Sichtbarkeit von Dellen und Beulen auf spiegelnden Oberflächen

Mathias Ziebarth[1], Michael Heizmann[2,3] und Jürgen Beyerer[1,3]

[1] Karlsruher Institut für Technologie, Institut für Anthropomatik und Robotik, Lehrstuhl für Interaktive Echtzeitsysteme, Adenauerring 4, D-76131 Karlsruhe
[2] Hochschule Karlsruhe – Technik und Wirtschaft, Moltkestraße 30, D-76133 Karlsruhe
[3] Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, Fraunhoferstraße 1, D-76131 Karlsruhe

Zusammenfassung Während sich Fehler auf matten Oberflächen anhand ihrer Ausdehnungen bewerten lassen, ist bei spiegelnden Oberflächen die Auswirkung eines Defekts auf die spiegelnde Abbildung maßgebend. In diesem Beitrag wird ein Zusammenhang zwischen der Sichtbarkeit eines Defekts für den Menschen und dessen Tiefe, Ausdehnung und Krümmung sowie der Reflektanz der Oberfläche hergestellt. Dazu wird eine Studie vorgestellt, bei der die Studienteilnehmer Defekte auf speziell präparierten Blechteilen in einer vorgegebenen Zeit erkennen und subjektiv nach ihrer Auffälligkeit bewerten sollten. Für die Schätzung der Detektionswahrscheinlichkeiten wurde das aus der probabilistischen Testtheorie bekannte Rasch-Modell verwendet. In der Studie wurden Bleche mit unterschiedlichen Glanzgraden untersucht und Defekte unterschiedlicher Größe und Ausprägung auf den Blechen zufällig verteilt. Zum Abgleich der Detektionsergebnisse der Testpersonen mit den deflektometrisch erfassten Eigenschaften der Defekte wird die psychometrische Funktion verwendet. Es wird gezeigt, dass sich die subjektiven Bewertungen der Studienteilnehmer mithilfe von Stevens Potenzgesetz ableiten lassen.

1 Einleitung

Im modernen Produktdesign haben spiegelnde Oberflächen eine wichtige Rolle. Neben ihrer Schutzfunktion ist die Beschaffenheit der Ober-

fläche maßgeblich für die Wahrnehmung eines Objekts verantwortlich. Beispiele hierfür sind lackierte Autokarosserien oder emaillierte Haushaltsgeräte. Während der Herstellung können Fehler auftreten, die erkannt und beurteilt werden müssen, um eine gleichbleibende Qualität zu gewährleisten. Studien zur Wahrnehmung von Defekten auf Oberflächen rücken daher immer mehr in den Fokus der klassischen Messtechnik. Da zunehmend automatische Inspektionssysteme die menschlichen Prüfer ablösen, ist es nötig, die bisher subjektiven Fehlermaße durch objektive Maße zu ersetzen. Diese sollten die Einschätzung eines menschlichen Prüfers bestmöglich nachbilden.

Insbesondere Fertigungstoleranzen, die eine reine Höhenabweichung zum Modell definieren, sind auf spiegelnden Oberflächen nicht geeignet. Im Gegensatz zu matten Oberflächen kann man in spiegelnden Oberflächen Teile der Umgebung wahrnehmen. Das virtuelle Bild der Umgebung, das auf der Oberfläche entsteht, hängt dabei von der Form und Reflektanz der Oberfläche ab. Da bereits kleine Änderungen in der Oberflächenkrümmung zu sichtbaren Verzerrungen des virtuellen Bildes der Umgebung führen, sind die Genauigkeitsanforderungen an solche Oberflächen höher als bei matten Oberflächen. Wie hoch diese Anforderungen für die Fehlermaße Tiefe, Ausdehnung und Krümmung eines Defekts sind und inwiefern sie von der Reflektanz der Oberfläche abhängen, wird in dieser Arbeit untersucht.

2 Deflektometrie

Bei der Inspektion spiegelnder Oberflächen treten im Vergleich zur Inspektion matter Oberflächen andere Problemstellungen auf. Zunächst lassen sich Messverfahren, die Muster auf die Oberfläche projizieren und die Oberfläche direkt beobachten, nur schwer anwenden. Hier finden deflektometrische Verfahren [1] ihre Anwendung. Zudem hat die Deflektometrie Vorteile, wenn es das Ziel ist, Fehler zu finden, die auch einem Menschen auffallen würden. Das Prinzip, mit dem die Oberfläche erfasst wird, ähnelt dem, wie auch der Mensch spiegelnde Oberflächen wahrnimmt. Statt die Oberfläche direkt zu beobachten, werden die Verzerrungen von Mustern, die sich in der Oberfläche spiegeln, gemessen. Die Verzerrungen sind durch die Form der Oberfläche vorgegeben und ändern sich dort am stärksten, wo auch die Krümmung der

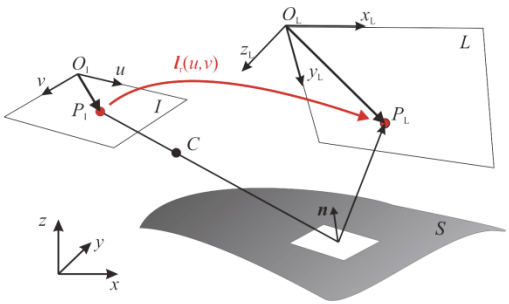


Abbildung 14.1: Prinzip der deflektometrischen Registrierung aus [1].

Oberfläche besonders groß ist. Diese Krümmungsinformationen sind im Normalenfeld der Oberfläche enthalten, welches von der Deflektometrie direkt erfasst wird. Dazu wird ein Messaufbau wie in Abbildung 14.1 dargestellt, bestehend aus einer Kamera mit der Bildebene I, der spiegelnden Oberfläche S als Prüfobjekt und dem Schirm L verwendet. Auf den Schirm werden sinusförmige Streifenmuster in horizontaler und vertikaler Richtung projiziert. Die Kamera ist so positioniert, dass die Muster auf dem Schirm über die spiegelnde Oberfläche beobachtet werden können. Durch die Abbildung über die Oberfläche werden die dargestellten Muster verzerrt. Anhand der Beobachtung einer ganzen Mustersequenz können die Schirmpunkte eindeutig den Kamerasichtstrahlen zugeordnet werden:

$$l : P_I \to P_L,\ l[u,v] = (x_L, y_L).$$

Diese Abbildung wird als deflektometrische Registrierung bezeichnet. Sie enthält bereits wesentliche Informationen über die Oberfläche, aus denen sich das Normalenfeld gewinnen lässt. Ohne Kenntnis der genauen Lage der Oberfläche lassen sich jedoch keine eindeutigen Oberflächennormalen angeben bzw. lässt sich keine eindeutige Rekonstruktion der Oberfläche erstellen [2].

3 Durchführung der Studie

Für die Studie wurden insgesamt 30 Mitarbeiter und Studenten am Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung

befragt. Wie im Überblick über das Teilnehmerfeld in Abbildung 14.2 zu sehen, ist die Testgruppe nicht repräsentativ für die Gesamtbevölkerung, gibt aber bereits Anhaltspunkte für eine Bewertung. Um Fremdeinflüsse weitestgehend auszuschließen, wurde die Studie in einem leeren Büroraum mit weiß gestrichenen Wänden, vollständig abgedunkeltem Fenster und geschlossener Tür durchgeführt. Die Studienteilnehmer saßen während ihrer Aufgabe an einem Tisch, auf dem ein 24"Monitor im Abstand von 20mm bis 110mm schräg im Winkel von 30° zur Tischplatte installiert wurde (siehe Abbildung 14.3). Auf dem Monitor wurde ein sinusförmiges Streifenmuster mit 50 Streifen angezeigt. Auf die Tischplatte unterhalb des Monitors wurde ein weißes Blatt mit aufgezeichnetem 10×10 - Raster gelegt, um es den Teilnehmern zu erleichtern, absolute Positionen auf dem Testblech bestimmen zu können. In die Mitte des Rasters wurden dann die 300mm $\times$ 300mm großen Testbleche gelegt. Die Testbleche wurden aus 1mm starkem Stahlblech gefertigt, schwarz lackiert und dann auf die Glanzgrade (nach DIN 67530) 60GE, 70GE, 80GE und 95GE poliert. Auf allen Testblechen wurden jeweils 8 Fehler (siehe Tabelle 14.1) an zufälligen Positionen (siehe Abbildung 14.4) aufgebracht. Zum Aufbringen der Fehler wurden die Testbleche in eine spezielle Halterung eingespannt und dann lokal mithilfe einer Gewindestange mit einem aufgesetzten Holzdübel verformt. Neben dem Monitor wurde der jeweils aktuelle Fragebogen abgelegt. Dazu wurde neben dem Monitor eine kleine Tischlampe mit 42W-Leuchtmittel angebracht, um in dem ansonsten dunklen Raum lesen zu können.

Defekt	1	2	3	4	5	7	8
Höhe	6mm	5mm	4mm	2mm	1mm	0,5mm	0,5mm

Tabelle 14.1: Höhe der auf den Testblechen aufgebrachten Defekte.

Der Testablauf war wie folgt: Zunächst wurden den Testpersonen allgemeine Testanweisungen und Fragebogen mit allgemeinen Fragen zur Person vorgelegt. Nachdem die allgemeinen Fragen beantwortet waren, bekamen die Testpersonen nacheinander vier Testbleche mit den Glanzgraden 60GE, 70GE, 80GE und 95GE vorgelegt und jeweils einen dazugehörigen Fragebogen mit zwei Aufgaben. Die erste Aufgabe bestand darin, alle Defekte, die auf dem Blech erkannt wordene waren, fortlaufend nummeriert in ein 10×10 - Raster einzutragen. Zum Be-

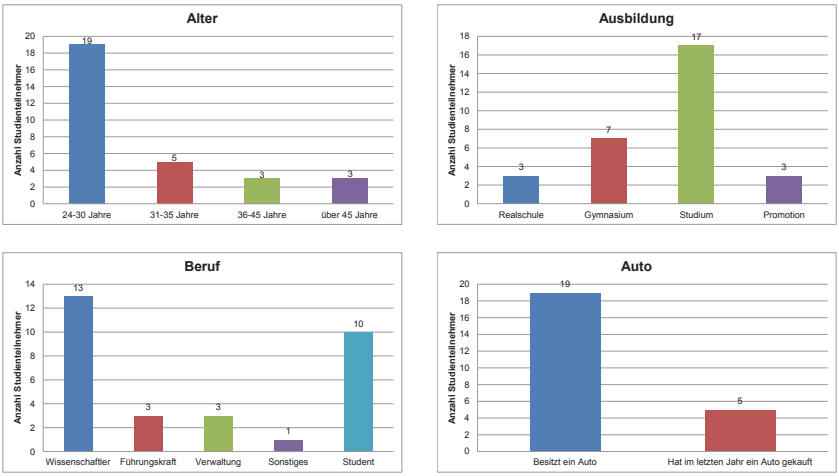


Abbildung 14.2: Überblick über die 30 Studienteilnehmer.

arbeiten dieser Aufgabe hatten die Testpersonen jeweils maximal eine Minute Zeit. Bei der zweiten Aufgabe sollte nun jedem erkannten Defekt eine Bewertung auf einer Skala (1–6, einfach zu finden – schwer zu finden) zugeordnet werden. Für die zweite Aufgabe gab es keine Zeitbeschränkung. Der gesamte Test dauerte etwa 20 Minuten pro Person.

4 Methoden

Um die Detektionswahrscheinlichkeiten der einzelnen Defekte zu bestimmen, wurde das Rasch-Modell [3] verwendet. Das Rasch-Modell ist ein Modell der probabilistischen Testtheorie und unterscheidet zwischen der (im Allgemeinen nicht direkt beobachtbaren) Fähigkeit einer Person, bestimmte Aufgaben zu lösen, und der Schwierigkeit der jeweiligen Aufgaben. Damit ist es möglich, die Lösungswahrscheinlichkeit der Aufgabe für eine durchschnittliche Person zu schätzen, unabhängig davon, wie genau das Teilnehmerfeld zusammengesetzt ist. Zur Schätzung der Rasch-Parameter wurde die bedingte Maximum-Likelihood-Methode verwendet (siehe [4], Kapitel 10 und [5]). Da für

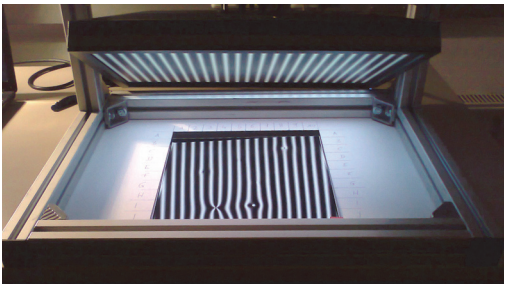
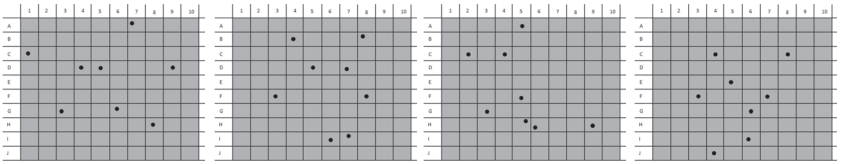


Abbildung 14.3: Testaufbau zur Wahrnehmungsstudie.



(a) 60GE (b) 70GE (c) 80GE (d) 95GE

Abbildung 14.4: Verteilung der Fehler auf den Testblechen.

Aufgaben, die von keinen bzw. allen Testpersonen gelöst werden, die Aufgabenschwierigkeit nur in eine Richtung abgeschätzt werden kann, wurde die Wahrscheinlichkeit für die Lösung einer solchen Aufgabe 0 bzw. 1 gesetzt.

Für die Beschreibung des Zusammenhangs der gemessenen Höhe, Fläche bzw. Krümmung eines Defekts und der Wahrscheinlichkeit, dass ein Studienteilnehmer diesen entdeckt, wurde die psychometrische Funktion [6] verwendet:

$$F(x) = \gamma + (1 - \lambda - \gamma) f(x) \text{ mit } f(x) = 1 - e^{-(\beta x)^{\alpha}}.$$

Dabei ist γ die Wahrscheinlichkeit dafür, eine richtige Lösung zu raten, und λ die Wahrscheinlichkeit dafür, auch bei einer sicheren Lösung der Aufgabe einen Fehler zu machen. Da die besonders einfachen Fehler von allen Testpersonen gefunden wurden, wurde hier $\lambda = 0$ angenommen. Die Wahrscheinlichkeit, einen der 8 Fehler auf dem 10×10 - Raster zu erraten, wird im Mittel mit $4/100$ abgeschätzt, daher wurde γ entsprechend gewählt. Als Wahrscheinlichkeitsverteilung $f(x)$ wurde hier

die in der Literatur häufig verwendete Weibull-Verteilung [6] mit $\alpha = 2$ angenommen, da eine gleichzeitige zuverlässige Schätzung von α und β aufgrund der wenigen Datenpunkte nicht möglich ist.

Um aus den gemessenen Größen Höhe h, Fläche A und Krümmung κ eines Defekts die subjektive Bewertung der Studienteilnehmer abzuleiten, wurde die Stevens'sche Potenzfunktion [7] für Reiz R und Empfindung E verwendet:

$$E = k(R - R_0)^n.$$

Sie verallgemeinert das Weber-Fechner'sche Gesetz, wonach der empfundene Sinneseindruck proportional zum Logarithmus der Stärke des physikalischen Reizes ist, was oftmals nur für kleine Intensitätsbereiche gültig ist. Da die Studienteilnehmer die Bewertungsskala 1–6 unterschiedlich nutzten, mussten die Bewertungen der Defekte zunächst normalisiert werden, um diese zwischen den Studienteilnehmern vergleichbar zu machen. Dafür wurden die Bewertungen auf eine Skala von $0 =$ „sehr schwer zu sehen“ und $1 =$ „sehr einfach zu sehen“ normalisiert. Zudem konnte in den Daten beobachtet werden, dass die Teilnehmer die Skala nicht gleichbleibend über die verschiedenen Bleche hinweg verwendet haben. So haben sie für Fehler gleicher Größe auf den unterschiedlichen Glanzgraden die gleichen Bewertungen vergeben, obwohl die Detektionswahrscheinlichkeiten zeigen, dass die Fehler auf den stärker spiegelnden Oberflächen einfacher zu sehen sind. Ein Vergleich der Glanzgrade aufgrund der Bewertungen ist somit nicht mehr möglich. Zudem wurde angenommen, dass ein Fehler, der von einem Teilnehmer nicht gefunden wurde und in Folge dessen auch nicht bewertet wurde, von dieser Person als maximal schwer bewertet worden wäre. Die Bewertungen von 2 Teilnehmern waren entgegengesetzt zu denen aller anderen Teilnehmer, so dass hier von einem Fehler bei der Beantwortung ausgegangen werden kann. Die entsprechenden Bewertungen wurden aus der Auswertung ausgeschlossen.

Die physikalischen Größen wurden deflektometrisch bestimmt, wobei hier nicht kalibrierte Messungen verwendet wurden. Somit sind die Werte für Fläche und Krümmung lediglich proportional zu den wahren Werten, aber für zwischen den verschiedenen Blechen vergleichbar. Die Fläche der Defekte ist außerdem von dem Schwellwert abhängig, ab dem eine Abweichung von der Ebene als Defekt angesehen wird.

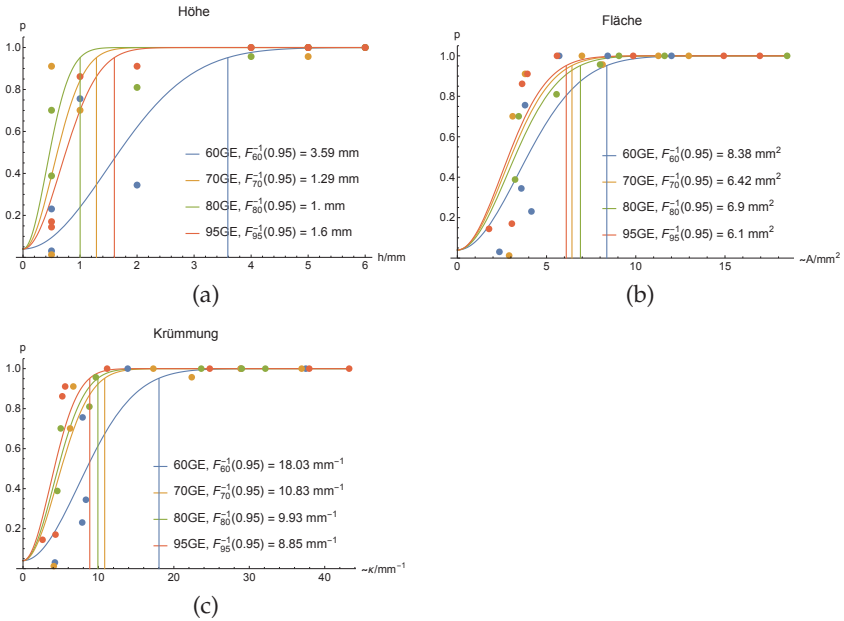


Abbildung 14.5: Detektionswahrscheinlichkeiten in Abhängigkeit von der Höhe, der Fläche und der Krümmung (jeweils unkalibriert).

Die Höhe der Defekte wurde über die Einschraubtiefe beim Auftragen der Defekte bestimmt.

5 Ergebnisse

In der Studie zeigten sich zunächst verschiedene Probleme in den Daten, die die Ergebnisse beeinflussten. Sobald zwei Defekte auf der Oberfläche sehr nah beeinander lagen, fiel es den Testpersonen deutlich schwerer, beide Defekte getrennt voneinander zu erkennen (z.B. Defekte auf H5/H6 in Abbildung 14.4c). Die entsprechenden Ergebnisse mussten aus der Auswertung entfernt werden, um das Ergebnis nicht zu verfälschen. Außerdem setzte eine Art Lernprozess bei der Bearbeitung der Aufgaben ein. Mit jedem Blech wurde die Zeit kürzer, die die

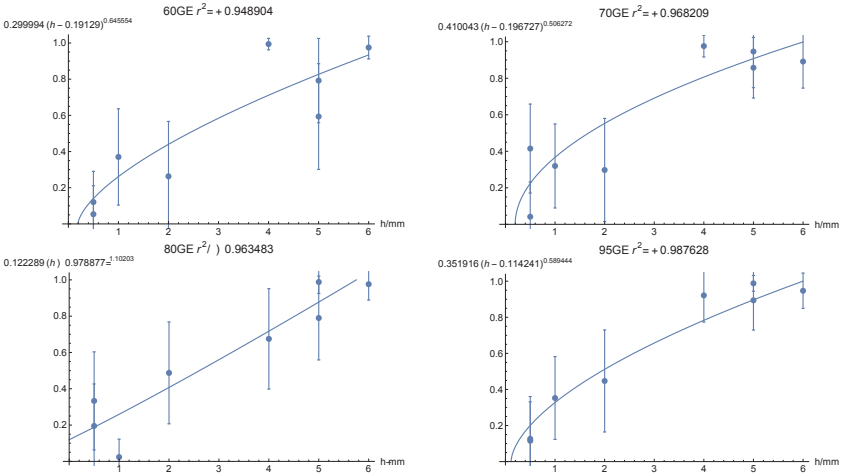


Abbildung 14.6: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Höhe (unkalibriert) für die 4 Testbleche.

Personen für das Suchen und Dokumentieren der Fehler benötigten. Da jedes Mal mit dem Blech 60GE angefangen wurde, kann davon ausgegangen werden, dass die Detektionsleistungen hier etwas zu niedrig eingeschätzt werden.

In den Grafiken in Abbildung 14.5 sind die Detektionswahrscheinlichkeiten der Defekte in Abhängigkeit von Höhe, Fläche und Krümmung der Defekte dargestellt. Zudem wurde die psychometrische Funktion an die Ergebnisse angepasst und der Wert der jeweiligen Größe bestimmt, ab dem die Erkennungswahrscheinlichkeit 95 % beträgt. Es ist zu beobachten, dass die Erkennungswahrscheinlichkeit für das schwach spiegelnde Blech (60GE) in allen drei Fällen stark abnimmt. Für die stärker spiegelnden Bleche zeigt sich kein eindeutiges Bild, da die Unterschiede recht gering sind und sich die Reihenfolge in den Abbildungen a, b und c ändert. Das erwartete Ergebnis, dass Defekte auf stärker spiegelnden Blechen eher gefunden werden, zeigt sich einzig in Abhängigkeit von der Krümmung (Abbildung 14.5c). Die exakten Werte für die Sichtbarkeit sollten mit Vorsicht behandelt werden, da die Modellparameter aufgrund der wenigen Abtastpunkte (8 Fehler pro Blech)

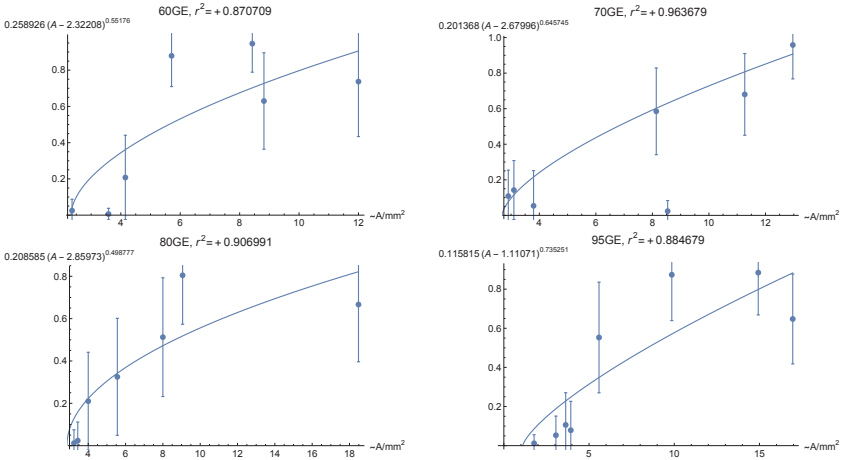


Abbildung 14.7: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Fläche (unkalibriert) für die 4 Testbleche.

statistisch nicht signifikant sind.

Die Abbildungen 14.6, 14.7 und 14.8 zeigen die normalisierten Bewertungen der Defekte durch die Testteilnehmer in Abhängigkeit von den physikalischen Messgrößen und dem Glanzgrad des Blechs. Zudem ist die Standardabweichung für die Bewertung jedes Defekts gegeben. Darüber hinaus ist das beste angepasste Modell (Stevens'sche Potenzfunktion) dargestellt und dessen Bestimmtheitsmaß[4] r^2 gegeben. Die geschätzten Exponenten n des Modells nehmen Werte kleiner als 1 an. Das bedeutet, dass Änderungen in der Defektausprägung gerade im Grenzbereich der Wahrnehmbarkeit der Defekte stärker wahrgenommen werden. Aufgrund der wenigen Datenpunkte und der hohen Varianz in den Bewertungen ist die statistische Signifikanz der Modellparameter allerdings zu gering, um damit verlässliche Vorhersagen zu treffen. Es zeigt sich jedoch, dass sich die Bewertungen mithilfe von Stevens Potenzfunktion aus den messbaren Größen Höhe, Ausdehnung und Krümmung ableiten lassen. Die Bestimmtheitsmaße für die Modelle, die den Zusammenhang von Höhe bzw. Krümmung eines Fehlers

[4] Kein Zusammenhang von Modell und Daten besteht für $r^2 = 0$. Ein perfekter Zusammenhang von Modell und Daten liegt für $r^2 = 1$ vor.

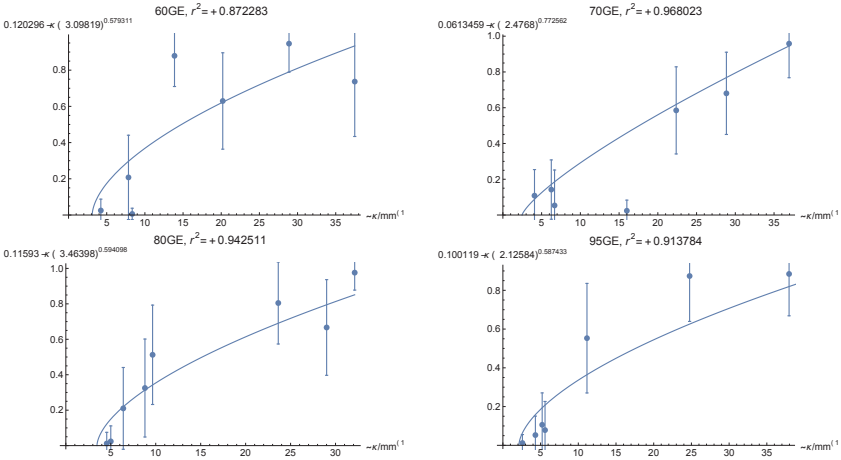


Abbildung 14.8: Normalisierte Bewertung und Standardabweichung in Abhängigkeit von der Krümmung (unkalibriert) für die 4 Testbleche.

und dessen Bewertung darstellen, sind besser als die Bestimmheitsmaße für die Modelle, die den Zusammenhang von Fläche und Bewertung darstellen.

6 Zusammenfassung

Unter Verwendung mehrerer Standardverfahren aus der Psychometrie und den empirischen Sozialstudien wurde der Zusammenhang deflektometrisch messbarer Größen spiegelnder Oberflächen und der Wahrnehmung des Menschen hergestellt. Dabei stellten sich die Höhe und die Krümmung als geeignete Maße für die Detektionswahrscheinlichkeit und die subjektive Bewertung von Fehlern auf spiegelnden Oberflächen heraus. Die Vermutung, dass die Krümmung besonders gut zur Vorhersage der Sichtbarkeit geeignet ist, konnte nicht eindeutig bestätigt werden. Für die Vorhersage von Detektionswahrscheinlichkeiten und Bewertungen allein aus Messgrößen der Deflektometrie und dem Glanzgrad der Oberfläche müssen jedoch deutlich mehr Defekte untersucht werden.

In der Studie hat sich eine Reihe von Verbesserungsmöglichkeiten gezeigt: Wird die Studie wie in diesem Artikel beschrieben durchgeführt, sollte die Reihenfolge der Bleche für jeden Studienteilnehmer anders sein, um den systematischen Einfluss von Lerneffekten auszuschließen. Außerdem sollte jeder Teilnehmer zunächst die Möglichkeit haben, auf einem Trainingsblech die Defektbewertung zu üben, bevor das erste Blech bewertet wird. Um auszuschließen, dass sich die Fehler gegenseitig beeinflussen, muss ein Mindestabstand zwischen den einzelnen Fehlern sichergestellt sein. Für die Auswertung der subjektiven Bewertungen wäre eine gleichbleibend, objektive Referenzskala hilfreich, z.B. in Form eines Referenzbleches. Außerdem kann man die genauen Wahrnehmungsschwellen schneller bestimmen, indem man einen adaptiven Test, d.h. einen Test mit Defekten, die nah an der Wahrnehmungsschwelle der jeweiligen Person liegen, durchführt.

7 Danksagung

Wir danken Benjamin Krüger, der diese Studie im Rahmen seiner Bachelorarbeit durchgeführt hat. Diese Veröffentlichung ist im Rahmen des Projekts MID-Wave entstanden. Dieses Projekt wird finanziert durch die Baden-Württemberg Stiftung gGmbH.

Literatur

1. S. Werling, „Deflektometrie zur automatischen Sichtprüfung und Rekonstruktion spiegelnder Oberflächen“, Dissertation, Karlsruher Institut für Technologie, 2011.
2. J. Balzer, „Regularisierung des Deflektometrieproblems – Grundlagen und Anwendung“, Dissertation, Universität Karlsruhe (TH), Universitätsverlag Karlsruhe, 2008.
3. G. Rasch, „On general laws and the meaning of measurement in psychology“, in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 4, 1961, S. 321–333.
4. H. Moosbrugger und A. Kelava, *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg, 2012.
5. P. Mair, R. Hatzinger und M. Maier, „Extended rasch modeling: The r package erm“, Wirtschaftsuniversität Wien, Tech. Rep., 2009.

6. F. A. Wichmann und N. J. Hill, „The psychometric function: I. fitting, sampling, and goodness of fit“, *Perception & psychophysics*, Vol. 63, S. 1293–1313, 2001.

7. S. S. Stevens, „On the psychophysical law.“ *Psychological review*, Vol. 64, S. 153–181, 1957.