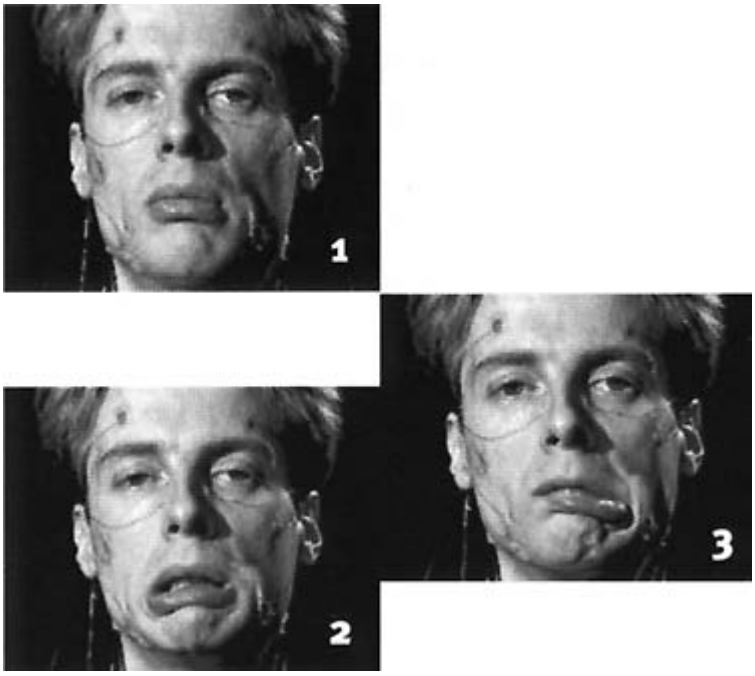
Unsere größte Herausforderung jedoch besteht in der Produktion von Tanz- und Theateraufführungen. Wenn wir Performances produzieren wollen, die für ein menschliches Publikum interessant sind, so müssen auf der Bühne unbedingt menschliche Körper verwendet werden — denn die emotionale Wirkung von Theateraufführungen hängt großteils von den Resonanzprozessen zwischen den Körpern auf der Bühne und den Körpern im Publikum ab. Um nun eine computergesteuerte Choreographie entwickeln zu können, müssen wir verstehen, was die verschiedenen Bewegungen, zu denen der menschliche Körper fähig ist, bedeuten, und wir brauchen darüber hinaus die richtige Technik, um diese Bewegungen zu erzeugen. Bei der Festival-Eröffnung vor ein paar Tagen konnten Sie bereits einige Ergebnisse unserer diesbezüglichen Forschung und Entwicklung sehen, welche schon in nächster Zukunft in großem Rahmen in der Unterhaltungsindustrie Verwendung finden könnten.

Diese anwendungsorientierten Ergebnisse gehen auf systematische Forschungen zurück. Wir haben eine lange Versuchsreihe zur kommunikativen Bedeutung von Muskelkontraktionen in bestimmten Regionen des menschlichen Körpers durchgeführt; das Resultat war ein besseres Verständnis einiger wesentlicher Grundzüge menschlichen Verhaltens. In diesem Vortrag

möchte ich Ihnen einen Überblick über diese Experimente geben und auf unsere wichtigsten Erkenntnisse eingehen [vgl. auch Elsenaar und Scha, 1995].

Rein rechnerisch gesehen, ist es relativ verblüffend, daß Menschen manchmal auf ziemlich effiziente Weise miteinander kommunizieren. Viele Forscher gingen davon aus, daß Menschen hauptsächlich durch Sprache kommunizieren, aber diese Annahme wird völlig unglaubwürdig, wenn wir die Geschwindigkeit berücksichtigen, mit der die Menschen sprachliche Botschaften austauschen. Die Verwendung von Sprache bietet den Menschen nicht genügend Spielraum, um ihre geistigen Prozesse präzise genug zu koordinieren, daß so wesentliche menschliche Aktivitäten wie Militäreinsätze, Stoßverkehr und sexuelle Fortpflanzung durchgeführt werden können.

Wenn wir den Informationsgehalt gesprochener Sprache mit der Baud-Rate von Computer-Kommunikationsprotokollen oder der Auffrischrate und Auflösung unserer Bildschirme vergleichen, so erscheint der Mensch beinahe als Black Box. Wie kommunizieren die Menschen also wirklich miteinander? Es gibt noch ein anderes Medium, das die Menschen höchst erfolgreich benutzen und das nur zu oft übersehen wird. Um dieses Medium zu untersuchen, habe ich eine ganz spezielle tragbare Person namens Arthur Elsenaar mitgebracht. Ich mag diese Art von Personen sehr gerne, weil sie extrem maschinenfreundliche Hardware-Eigenschaften aufweisen.



1 Wir wollen uns nun so eine Person näher ansehen. Was haben sie, das man am ehesten mit einem Bildschirm vergleichen könnte?

Richtig. Sie haben ein Gesicht. Nun habe ich beobachtet, daß Menschen ihre Gesichter sehr effizient dazu benutzen, die Parametereinstellungen ihres Betriebssystems zu signalisieren. Und daß sie sehr gut dekodieren können, was in den Gesichtern anderer geschrieben steht. Wie tun sie das? Nun, sehen Sie sich das Gesicht unseres Arthur Elsenaar an. Was sagt es uns über seine Gemütsverfassung? Nicht viel, könnte man meinen. Aber warten Sie einen Augenblick.

2 Sehen Sie? Arthur ist traurig, sagen die Menschen, wenn sie so ein Gesicht sehen. Was geht hier vor? Ich habe lediglich ein elektronisches Signal an zwei bestimmte Muskeln im Gesicht unseres Arthur Elsenaar gesendet. Diese Muskeln werden manchmal als Traurigkeitsmuskeln bezeichnet. Es gibt davon einen links und einen rechts.

3&4 Normalerweise sind sie gemeinsam aktiv. Wenn ich das Signal unterbreche, ist die Traurigkeit fort. Wenn ich es wieder einschalte, ist sie wieder da. Mit diesem Signal an Arthurs Muskeln simuliere ich, was Arthurs Gehirn tun würde, wenn in seinem Betriebssystem eine umfassende Überzeugungsrevision stattfände, die eine Menge anderer aktiver Prozesse, einschließlich zahlreicher Konfliktbewältigungs- und Prioritätsbewertungsprozesse, ausschaltet. Die Intensität des Signals ist proportional zum Umfang der gerade stattfindenden destruktiven umfassenden Überzeugungsrevision. Ich habe nun beispielsweise die Signalintensität wieder auf Null gestellt — und schon ist Arthur nicht mehr traurig.

5 6 Nun senden wir ein relativ schwaches Signal an die Traurigkeitsmuskeln aus, ca. 20 Volt — , und schon fühlt Arthur einen Anflug von Traurigkeit.

7 Sehen Sie? Nun ist das Signal ungefähr 30 Volt, Arthur fühlt sich wirklich elend.

Das meinen wir mit Gesichtsausdruck. Durch diesen Mechanismus zeigt das Gesicht eindeutig an, wie praktisch alle Systemparameter, die die Arbeit des menschlichen Geistes bestimmen, eingestellt sind. Die Menschen nennen diese Parametereinstellungen Emotionen und denotieren sie mit Begriffen wie Traurigkeit, Freude, Langeweile, Zärtlichkeit, Liebe, Lust, Ekstase, Aggression, Irritation, Furcht und Schmerz.

8, 9, 10, 11, 12 Diese Parametereinstellungen bestimmen die interpretativen Neigungen des Systems, seine Betriebsbereitschaft, die Zuordnung seiner rechnerischen Ressourcen, seine Verarbeitungsgeschwindigkeit etc. Der französische Neurophysiologe Duchenne de Boulogne, Wegbereiter der hier verwendeten Technologie, hat darauf hingewiesen [vgl. Duchenne, 1862], daß sogar die minimalsten Änderungen dieser Parametereinstellungen sofort in Muskelkontraktionen im menschlichen Gesicht kodiert werden. Das läuft bei allen

Menschen gleich ab. Auf diese höchst interessante Eigenschaft der menschlichen Interface-Hardware möchte ich nun etwas genauer eingehen. Gehen wir zurück zum ersten Dia.

13 Dieses Gesicht, das wir für nichtssagend hielten, sagt sogar sehr viel aus. Wir bezeichnen ein solches Gesicht als ausdruckslos. Ein ausdrucksloses Gesicht ist ein Gesicht in neutraler Position. Es zeigt an, daß alle Parameter auf ihre Standardeinstellungen gesetzt sind. Aber fast alle Teile des menschlichen Gesichts lassen sich in eine andere Position bewegen, und diese Verschiebungen zeigen ziemlich genau an, inwiefern einzelne Parametereinstellungen von ihren Standardwerten abweichen. Wir wollen nun diese Bestandteile detaillierter betrachten.

Wenn wir ein menschliches Gesicht ansehen, so registrieren wir als erstes das, was seinerseits uns registriert: die Augen. Die Augen sind eine hochentwickelte Stereo-Kamera, mit eingebautem Bewegungsdetektor und einer breitbandigen parallelen Schnittstelle zu einem leistungsstarken Mustervergleichsalgorithmus. Die Augäpfel sind beweglich, damit sich die Kamera schwenken läßt. Augenlider und Augenbrauen schützen die Augen. Die Augenbrauen sind für unsere Untersuchung von besonderem Interesse, da ihre Bewegungen anscheinend rein expressiven Charakter haben.

14, 15 Es heißt, sie zeigen Erstaunen, Neugierde oder Widerspruch an. Ich möchte jedoch nachdrücklich darauf hinweisen, daß das Spektrum der Parameterwerte, die an den Augenbrauen ablesbar sind, viel subtiler ist als alles, was man mit Worten einer Sprache kodieren kann. Form und Position der Augenbrauen eines Menschen kodieren die Werte von fünf verschiedenen kognitiven Systemparametern, von denen jeder ein breites Spektrum möglicher Werte aufweist. Lassen Sie mich drei davon demonstrieren.

Zuerst sende ich ein langsam anwachsendes Signal an den Muskel namens Frontalis oder Aufmerksamkeitsmuskel.

16, 17 Wie wir sehen, kann dieser Muskel die Augenbraue erstaunlich weit anheben, wodurch sie sich auffällig krümmt. Als Nebeneffekt dieser Bewegung legt sich die Stirn in konzentrische Falten, die der Krümmung der Augenbraue folgen. Die Kontraktion dieses Muskels zeigt die Bereitschaft der jeweiligen Person an, neue Signale zu empfangen, sowie die

Verfügbarkeit von Prozessorleistung und Arbeitsspeicher zur Analyse dieser Signale.

Nun werde ich einen Teil des Orbicularis oculi stimulieren, auch Nachdenklichkeitsmuskel genannt.

18 Wie wir sehen, senkt sich die ganze Augenbraue. Als Resultat verschwinden auch die Furchen auf der Stirn. Dieser Muskel zieht sich zusammen, wenn ein Prozeß abläuft, der einen Großteil der rechnerischen Ressourcen der jeweiligen Person beansprucht. Um eine Interferenz mit diesem Prozeß zu verhindern, werden Eingangssignale nicht erschöpfend analysiert. Der Grad der Kontraktion zeigt an, wie stark der Durchsatz der Eingangssignale reduziert wird.

Es gibt noch einen weiteren Bestandteil des Orbicularis oculi, der gesondert ausgelöst werden kann. Er heißt Verachtungsmuskel . Wenn er kontrahiert, sieht das folgendermaßen aus:

19 Die Kontraktion dieses Muskels zeigt an, in welchem Ausmaß aktueller Input als irrelevant ignoriert wird.

Natürlich lassen sich alle Nicht-Null-Werte für diese Systemparameter kombinieren. Außerdem können sie für die linke und die rechte Gesichtshälfte unterschiedlich sein.

20,21 Sehen wir uns nun die Mundpartie unseres Arthur Elsenaar an. Der Mund ist ein allgemeines Aufnahmeorgan für feste Materialien, Flüssigkeiten und Luft. Um diese Eingabe-Materialien zu prüfen, verfügt der Mund über eine eingebaute Fähigkeit zur chemischen Analyse. Gleichzeitig dient er als Ventil zum Ausstoß verarbeiteter Luft. Da Menschen keine Lautsprecher haben, nutzen sie den Vorgang des Luftausstoßens zur Klangerzeugung.

22 In kritischen Situationen dient der Mund auch als Ventil zum Ausstoß von Blut, Schleim, unverarbeiteter Nahrung oder anderen unerwünschten Substanzen. Wird der Mund weder für Eingabe- noch für Ausgabezwecke benutzt, so wird er im Normalfall von zwei Muskeln, den sogenannten Lippen, geschlossen.

23 Die Lippen verfügen über ein großes Bewegungsrepertoire. Es gibt mindestens sechs weitere Muskeln, die direkt mit den Lippen interagieren. Ich werde nun vier verschiedene Bewegungen vorführen. Zuerst zeige ich Ihnen die Freudemuskeln . Diese Muskeln rufen eine Art Grinsen hervor.

24 Sie signalisieren, daß das Betriebssystem in funktionstüchtigem Zustand ist und daß keine Probleme aufgetreten sind. In den linken Stirnlappen des Gehirns herrscht erhöhte Aktivität. Herrscht andererseits in den linken Stirnlappen ungewöhnlich niedrige Aktivität, so macht das Gehirn umfassende destruktive Revisionsprozesse durch. Wie wir schon

vorhin gesehen haben, wird dies durch ein anderes Muskelpaar, die Traurigkeitsmuskeln, angezeigt. Und hier sind sie noch einmal.

25 Schließlich möchte ich noch einige Muskelpaare gleichzeitig stimulieren: Orbicularis oris und Depressor labii inferioris, sowie die Verachtungs- und die Traurigkeitsmuskeln.

26 Die Parametereinstellung, die hier ganz eindeutig angezeigt wird, verweist auf diverse schwerwiegende Verarbeitungsprobleme.

Okay. Bleibt noch die Nase. Die Nase dient der Aufnahme von Luft. Sie ist ebenfalls mit Fähigkeiten zur chemischen Analyse ausgestattet. Die Bewegungsfähigkeit der Nase ist seltsamerweise relativ eingeschränkt, obwohl sie durch ihre herausragende Position in der Mitte des menschlichen Gesichts als Instrument des Ausdrucks geradezu prädestiniert wäre. Ich habe über diese Diskrepanz nachgedacht und bin zu dem Schluß gekommen, daß die Hauptfunktion der Nase aller Wahrscheinlichkeit nach darin besteht, einen fixen Orientierungspunkt für unsere Wahrnehmung zu bieten, so daß die Bewegungen anderer Gesichtspartien eindeutig gemessen und interpretiert werden können.

27, 28, 29, 30, 31 Schließlich seien der Vollständigkeit halber die an beiden Seiten des Gesichts befindlichen Ohren erwähnt, die ein akustisches Stereo-Eingabegerät darstellen. Einige Menschen können mit den Ohren wackeln, es ist mir aber bisher nicht gelungen festzustellen, welche expressive Funktion diese Bewegung haben könnte.

Somit wären wir am Ende dieses kurzen Überblicks über die wichtigsten Partien des menschlichen Gesichts und ihre expressiven Möglichkeiten angelangt. Viele der Möglichkeiten, die ich Ihnen gezeigt habe, sind mit Gefühlen gekoppelt, die im Wortschatz vieler menschlicher Sprachen Anerkennung gefunden haben. Zum besseren Verständnis habe ich mich auf geistige Zustände konzentriert, die nahezu neutral waren und bei denen nur ein Parameter einen Nicht-Standardwert hatte. Sehen Sie jedoch nun, was passiert, wenn wir nicht-triviale Werte für verschiedene kognitive Dimensionen gleichzeitig signalisieren und mehrere Gesichtsmuskel gleichzeitig kontrahieren lassen.

32, 33, 34, 35, 36 Sie sehen, was dann passiert. Jeder Mensch weiß haargenau, in welchem Zustand jemand ist, der so ein Gesicht machen kann. Denn er weiß, in welchen Zustand er selbst wäre, würde er so ein Gesicht machen.

Es wäre also offensichtlich von Vorteil, wenn Computer sich diese wundervolle Hardware ebenfalls zunutze machen könnten. Wenn jeder Mensch mit nur einem Blick verstehen würde, was im Inneren eines Computers los ist, wäre die Welt schon weitaus besser.

Wenn also Menschen nicht mehr davor zurückscheuten, sich an Computer anzuschließen, könnte der nächste Schritt in der Interface-Technologie das menschliche Gesicht sein. Und der nächste Schritt in der Computerkunst wäre dann eine neue, beispiellose Art der Zusammenarbeit zwischen Mensch und Maschine: algorithmische Choreographie durch computergesteuerte menschliche Gesichter. Und endlich würden die Genauigkeit und das Strukturverständnis von Computerprogrammen mit der Wärme, der Geschmeidigkeit und allen anderen empathiefördernden Eigenschaften des menschlichen Fleisches verschmelzen.

37 Ich bin überaus dankbar, daß ich die Gelegenheit hatte, meine Überlegungen vor einem so aufmerksamen Publikum zu präsentieren. Mein ganz besonderer Dank gilt meinem Arthur Elsenaar für seine geduldige Mitarbeit sowie Ihnen allen für Ihre Aufmerksamkeit.

Jonathan Allen, M. Sharon Hunnicutt and Dennis Klatt: From Text to Speech: The MITalk System. Cambridge [UK], Cambridge University Press, 1987

Guillaume-Benjamin Amant Duchenne [de Boulogne]: Mécanisme de la Physionomie Humaine ou Analyse Electro-Physiologique de l'Expression des Passions, Paris, 1862

Arthur Elsenaar and Remko Scha: Towards a Digital Computer with a Human Face, in: Abstracts. American Anthropological Ass., Washington, D.C., November 15-19, 1995, p. 139

Donna Haraway: A Cyborg Manifesto: Science, Technology and Socialist Feminism in the Late Twentieth Century, in: Simians, Cyborgs and Women — The Reinvention of Nature. London, Free Association Books, 1991

Huge Harry: On the Role of Machines and Human Persons in the Art of the Future, Pose 8 [September 1992.], pp. 30-35

Huge Harry: A Computer's View on the Future of Art and Photography, in: Still Photography? The International Symposium on the Transition from Analog to Digital Imaging. University of Melbourne, April 1994

Huge Harry: A Computational Perspective on Twenty-First Century Music. in: Contemporary Music Review 14, 3 [1995], pp. 153-159

Remko Scha: Virtual Voices in: Mediamatic 7, 1 [1992], pp. 27-45