

Extraktion gehörrelevanter Schallsignalparameter aus Flügelklängen

M. N. Valenzuela

Fachgebiet Akustische Kommunikation, Technische Universität München

1. Einleitung

Auf der Grundlage einer gehörangepaßten Schallsignalrepräsentation, dem sogenannten Teiltonzeitmuster (TTZM) [1], wurde ein Verfahren entwickelt, das mit Hilfe psychoakustischer Erkenntnisse gehörrelevante Spektralkomponenten aus Konzertflügelklängen extrahiert. Durch die dabei verwendete Spektralanalyse werden die Verdeckungseffekte nur unzulänglich berücksichtigt. Daher ist es erforderlich, verdeckte Spektralkomponenten, die im TTZM noch enthalten sind, durch Modellierung der Maskierung zu entfernen. Die Reduktion der TTZM durch Entfernen gehörrelevanter Teiltöne wird im folgenden kurz *Reduktion* genannt. Mit einem Hörversuch, in dem die Resynthesen reduzierter TTZM von Flügelklängen mit den Originalklängen verglichen wurden, konnte gezeigt werden, daß mit dem Extraktionsverfahren in der Tat nur gehörirrelevante Teiltöne entfernt werden.

2. Das Extraktionsverfahren

Die einzelnen Spektralkomponenten eines Flügelklangs müssen gleichzeitig als Testton und Maskierer betrachtet werden. Jede Komponente wird dabei von einem Gesamtpegel verdeckt, der sich aus den Erregungen der übrigen Teiltöne und der Intensität, die der absoluten Hörschwelle des Gehörs entspricht, bei der Frequenz der betrachteten Komponente zusammensetzt [4].

Den Verlauf des Erregungspegels eines maskierenden Sinustons erhält man nach Zwicker [6, S. 115], indem man die Mithörschwelle um den Betrag des Verdeckungsmaßes a_v nach oben verschiebt. Verwendet man als Näherung für den Verlauf der Mithörschwellen maskierender Sinustöne die von Terhardt [4] angegebene Dreiecksfunktion, so wird insbesondere für Spektralkomponenten unterhalb ca. 700 Hz im Bereich direkt oberhalb der Komponente eine zu weitgehende Verdeckung vorgenommen. Die Reduktion eines TTZM mit dieser Näherung entfernt auch gehörrelevante Teiltöne. Dies führt zu wahrnehmbaren Unterschieden zwischen dem originalen Flügelklang und dem resynthetisierten, mit der Dreiecksnäherung reduzierten TTZM. Um das Entfernen gehörrelevanter Spektralkomponenten zu verhindern, wurde der Verlauf der oberen Dreiecksflanke wie folgt optimiert.

Näherung für die Erregungsverteilung einer Sinustonkomponente. In Anlehnung an die Ergebnisse verschiedener Autoren für die Mithörschwellen maskierender Sinustöne [2, 7] und unter Beachtung der von Sonntag [3] hervorgehobenen Abhängigkeit der Mithörschwellen von der Tonheit des Maskierers z_M , wird folgende Näherung für die Mithörschwellen verdeckender Sinustöne (f_M, L_M) vorgeschlagen.

Die Steigung S_1 der unteren Flanke der Mithörschwelle wird von Terhardt [4] übernommen:

$$S_1 = 27 \text{ dB/Bark} \quad \text{für} \quad z \leq z_M \quad (1)$$

Der Bereich der oberen Flanke wird durch zwei Geradenstücke angenähert, deren Steigungen von der Frequenz und dem Pegel des Maskierers abhängen. Damit wird die Formvarianz der Mithörschwellen-Tonheitsmuster berücksichtigt. Für die Steigungen S_2 und S_3 ergibt sich:

$$S_2 = \left[-18,9 - \frac{6 \text{ kHz}}{f_M} + 0,19 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}} \quad ; \quad z_M < z \leq z_M + 1 \text{ Bark} \quad (2)$$

$$S_3 = \left[-14 - 0,33 \cdot \frac{f_M}{\text{kHz}} + 0,13 \cdot \frac{L_M}{\text{dB}} \right] \frac{\text{dB}}{\text{Bark}} \quad ; \quad z > z_M + 1 \text{ Bark} \quad (3)$$

Mit dieser Näherung ergibt sich der bei einer beliebigen Tonheit z durch einen Sinuston (f_M, L_M) erzeugte Erregungspegel $L_{EM}(z)$ zu

$$L_{EM}(z) = \begin{cases} L_M - a_0(z_M) - S_1 \cdot (z_M - z) & ; z \leq z_M \\ L_M - a_0(z_M) - S_2 \cdot (z_M - z) & ; z_M < z \leq z_M + 1 \text{ Bark} \\ L_{EM}(z_M + 1 \text{ Bark}) - S_3 \cdot (z_M - z + 1 \text{ Bark}) & ; z_M + 1 \text{ Bark} < z \end{cases} \quad (4)$$

wobei a_0 den Frequenzgang des Außen- und Mittelohres bezeichnet. Man erhält ihn aus dem Verlauf der Absoluthörschwelle bei Frequenzen über 1 kHz [6, S. 114]. In Abb. 1 sind die auf diese Weise berechneten Erregungspegelverläufe von zwei Sinustönen unterschiedlicher Frequenz für verschiedene Pegel als Funktion der Tonheit dargestellt. Der Frequenzgang a_0 wurde in dieser Darstellung vernachlässigt.

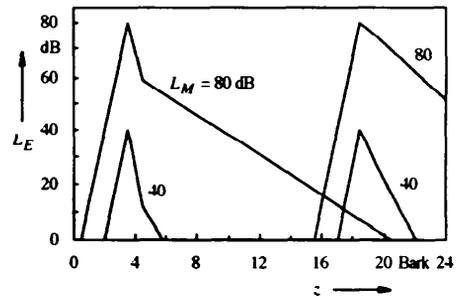


Abb. 1: Berechnete Erregungspegelverläufe von zwei Sinustönen unterschiedlicher Frequenz ($f_1 = 350 \text{ Hz}$, $f_2 = 4,8 \text{ kHz}$) als Funktion der Tonheit. Parameter ist der Pegel L_M des jeweiligen Sinustons; der Frequenzgang a_0 des Ohres wurde vernachlässigt.

Das Extraktionsverfahren berechnet anhand dieser Näherung bei jeder Spektralkomponente die Erregungspegel der anderen Komponenten. Neben ihrer Überlagerung wird auch noch die absolute Hörschwelle berücksichtigt. Im nächsten Schritt überprüft das Verfahren mit Hilfe der nachfolgend formulierten Bedingung, ob eine Spektralkomponente verdeckt wird und somit als gehörrelevant entfernt werden kann.

Maskierungsbedingung. Nach Zwicker [6, S. 105] ist ein Testton eben wahrnehmbar, wenn sein Pegel um das Verdeckungsmaß a_v unterhalb des Störpegels in der Frequenzgruppe liegt. Drückt man diese Maskierungsbedingung mit Hilfe von Erregungen bzw. Erregungspegeln aus, so gilt eine Spektralkomponente (z_K, L_K) als maskiert, falls ihr Erregungspegel $L_{EK} = L_K - a_0(z_K)$ bei der Tonheit z_K um mehr als das Verdeckungsmaß $a_v(z_K)$ unterhalb dem Gesamterregungspegel $L_{EMask}(z_K)$ liegt, der durch andere Teiltöne und der absoluten Hörschwelle bei der Tonheit z_K hervorgerufen wird:

$$L_K - a_0(z_K) - L_{EMask}(z_K) < a_v(z_K) \quad (5)$$

Auf Grund der Zeitvarianz von Flügelklängen erfolgt die Bestimmung maskierter Teiltöne in äquidistanten Zeitabständen, die entsprechend dem Auswertintervall nach [1] festgelegt sind. Dabei können mehrere das Abklingverhalten eines Partialtons repräsentierenden Teiltöne die Maskierungsbedingung (5) erfüllen. Durch das Entfernen dieser Teiltöne aus dem TTZM wird die Zeitstruktur

des entsprechenden Partialtons unterbrochen. Auf die gleiche Weise werden die Zeitverläufe verschiedener Partialtöne unterbrochen. Diese Unterbrechungen können wiederum als eine Aufspaltung des Flügelklangs in mehrere Einzelobjekte verstanden werden. Beim resynthetisierten Signal eines maskierten TTZM, das solche zeitlichen Unterbrechungen einzelner Partialtöne enthält, sind neben dem Flügelklang störende Einzelobjekte wahrnehmbar. Deshalb wurde eine Extraktionsbedingung für Partialtöne formuliert, die im letzten Schritt des Verfahrens überprüft wird.

Extraktionsbedingung für Partialtöne. Anhand der nachfolgend angegebenen Bedingung wird entschieden, ob ein Partialton als vollständig verdeckt betrachtet werden kann. Ist das der Fall, werden alle ihn repräsentierenden Teiltöne als gehörirrelevant entfernt. Ansonsten müssen alle zugehörigen Teiltöne, inklusive derer, die entsprechend (5) maskiert sind, erhalten bleiben. Die Extraktionsbedingung lautet:

Wird ein Partialton durch wenigstens x nicht maskierte Teiltöne (TT) repräsentiert, für die $L_{TT} - a_0(z_{TT}) - L_{EMask}(z_{TT}) \geq c$ erfüllt ist, so gilt der Partialton als nicht vollständig verdeckt und muß extrahiert werden.

$L_{EMask}(z_{TT})$ bezeichnet hierbei den Gesamterregungspegel, der durch andere Komponenten und der absoluten Hörschwelle bei der Tonheit z_{TT} des betreffenden Teiltones hervorgerufen wird. Entsprechend der in [4] genannten Voraussetzung für die Existenz einer Spektraltonhöhe wird die Konstante $c = 3$ dB gesetzt. Dieser Wahl liegt die Annahme zu Grunde, daß ein Partialton als vollständig verdeckt betrachtet werden kann, wenn er zu keinem Zeitpunkt eine Spektraltonhöhe ausbildet. Die frei wählbare Anzahl nicht maskierter Teiltöne x kann so gewählt werden, daß eine bestimmte prozentuale Reduktion erreicht wird. Die *prozentuale Reduktion* ist als das Verhältnis der Anzahl entfernter Partialtöne zur Gesamtzahl der ursprünglich vorhandenen Partialtöne definiert.

3. Auditive Überprüfung des Extraktionsverfahrens

Um die Brauchbarkeit des Extraktionsverfahrens zu untersuchen, wurden aus 3 Tonlagen (C2, C4 und C6) jeweils vier Klänge verschiedener Flügel (b, i, s, y) mit diesem Verfahren bearbeitet. Die auf diese Weise reduzierten TTZM der zwölf Flügelklänge wurden anschließend resynthetisiert und in Paarvergleichen mit den entsprechenden Originalklängen sieben Versuchspersonen dargeboten. Die drei Tonlagen wurden dabei in getrennten Hörversuchen untersucht. Jedes Paar wurde einmal wiederholt, bevor die Versuchsperson ein Distanzmaß für die empfundene Unähnlichkeit anzugeben hatte, wobei hierfür der Zahlenbereich von 0 (kein Unterschied) bis 9 (deutlichster Unterschied) vorgegeben war. Alle Paarvergleiche wurden jeweils viermal in randomisierter Folge dargeboten. Als Referenz wurden außerdem noch Paarvergleiche zwischen Klängen unterschiedlicher Flügel dargeboten, um somit anhand der Ergebnisse früherer Hörversuche die unterschiedlichen Skalen der drei Tonlagen miteinander vergleichen zu können. In einer vorangestellten Übung konnten die Versuchspersonen die wahrgenommenen Unterschiede an den vorgegebenen Wertebereich anpassen. An den Experimenten nahmen normalhörende Versuchspersonen im Alter von 25 bis 35 Jahren teil, die über eine gute bis sehr umfassende musikalische Bildung verfügten.

In Tab.1 sind die Ergebnisse der drei Tonlagen zusammengefaßt. Für jeden Flügelklang ist die mit dem Extraktionsverfahren erreichte prozentuale Reduktion angegeben. Weiterhin ist der Median und Interquartilbereich aller Bewertungen angegeben, die beim Vergleich zwischen dem mit dem Verfahren erzeugten Klang (bV, iV, sV, yV) und dem entsprechenden originalen Flügelklang abgegeben wurden.

Tab.1: Versuchsergebnisse der drei Tonlagen C2, C4 und C6. Für jeden durch das Verfahren erzeugte Klang (bV, iV, sV, yV) ist die erreichte prozentuale Reduktion und der beim Vergleich mit dem originalen Flügelklang ermittelte Median und Interquartilbereich aller abgegebenen Bewertungen angegeben.

Tonlage	bearb. Klang	Reduktion [%]	Median	1. Quartil	3. Quartil
C2	bV	58	0	0	1
	iV	57	0	0	1
	sV	62	0	0	1
	yV	46	0	0	1
C4	bV	20	0	0	1
	iV	30	0	0	1
	sV	35	0	0	1
	yV	28	0	0	1
C6	bV	43	0	0	1
	iV	43	5	4	6
	sV	33	0	0	2
	yV	14	0	0	1,5

In allen drei Tonlagen zeigt sich bis auf einer Ausnahme in der Tonlage C6 (iV), die im nachfolgenden noch genauer diskutiert wird, daß trotz der hohen Reduktionen praktisch keine Unterschiede zu den Originalklängen wahrgenommen werden können. Dies bestätigt also, daß das Extraktionsverfahren wirklich nur irrelevante Informationen entfernt.

Die in der hohen Tonlage C6 wahrgenommenen Unähnlichkeiten sind entsprechend den Aussagen der Versuchspersonen und den Ergebnissen eines früheren Hörversuchs [5], als sehr klein einzustufen. Dabei ist es sinnvoll den für den Klang "iV" angegebenen Median der Distanzangaben mit Hilfe der Referenzmaße, die beim Vergleich unterschiedlicher Flügelklänge abgegeben wurden, an die Skalen der beiden anderen Tonlagen anzupassen. Für den Median des Klangs "iV" ergibt sich dabei der Wert 1,5, für den 1. Quartil 1,2 und für das 3. Quartil 1,8. Dementsprechend ist also auch für diesen Klang "iV" trotz der hohen Reduktion von 43% nur ein sehr kleiner Unterschied zum originalen Flügelklang wahrzunehmen, wobei dieser entsprechend den Aussagen einiger Versuchspersonen nur im direkten Paarvergleich zu hören ist.

Zusammenfassung

Mit dem beschriebenen Extraktionsverfahren können hohe Reduktionen der TTZM von Flügelklängen durch das Entfernen verdeckter Spektralkomponenten erreicht werden. Die praktisch unmögliche Unterscheidung derart reduzierter Klänge von ihren Originalaufnahmen bestätigt hierbei, daß das Verfahren tatsächlich nur gehörirrelevante Informationen entfernt.

Herrn Prof. Dr.-Ing. E. Terhardt danke ich für die vielen wertvollen Anregungen und die kritische Durchsicht des Manuskripts.

Literatur

- [1] Heinbach, W. (1988). *Acustica*, 67: 113-121.
- [2] Schöne, P. (1979). *Acustica*, 43: 197-204.
- [3] Sonntag, B. (1982/83). *Acustica*, 52: 95-97.
- [4] Terhardt, E. (1979). *Hearing Research*, 1: 155-182.
- [5] Valenzuela, M. N. (1995). *Fortschritte der Akustik, DAGA 95*: 587-590.
- [6] Zwicker, E., Feldtkeller, R. (1967). *Hürzel*, Stuttgart, 2. Edition.
- [7] Zwicker, E., Jaroszewski, A. (1982). *J. Acoust. Soc. Am.*, 71: 1508-1512.