

# Verständlichkeitsmessungen und Qualitätsbeurteilungen von Vokalen mit geglätteter Formantstruktur

STEFAN WARTINI

Fachgebiet Akustische Kommunikation, Technische Universität München

Im Rahmen von Untersuchungen zum Einfluß von Akzentuierung [3] auf die Sprachwahrnehmung wurden Verständlichkeitsmessungen und Qualitätsbeurteilungen synthetischer Vokale durchgeführt, deren Formantstruktur auf den Verlauf einer Zielfunktion hin geglättet wurde. Die vorliegende Arbeit beschreibt den Einfluß unterschiedlicher Zielfunktionen auf die Wahrnehmung derart modifizierter Vokale.

## 1. Versuchsschalle

Als Versuchsschalle dienten die durch Fouriersynthese erzeugten Vokale /a/, /e/, /i/, /o/ und /u/ der Grundfrequenz 100 Hz und der Bandbreite 5 kHz. Der Verlauf ihrer Spektralhüllkurve wurde nach einer auf dem Vokaltraktmodell von Fant basierenden Formel [2] bestimmt. Zur Glättung der Spektralhüllkurven wurde ein Verfahren gewählt, bei dem die Pegel der Teiltöne um einen bestimmten Prozentsatz  $g$  angehoben bzw. abgesenkt werden, je nachdem, ob sie größer oder kleiner als der Pegel  $L_{ziel}(f)$  einer gewählten Zielfunktion bei der Teiltonfrequenz sind. Auf diese Weise paßt sich die ursprüngliche Hüllkurve des Vokals in zunehmendem Maß an die Zielfunktion an. Die Berechnung der Teiltonpegel des geglätteten Vokals erfolgt gemäß der Gleichung:  $L_{glätt}(f) = [L_{vok}(f) - L_{ziel}(f)] \cdot (1 - g) + L_{ziel}(f)$  mit  $0 \leq g \leq 1$ . Neben einer gemeinsamen Zielfunktion für alle Vokale wurde auch eine für jeden einzelnen Vokal spezifische Zielfunktion untersucht.

Als vokalunabhängige Zielfunktion wurde eine Funktion gewählt, die mit 20 dB/Dekade über der Frequenz abfällt, um die Spektralhüllkurve des Langzeitspektrums von stimmhafter Sprache anzunähern. Die Spektralhüllkurve eines vollständig geglätteten Vokals mit dem Pegel 60 dB besitzt somit den Verlauf:  $L_{ziel}(f) = 57,89 \text{ dB} - 20 \log(f/100 \text{ Hz}) \text{ dB}$ .

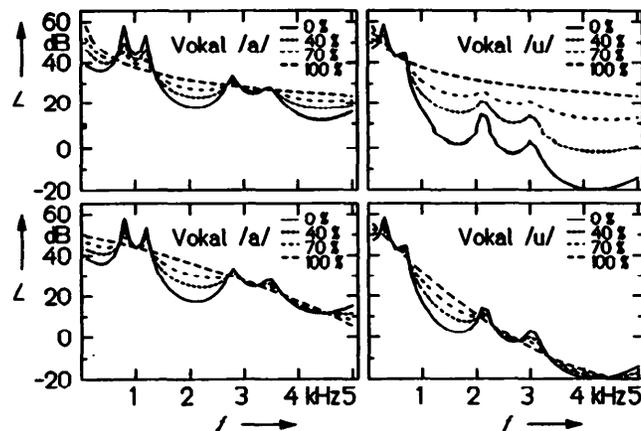


Fig. 1. Verlauf der Spektralhüllkurven der Vokale /a/ und /u/. Parameter der Kurven ist der Glättungsgrad  $g$ . Oben: Vokalunabhängige Zielfunktion. Unten: Vokalspezifische Zielfunktion. Deutlich ist der Einfluß der Zielfunktion auf den Vokal /u/ zu erkennen.

Fig. 1 (oben) zeigt schematisch den Verlauf der Spektralhüllkurven der Vokale /a/ und /u/ bei unterschiedlichen Glättungsgraden. Mit zunehmender Glättung werden im Vokal /a/ sowohl Teiltöne im Bereich der Formanten abgesenkt als auch Teiltöne zwischen den Formanten angehoben. Dagegen werden im Vokal /u/ nur Teiltöne im ersten Formanten abgesenkt, während alle anderen Teiltöne angehoben werden.

Da die vokalunabhängige Zielfunktion die Vokale sehr unterschiedlich verändert, wurden mittels nichtlinearer Regression vokalspezifische Zielfunktionen bestimmt, die zwischen den Maxima und Minima der spektralen Vokalhüllkurven verlaufen. Als Stützwerte für die Regression dienten die Formanten und die Minima zwischen den Formanten. Mit diesen Stützwerten konnte für eine Potenzfunktion zweiten Grades für jeden Vokal ein passender Parametersatz gefunden werden. Die vokalspezifische Zielfunktion verläuft gemäß:

Vokal	a	b	c
/a/	49,90	-0,5258	-0,0070743
/e/	46,61	-0,0626	-0,0082513
/i/	52,89	-1,3007	-0,0131423
/o/	59,79	-3,4162	-0,0409242
/u/	57,58	-2,8256	-0,0248141

Tab. 1. Parameter der Potenzfunktion zweiten Grades zur Festlegung der vokalspezifischen Zielfunktionen der Vokale /a/, /e/, /i/, /o/ und /u/.

$L_{\text{Ziel}}(f) = (a + b f/100 \text{ Hz} + c (f/100 \text{ Hz})^2)$  dB. Tab. 1 zeigt die Parameter a, b und c zur Glättung der einzelnen Vokale. Der Verlauf der Spektralhüllkurven der unterschiedlich geglätteten Vokale /a/ und /u/ ist in Fig. 1 (unten) dargestellt. Deutlich ist bei beiden Vokalen die gleichmäßige Glättung durch die vokalspezifischen Zielfunktionen zu erkennen.

## 2. Versuchsablauf

Die Vokale wurden mit einem Pegel von 60 dB über einen freifeldentzerrten Kopfhörer (Beyer DT48) in einer schallisolierten Meßkabine dargeboten. Die Dauer der Testvokale betrug 200 ms.

Zur Bestimmung der Erkennungsraten wurden die Testvokale monotisch dargeboten. Die Versuchspersonen hatten die Aufgabe, über gekennzeichnete Tasten eines Terminals in der Meßkabine den erkannten Vokal anzugeben. Um zu verhindern, daß sie eine Zuordnung zwischen Schalldarbietung und richtiger Antwort erlernen, erhielten die Versuchspersonen keine Rückmeldung über die Richtigkeit ihrer Antwort. Jeder Testvokal wurde verteilt über mehrere Versuchssitzungen fünfundzwanzigmal dargeboten.

Die Qualitätsbeurteilungen wurden nach der Methode der Größenschätzung mit Ankerschall durchgeführt. Als Anker diente jeweils der ungeglättete Testvokal (Dauer 200 ms), dessen Qualität der Referenzwert 100 zugewiesen wurde. Der Anker wurde über einen Kanal des Kopfhörers dargeboten. Nach einer Pause von 700 ms wurde der geglättete Testvokal (200 ms) kontralateral dargeboten. Die Versuchsperson sollte zur Beurteilung des Testvokals den Wert angeben, der bezüglich des Referenzwertes das Verhältnis ihrer Qualitätsempfindungen von Anker und Testvokal beschreibt. Vor jeder Darbietung wurde der Versuchsperson auf einem Monitor mitgeteilt, welcher Vokal ihr dargeboten wird. Auf diese Weise wurde verhindert, daß die Versuchsperson ihre Beurteilung ausschließlich an der Erkennung des Vokals orientiert oder einen anderen Vokal erkennt und diesen bewertet. Jede Beurteilung wurde verteilt über mehrere Versuchssitzungen sechsmal durchgeführt.

An den Hörversuchen mit vokalunabhängiger Zielfunktion nahmen vier, an denen mit vokalspezifischer Zielfunktion zwei Versuchspersonen im Alter zwischen 24 und 31 Jahren teil.

## 3. Erkennung geglätteter Vokale

Die Auswahl der verwendeten Glättungsgrade erfolgte aufgrund informeller Hörversuche und sollte einen möglichst breiten Bereich zwischen *geringer* und *großer* Beeinflussung der Erkennung abdecken.

### Vokalunabhängige Zielfunktion

Fig. 2 (links) zeigt die Erkennungsraten *h* der Vokale als Funktion des Glättungsgrades *g* gemittelt über vier Versuchspersonen. Die ungeglätteten Vokale (Glättungsgrad *g* = 0%) werden immer erkannt. Mit zunehmender Glättung nimmt die Erkennungsrate ab. Der Glättungsgrad 100% wurde bei diesen Messungen nicht verwendet, da alle Vokale bei dieser Glättung dieselbe Spektralhüllkurve besitzen. Der Vokal /a/ wird selbst beim Glättungsgrad 95% noch in über 70% aller Darbietungen erkannt, während die übrigen Vokale diesen Wert bereits bei Glättungsgraden zwischen 60% und 85% unterschreiten. Wie die Verwechslungsmatrix (Tab.2, links) für

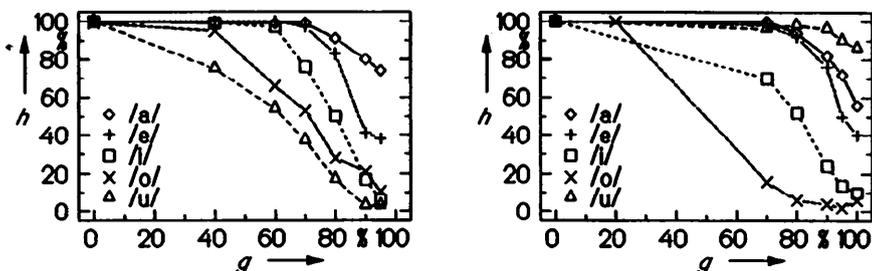


Fig. 2. Relative Häufigkeit  $h$  der Erkennung der Testvokale in Abhängigkeit des Glättungsgrades  $g$ . Mit zunehmendem Glättungsgrad nehmen die Erkennungsraten der Vokale unterschiedlich ab. Links: Vokalunabhängige Zielfunktion: Daten gemittelt über vier Versuchspersonen, 25 Darbietungen je Testvokal und Glättungsgrad. Rechts: Vokalspezifische Zielfunktion: Erkennungsdaten gemittelt über zwei Versuchspersonen.

den Glättungsgrad 95% zeigt, entstehen die hohen Erkennungsraten der Vokale /a/ und /e/ durch häufige Verwechslungen mit den anderen Vokalen. Die Versuchspersonen zeigen eine deutliche Präferenz für diese beiden Vokale. Anscheinend gleicht ein Harmonischer Komplexer Ton, dessen Spektralhüllkurve mit 20 dB / Dekade abfällt, den Vokalen /a/ oder /e/ sehr stark.

Vokalunabhängige Zielfunktion						Vokalspezifische Zielfunktion					
Antworten /%						Antworten /%					
Testvokal	a	e	i	o	u	Testvokal	a	e	i	o	u
a (95%)	74,0	20,0	2,0	3,0	1,0	a(100%)	56,0	26,0	2,0	12,0	4,0
e (95%)	48,0	38,0	6,0	6,0	2,0	e(100%)	60,0	40,0	0,0	0,0	0,0
i (95%)	55,0	31,0	6,0	8,0	0,0	i(100%)	8,0	82,0	10,0	0,0	0,0
o (95%)	40,0	38,0	8,0	11,0	3,0	o(100%)	0,0	0,0	0,0	6,0	94,0
u (95%)	51,0	32,0	7,0	7,0	3,0	u(100%)	2,0	0,0	0,0	12,0	86,0

Tab. 2. Verwechslungsmatrizen der geglätteten Vokale gemittelt über alle Versuchspersonen. Links: Vokalunabhängige Zielfunktion: Glättungsgrad  $g = 95\%$ . Die Einträge zeigen die häufige Zuordnung der stark geglätteten Vokale zu den Vokalen /a/ oder /e/. Rechts: Vokalspezifische Zielfunktion: Glättungsgrad  $g = 100\%$ . Die Einträge zeigen eine häufige Verwechslung der Vokale /o/ und /u/.

### Vokalspezifische Zielfunktion

Fig. 2 (rechts) zeigt die Erkennungsdaten  $h$  der mit vokalspezifischer Zielfunktion geglätteten Vokale als Funktion des Glättungsgrades  $g$ . Auch bei der vokalspezifischen Glättung sinkt die Verständlichkeit mit zunehmendem Glättungsgrad. Aufgrund des ähnlichen Verlaufs der Zielfunktionen mit dem der vokalunabhängigen Zielfunktion gleichen die Erkennungsdaten der Vokale /a/, /e/ und /i/ weitgehend denen aus Fig. 2 (links). Auffällig verlaufen die Erkennungsdaten der Vokale /o/ und /u/. Bei den Untersuchungen mit vokalunabhängiger Zielfunktion lagen die Erkennungsdaten des Vokals /o/ immer über denen des /u/. Mit den vokalspezifischen Zielfunktionen erreicht der Vokal /u/ bei praktisch allen Glättungsgraden die höchste Erkennungsrate. Selbst bei vollständiger Glättung (100%) sinkt diese Rate nicht unter 85%. Die Versuchspersonen assoziieren mit den steil abfallenden Spektralhüllkurven der geglätteten Vokale /o/ und /u/ anscheinend immer den Vokal /u/. Diesen Umstand zeigt auch die in Tab. 2 (rechts) dargestellte Verwechslungsmatrix.

### 4. Qualitätsbeurteilung geglätteter Vokale

Die Beurteilung der Qualität eines Vokals stellt die Versuchsperson vor eine schwierige Aufgabe, da sie ein Urteil fällen soll, das auf einer nicht definierten Empfindung von Tonhöhen und Klangfarben basiert. Die Untersuchungen zur Qualitätsbeurteilung geglätteter Vokale ergaben zum Teil erhebliche individuelle, vor allem aber interindividuelle Schwankungen. Aus Gründen der Übersichtlichkeit wird im folgenden auf die Darstellung der Interquartilbereiche verzichtet.

### Vokalonabhängige Zielfunktion

Fig. 3 (links) zeigt den Median der Qualitätsbeurteilungen  $Q_v$  der Testvokale als Funktion des Glättungsgrades  $g$ . Die Qualitätsbeurteilungen aller Vokale nehmen wie die Erkennungsraten

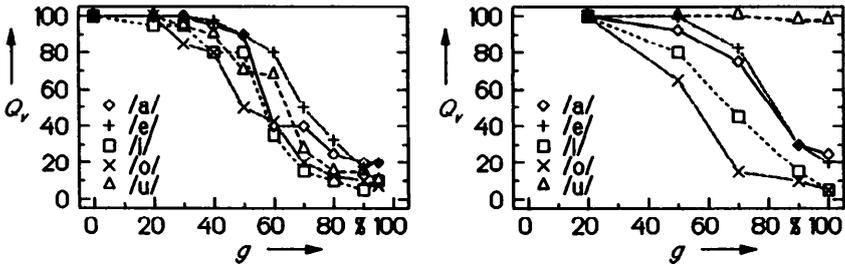


Fig. 3. Qualitätsbeurteilung  $Q_v$  der Testvokale als Funktion des Glättungsgrades  $g$  ermittelt durch Größenschätzung mit Ankerschall (Anker: 100). Links: Vokalonabhängige Zielfunktion. Median von vier Versuchspersonen bei 6 Darbietungen je Testvokal und Glättungsgrad. Rechts: Vokalspezifische Zielfunktion. Median von zwei Versuchspersonen.

mit zunehmender Glättung ab. Für Glättungsgrade über 90% sinken sie unter den Wert 20. Die größten Schwankungen treten im Übergangsbereich hoher und niedriger Vokalqualität bei Glättungsgraden zwischen 50% und 80% auf. Insgesamt verlaufen die Bewertungen der Vokale vergleichsweise ähnlich. Im Gegensatz zu den Verständlichkeitsmessungen, bei denen der Vokal /a/ auch bei hoher Glättung noch gut erkannt wurde, sinken seine Qualitätsbeurteilungen mit zunehmender Glättung ebenfalls ab. Die Präferenz der Versuchspersonen, die bei den Verständlichkeitsmessungen bei großer Glättung einem Klang einen Vokal zuordnen müssen, wird durch die veränderte Meßmethode eingeschränkt.

### Vokalspezifische Zielfunktion

Fig. 3 (rechts) zeigt den Median der Qualitätsbeurteilungen  $Q_v$  von zwei Versuchspersonen. Mit Ausnahme des Vokals /u/ nehmen die Beurteilungen mit zunehmendem Glättungsgrad ab. Die Qualität des Vokals /u/ wird selbst beim Glättungsgrad  $g = 100\%$  noch mit Werten um 100 beurteilt. In einigen Fällen bewerteten die Versuchspersonen den vollständig geglätteten Vokal sogar erheblich besser als den unbearbeiteten Vokal. Im Vergleich zu den Ergebnissen mit vokalonabhängiger Zielfunktion (Fig. 3 links) zeigen die Kurven eine stärkere Auffächerung. Die Qualitätsbeurteilungen beginnen erst bei höheren Glättungsgraden abzusinken.

### 5. Schlußbemerkung

Die beiden Zielfunktionstypen zeigen einen deutlichen Einfluß auf die Erkennungsraten und die Qualitätsbeurteilungen. Anscheinend rufen die geglätteten Vokale durch die unterschiedlichen Steigungen ihrer Spektralhüllkurven unterschiedliche Klangfarben hervor, welche die Versuchspersonen bei ihren Beurteilungen mit auswerten. Besonders deutlich wird dies beim Vokal /u/, der beim Übergang von der vokalonabhängigen auf die vokalspezifische Zielfunktion deutlich besser beurteilt wird. Die stark abfallende Spektralhüllkurve des ungegätteten Vokals /u/ wird weitgehend von den ersten beiden dicht benachbarten Formanten dominiert. Die vokalspezifische Zielfunktion des Vokals /u/ bildet diesen „Formanten“ nach, was zu den guten Beurteilungen und Erkennungsraten führt.

Die Untersuchungen wurden im SFB 204 „Gehör“ mit Unterstützung der DFG durchgeführt.

- [1] Fant, C. G. M. (1973). On the predictability of formant levels and spectrum envelopes from formant frequencies. In: Flanagan, J.L., Rabiner, L. R., Editoren, Speech Synthesis, Seiten 216-228. Dowden, Hutchinson & Ross, Inc., Stroudsburg, Pennsylvania.
- [2] Terhardt, E. (1979). Calculating virtual pitch. Hearing Research, 1: 155-182.
- [3] Wartini, S. (1994). Akzentuierung von Teiltönen in sprachähnlichen Schallen. In: Fortschritte der Akustik - DAGA '94, Dresden S. 1021-1024.