# Mechanismen für die Detektion von granular-periodischen Wiederholungen

Lara Stürenburg[1], Arne Oetjen[2], Steven van de Par[2]

[1] *Institut für Hörtechnik und Akustik, RWTH Aachen University, Deutschland, E-Mail: lst@akustik.rwth-aachen.de*

[2] *Abt. Akustik, Dept. für Med. Physik und Akustik, CvO Universität Oldenburg, Deutschland, E-Mail: arne.oetjen@uol.de*

## Einleitung

Ein wichtiges Element der auditorischen Wahrnehmung sind Mechanismen für die Detektion zeitlicher Strukturen von Geräuschen, wodurch Töne wahrgenommen und Sprache verstanden werden kann. Die zeitliche Struktur von weißem Rauschen ist zufällig verteilt und Menschen sind nicht in der Lage Informationen aus weißem Rauschen herauszuhören. Wenn ein weißes Rauschen periodisch ohne Pause und mit einer Frequenz von $\geq$1 Hz wiederholt wird, werden Details in der Struktur des Rauschens verstärkt und wahrnehmbare Merkmale werden informativ [1][2]. Auch unregelmäßige, wenige Wiederholungen können wahrgenommen werden. So gab es Experimente bei denen die Unterscheidbarkeit zwischen zwei Rauschabschnitten untersucht wurde [3][4]. In einer vorangeganenen Studie [4] wurde entdeckte, dass sich die Fähigkeit, zwei Rauschstimuli zu unterscheiden, verbesserte, wenn die zeitliche Länge der zwei Stimuli größer wird (bis ca. 25 ms). Es wurde gefolgert, dass die Unterscheidbarkeit sowohl von der peripheren Informationsmenge des auditorischen Objekts abhängt, als auch von der Kapazität, diese Informationen zu verarbeiten [3]. Bei weiteren Experimenten wurde gezeigt, dass Menschen nicht nur schnell erlernen Wiederholungsrauschen zu erkennen [5], sondern dass sie auch uninformative Stimulusdetails ignorieren und Stimuli eher objektbasiert verarbeiten [6]. Es gibt viel Forschung zur Detektion und Diskriminierung von Wiederholungsrauschen und viele Studien über das Wahrnehmen und Lernen von Wiederholungsrauschen. Trotzdem bleibt die Frage: Wie werden solche zeitliche Strukturen im auditorischen System wahrgenommen und verarbeitet? In dieser Studie bestehen die präsentierten Stimuli aus mehreren Rauschabschnitten, die jeweils einmal wiederholt werden. Zwischen den Rauschabschnitten der Stimuli wurden Lückenelemente eingefügt, um den Einfluss der Objektformierung zu beobachten [6]. Der Bereich der Wiederholungsrate (10-30 Hz) der Stimuli ist inspiriert durch die Studie von Hanna [4].

## Experiment I

### Vorrichtung, Verfahren und Teilnehmende

Alle Stimuli waren künstliche Geräusche, die digital bei einer Abtastrate von 44100 Hz an einem Computer mit der Software Matlab 2019b generiert wurden. Ein Audio-Interface RME Fireface UC wurde verwendet, um die digitalen Signale zu analogen Signalen zu transformieren. Die Stimuli wurden diotisch über die Kopfhörer (Sennheiser HD 650) in einer schallarmen Hörkabine bei einem Pegel von 65 dB SPL präsentiert. Der Versuch wurde in einem 3-AFC Verfahren durchgeführt mit der „2-down 1-up“-Regel und dem Verhältnis von Rauschabschnittsgröße zu Lückengröße als adaptive Variable. Es nahmen 16 Personen am Experiment teil, 8 Frauen und 8 Männer. Darunter waren die Erstautorin und der Zweitautor. Die Teilnehmenden hatten ein medianes Alter von 24,5 Jahren und gaben an normalhörend zu sein. Die Teilnehmenden wurden für ihre Teilnahme vergütet oder waren Teil der Arbeitsgruppe.

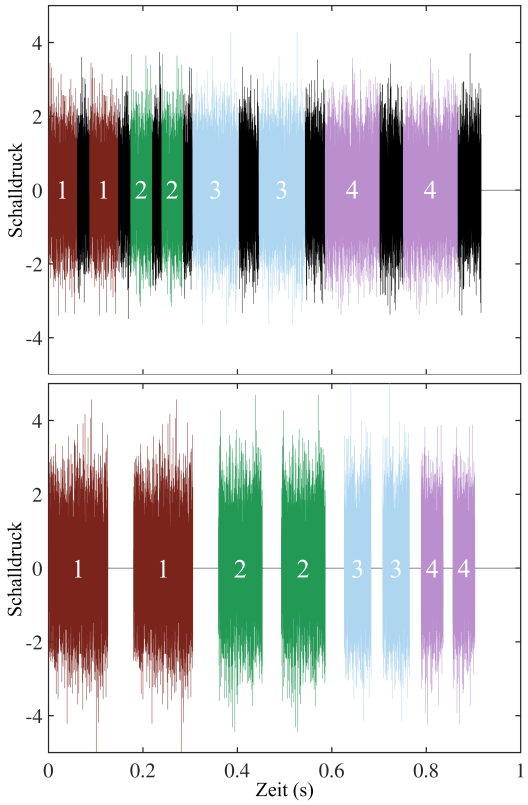


**Abbildung 1:** Struktur zweier Stimulus-Beispiele mit einer Wiederholungsrate von 10 Hz. Die sich wiederholenden Rauschabschnitte haben die gleiche Nummer und Farbe. Die Rauschabschnittsgröße beträgt hier 70%. Oben ist die Noise-Kondition und unten die Zero-Kondition abgebildet.

### Stimuli

Jeder Stimulus war maximal eine Sekunde lang und enthielt mehrere Anteile von weißen Rauschen und Lücken. Die maximale Kondition dieses Verhältnisses würde in nur Rauschen und die minimale Kondition in nur Lücken resultieren. Die rhythmische Struktur der Stimuli wurde zufällig generiert, sodass diese für jede Präsentation unterschiedlich sein konnte. In Abbildung 1 ist die Struktur von zwei Stimulus-Beispielen zu sehen. Die Rauschabschnitte wurden mit einer bestimmten Wiederholungsrate einmal wiederholt und sind in Abbildung 1 nummeriert und farblich markiert. Wie viele Rauschabschnitte ein

**Tabelle 1:** Messungen 1-12 des Experiments I. Spezifiziert ist die Wiederholungsrate in Hz und die Art der Lückenfüllung.

| Nr. | Wdh.-rate (Hz) | Lückenfüllung |
|---|---|---|
| 1 | 10 | Noise |
| 2 | 15 | Noise |
| 3 | 20 | Noise |
| 4 | 30 | Noise |
| 5 | 10 | Zero |
| 6 | 15 | Zero |
| 7 | 20 | Zero |
| 8 | 30 | Zero |
| 9 | 10 | Sinusoid |
| 10 | 15 | Sinusoid |
| 11 | 20 | Sinusoid |
| 12 | 30 | Sinusoid |

Stimulus enthält, hängt von der Wiederholungsrate ab (vgl. Tab. 1). Die Wiederholungsrate der Stimuli wurde für die einzelnen Rauschabschnitte mit 0,5 Oktaven gejittert, wodurch die zeitliche Länge jedes Rauschabschnittes unterschiedlich war. Die adaptive Variable in diesem Experiment war die Rauschabschnittsgröße. Wenn die Rauschabschnittsgröße kleiner als 100% war, sind Lücken zwischen den Rauschabschnitten aufgekommen. Die Breite der Lücken stand im Verhältnis zur zeitlichen Länge der Rauschabschnitte. Wenn die Rauschabschnittgröße bei 0% lag, bestand der Stimulus nur noch aus den Lücken. Die Rauschabschnittsgröße der Startkondition war immer 70% (vgl. Abb. 1). Die Lücken wurden unterschiedlich gefüllt (s. Tab. 1). In der „Noise"-Kondition wurde jede Lücke mit neu generiertem weißen Rauschen gefüllt. Für die „Zero"-Kondition enthielten die Lücken kein Signal. In der „Sinusoid"-Kondition wurden die Lücken mit einem Sinuston von 1 kHz gefüllt, welcher immer bei einer Phase von 0° begann. Bei 0% Rauschabschnittsgröße bestand der Stimulus entweder aus nur weißem Rauschen, Stille oder einem Sinuston. Die Referenzstimuli hatten eine ähnliche Struktur wie der Zielstimulus, enthielten jedoch keine Wiederholungen. In der „Noise"-Kondition bestanden die Referenzstimuli aus kontinuierlichem weißen Rauschen. Für die „Zero"-Kondition waren die Referenzstimuli weißes Rauschen mit kurzen Pausen. In der „Sinusoid"-Kondition bestanden die Referenzstimuli aus weißem Rauschen mit Sinustönen dazwischen. Für jede Stimulipräsentation wurden die Ziel- und Referenzstimuli neu, aber auf gleiche Weise generiert. Dies konnte dazu führen, dass die Anzahl der Rauschabschnitte der Ziel- und Referenzstimuli unterschiedlich war. Nach der Generierung wurden die Ziel- und Referenzstimuli durch ihren Effektivwert (RMS) geteilt, mit einem Hann-Fenster gefenstert und auf 65 dB kalibriert.

### Ergebnisse

Nach dem Experiment haben sich zwei Gruppen von Teilnehmenden gebildet, welche in einigen Konditionen ein ausgeprägtes, unterschiedliches Antwortverhalten aufwiesen. In Abbildung 2 sind die Ergebnisse der

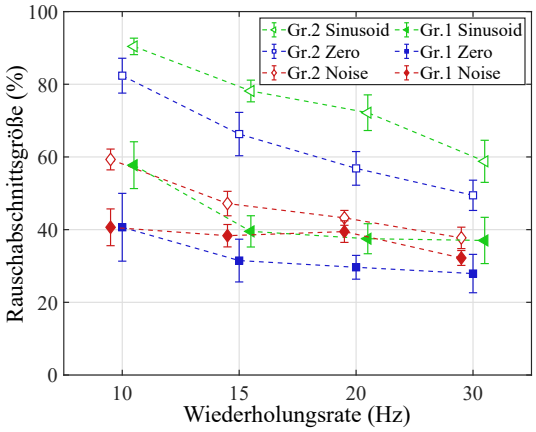


**Abbildung 2:** Rauschabschnittsgröße in % über die Wiederholungsrate in Hz. Abgebildet sind die interindividuellen Mittelwerte der zwei Gruppen (offene und gefüllte Symbole). Die roten Diamanten zeigen die Messungen in der Noise-Kondition. Die blauen Quadrate zeigen die Messungen in der Zero-Kondition. Die grünen Dreiecke zeigen die Messungen in der Sinusoid-Kondition.

zwei Gruppen mit offenen bzw. gefüllten Symbolen für alle drei Konditionen dargestellt. Die Abbildung zeigt die interindividuellen Mittelwerte der Rauschabschnittsgröße über die Wiederholungsrate für die Messungen 1-12 (s. Tab. 1). Mit roten Diamanten sind die Ergebnisse der Noise-Kondition zu sehen. Die Ergebnisse der Zero-Kondition sind mit blauen Quadraten und die Ergebnisse der Sinusoid-Kondition sind mit grünen Dreiecken gekennzeichnet. Die Rauschabschnittsgröße, die gebraucht wird, um das granular-periodische Rauschen zu detektieren, ist am größten für 10 Hz Wiederholungsrate und am niedrigsten für 30 Hz. Der Verlauf der Kurven aller drei Kondition ist ähnlich. In allen drei Konditionen sind die Ergebnisse der Gruppe 2 höher, als die der Gruppe 1. Während Gruppe 1 die niedrigsten Rauschabschnittsgrößen in der Zero-Kondition erzielte, sind die niedrigsten Werte der Gruppe 2 in der Noise-Kondition. Die Ergebnisse der zwei Gruppen liegen in der Noise-Kondition nah beieinander, während sie in der Zero- und Sinusoid-Kondition bis zu 42% auseinanderliegen.

## Experiment II

### Vorrichtung, Verfahren und Teilnehmende

Die Vorrichtung und das Verfahren waren in diesem Experiment die gleichen wie in Experiment I. Die adaptive Variable war jedoch nicht die Lückengröße, sondern die Trennfrequenz oder der Maskierungspegel. Es nahmen 10 Personen am Experiment teil, 3 Frauen und 7 Männer. Alle Personen haben auch schon an Experiment I teilgenommen und alle Personen aus Gruppe 1 von Experiment I waren darunter. Zwei der Teilnehmenden waren die Erstautorin und der Zweitautor. Die Teilnehmenden hatten ein medianes Alter von 25,5 Jahren und gaben an normalhörend zu sein. Die Teilnehmenden wurden für ihre Teilnahme vergütet oder waren Teil der Arbeitsgruppe.

### Spektralgetrennte Maskierung

Die Ziel- und Referenzstimuli enthielten ein Maskierungssignal. D.h. ein Ziel- und ein Maskierungssignal wur-

**Tabelle 2:** Messungen 1-12 des Experiments II. Spezifiziert ist die Wiederholungsrate des Ziel- und Maskierungssignals

| Nr. | Wdh.-rate Ziel (Hz) | Wdh.-rate Maskierer (Hz) |
|---|---|---|
| 1 | 15 | 15 |
| 2 | 15 | 30 |
| 3 | 30 | 30 |
| 4 | 30 | 15 |
| 5 | 15 | w. Rauschen |
| 6 | 30 | w. Rauschen |
| 7 | 15 | 15 |
| 8 | 15 | 30 |
| 9 | 30 | 30 |
| 10 | 30 | 15 |
| 11 | 15 | w. Rauschen |
| 12 | 30 | w. Rauschen |

den zu einem Zielstimulus addiert und ein Referenz- und Maskierungssignal zum Referenzstimulus. Die Ziel- und Maskierungssignale wurden wie die Stimuli der Noise-Kondition aus Experiment I generiert, jedoch ohne Jitteranteil. Die Zielsignale hatten eine feste Rauschabschnittsgröße von 100% und die Maskierungssignale von 70%. Tabelle 2 zeigt die Wiederholungsrate des Ziel- und Maskeriungssignals. Bei 4 Messungen hatte das Maskierungssignal keine Wiederholungsrate und bestand aus kontinuierlichem weißen Rauschen. In den Messungen 1-6 war das Zielsignal (Target) tiefpassgefiltert und das Maskierungssignal (Masker) hochpassgefiltert. In den Messungen 7-12 war es andersherum (vgl. Abb. 3). Die Variable der Messungen war die Trennfrequenz. D.h. die Abschneidefrequenz von Hoch- und Tiefpassfilter wurde während der Messung angepasst. In Abbildung 4 ist

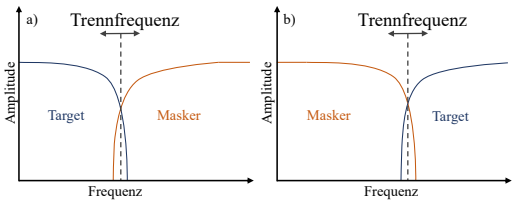


**Abbildung 3:** Illustration für Messung 1-6 (links) und 7-12 (rechts). Ziel- und Maskierungssignal sind hoch- bzw. tiefpassgefiltert.

die Trennfrequenz über die Wiederholungsrate von Ziel- und Maskierungssignal aufgetragen. Der interindividuelle Mittelwert der Teilnehmenden ist für die Messungen 1-6 mit roten offenen Symbolen markiert und für die Messungen 7-12 mit blauen gefüllten Symbolen. Die schmalste Bandbreite des Targetsignals wurde in beiden Messvarianten für die Kondition 30-WN erreicht und die breiteste Bandbreite für die Kondition 15-30.

### Breitbandige Maskierung

Tabelle 3 zeigt die Eigenschaften der Messungen 13-18. Die Zielsignale hatten wieder eine Rauschabschnittsgröße von 100% und die Maskierungssignale von 70%. Die Wiederholungsrate der Maskierungssignale war bei den Mes-

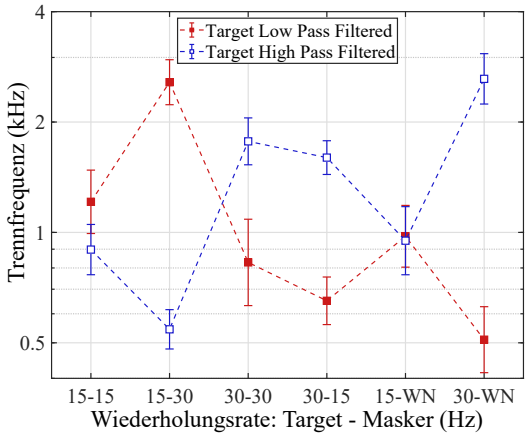


**Abbildung 4:** Trennfrequenz in kHz über die Wiederholungsrate von Ziel- und Maskeriungssignal in Hz. In rot Messungen 1-6 und in blau Messungen 7-12. WN: weißes Rauschen.

**Tabelle 3:** Messungen 13-16 des Experiments II. Spezifiziert ist die Wiederholungsrate des Ziel- und Maskierungssignals sowie der Jitteranteil des Maskierungssignals.

| Nr. | Wdh.-rate Ziel (Hz) | Wdh.-rate Maskierer (Hz) | Jitteranteil (Okt.) |
|---|---|---|---|
| 13 | 15 | 15 | ±0 |
| 14 | 15 | 30 | ±0 |
| 15 | 30 | 30 | ±0 |
| 16 | 30 | 15 | ±0 |
| 17 | 15 | 15 | ±0,5 |
| 18 | 15 | 30 | ±0,5 |

sungen 17 und 18 gejittert. Die Signale wurden nicht spektral gefiltert. Es wurde breitbandig maskiert. Die Variable war der Maskierungspegel in dB, wie in Abb. 5 illustriert. Abbildung 6 zeigt den SNR zwischen Ziel- und Maskierungspegel über die Wiederholungsrate von Ziel- und Maskierungssignal. Die interindividuellen Mittelwerte der Messungen 13-18 sind zu sehen. Wenn die Wiederholungsrate von Ziel- und Maskierungssignal gleich sind, ist der SNR negativ. Sind die Wiederholungsraten der Signale unterschiedlich und das schon minimal (vgl. mit und ohne Jitteranteil) ist der resultierende SNR positiv.

## Diskussion

Die Ergebnisse aus Experiment I zeigen, dass die Detektion granular-periodischer Stimuli leichter wird, wenn die Wiederholungsrate des Zielstimulus höher ist. Basie-

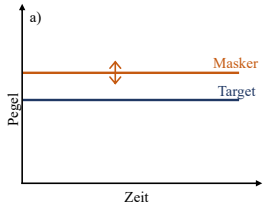


**Abbildung 5:** Illustration der Messungen 13-18. Der Pegel des Maskierers wird variiert.

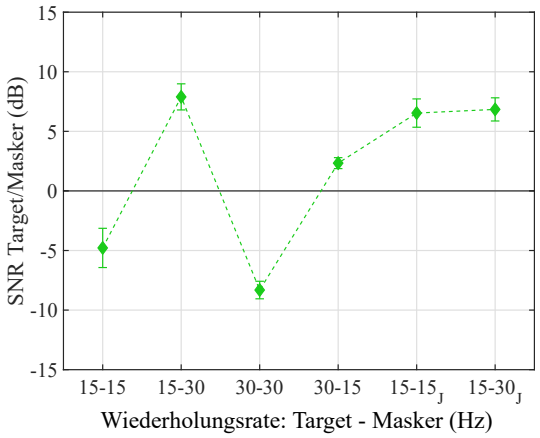


**Abbildung 6:** SNR in dB über die Wiederholungsrate von Ziel- und Maskierungssignal in Hz für die Messungen 13-18. J: gejitterte Wiederholungsrate.

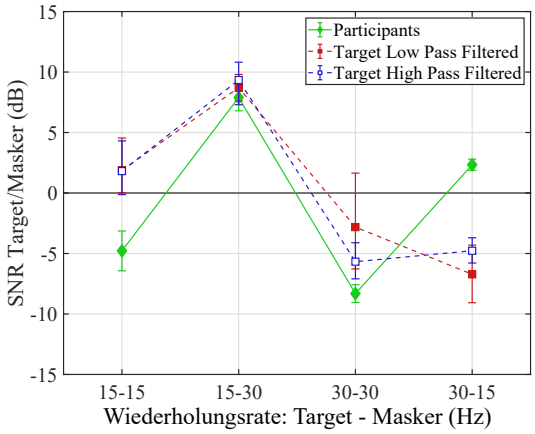


**Abbildung 7:** SNR in dB über die Wiederholungsrate von Ziel- und Maskierungssignal in Hz. In grün sind die Messungen 13-16. In rot sind die errechneten SNRs für die resultierende Trennfrequenz der Messungen 1-4 und in blau der Messungen 7-10.

rend auf das Antwortverhalten der Teilnehmenden konnten diese in zwei Gruppen eingeteilt werden. Es kann vermutet werden, dass für die Detektion des Zielstimulus die zwei Gruppen auf unterschiedliche Erkennungsmerkmale sensitiv sind. Eventuell nahmen die Teilnehmenden in einigen Konditionen ein weiteres akustisches Objekt wahr, welches für die granular-periodische Detektion ignoriert werden musste.

Experiment II zeigte, dass bei einer spektralen Maskierung die zum Detektieren gebrauchte Anzahl an Frequenzkanälen abhängig von den Wiederholungsraten der Ziel- und Maskierungssignalen ist. Bei breitbandiger Maskierung ist die Detektion ebenfalls von den Wiederholungsraten abhängig. Für weitere Untersuchungen wurde der SNR für die Ergebnisse der Messungen 1-4 und 7-10 aus Experiment II bestimmt. Der SNR wurde errechnet, indem die tief- und hochpassgefilterten Ziel- und Maskierungssignale mit der ermittelten Trennfrequenz reproduziert wurden. Die Signale wurden mit einem Mittelohrfilter gefiltert [7], bevor der SNR berechnet wurde. In Abbildung 7 sind die gemessenen (grün) und errechneten (rot & blau) SNRs zu sehen, welche zum Teil übereinstimmen (Kondition: 15-30). Die geringe Anzahl an Frequenzkanälen führt also zu den gleichen SNRs wie eine Pegelanpassung. Daraus folgt, dass für granular-periodische Stimuli die auditorischen Erkennungsmerkmale über die Frequenzkanäle summiert werden müssen. Dies steht in einem Widerspruch zum Prinzip der spektralen Trennung in der auditorischen Verarbeitung.

## Modellierung

Es wurde versucht, die Ergebnisse aus Experiment I mit dem Perzeptionsmodell nach Dau et al. [8] zu detektieren und festgestellt, dass die Modulationsmerkmale der internen Repräsentation für den optimalen Detektor des Modells nicht ausreichen, um granulare Periodizität zu detektieren. Nur durch Erhöhung der Stimuluslänge um mehr als das Zehnfache konnten 16 von 27 Stimuluskonditionen detektiert werden. Die Ergebnisse von Experiment II legen nahe, dass eine frequenzübergreifende Integration der Periodizitätsverarbeitung im auditorischen System stattfindet. Es gibt einen ersten Modellansatz, der die Ergebnisse für alle granular-periodischen Stimuli detektieren und zum Teil reproduzieren kann. Dieses Modell basiert auf einer schnellen Fouriertransformation (FFT) und auf der Summation über alle Frequenzkanäle. Das wiederum verstärkt die Hypothese, dass durch Summation über die Modulationspektren der Frequenzkanäle granular-periodische Stimuli detektiert werden.

## Literatur

[1] Kaernbach, C.: On the consistency of tapping to repeated noise. J. Acoust. Soc. Am. 92 (1992), 788-793

[2] Kaernbach, C.: Temporal and spectral basis of the features perceived in repeated noise. J. Acoust. Soc. Am. 94 (1993), 91-97

[3] Goossens, T., van de Par, S., and Kohlrausch, A.: On the ability to discriminate gaussian-noise tokens or random tone-burst complexes. J. Acoust. Soc. Am. 124 (2008), 2251-2262

[4] Hanna, T. E.: Discrimination of reproducible noise as a function of bandwidth and duration. Perception & Psychophysics 36 (1984), 409-416

[5] Agus, T. R., and Pressnitzer, D.: The detection of repetitions in noise before and after perceptual learning. J. Acoust. Soc. Am. 134 (2013), 464-473

[6] Goossens, T., van de Par, S., and Kohlrausch, A.: Gaussian-noise discrimination and its relation to auditory object formation. J. Acoust. Soc. Am. 125 (2009), 3882-3893

[7] Jepsen, M. L., Ewert, S. D., and Dau, T.: A computational model of human auditory signal processing and perception. J. Acoust. Soc. Am. 124 (2008), 422-438

[8] Dau, T., Kollmeier, B., and Kohlrausch, A.: Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102 (1997a), 2892-2905