Gaussian-noise discrimination and auditory object formation

The work described in this thesis was financially supported by Philips Research Laboratories Eindhoven and was carried out under the auspices of the J.F. Schouten School for User-System Interaction Research.

An electronic copy of this thesis in PDF format is available from the website of the library of the Technische Universiteit Eindhoven (http://www.tue.nl/bib).

© 2008, Tom L.J. Goossens, The Netherlands
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the author.

Cover design: Tom Goossens

Printing: Universiteitsdrukkerij Technische Universiteit Eindhoven

A catalogue record is available from the Eindhoven University of Technology Library

ISBN 978-90-386-1346-8

Gaussian-noise discrimination and auditory object formation

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op woensdag 3 september 2008 om 14.00 uur

door

Tom Lambertus Johannes Goossens

geboren te Uden

Dit proefschrift is goedgekeurd door de promotoren:

prof.Dr. A.G. Kohlrausch
en
prof.Dr. T. Dau

Copromotor:
dr.ir. S.L.J.D.E. van de Par

Contents

1 Introduction

When someone strikes a string of a guitar, this string starts to vibrate. Via the guitar's body, the vibrations of the string are transduced to the air, and radiate through the air away from their physical source. When these vibrations arrive at the eardrums of a listener, they enter the hearing system and may cause a sensation. These sensations are translated by us into mental descriptions of the physical events by which they were caused (Bregman, 1990). We can do this because our hearing system has evolved in a way that enables us to make inferences about happenings in the physical world from the air vibrations they cause (Dennett, 1997). Within certain limits, we are capable of detecting the occurrence of a physical event and of localizing its source. Furthermore, we are able to discriminate between different events, to categorize them, and to order them according to some physical or perceptual parameter (cf. Yost and Sheft, 1993). These abilities help us to interpret these vibrations and make inferences about the events that occur around us.

In every day situations, it is common that several sound-producing events happen at the same time. For instance, a car driving by the window, the computer fan, a person typing on a keyboard, and your colleague talking to you in the office. Their waveforms are mixed in the air before they arrive at the eardrums of a listener. One of the very difficult tasks that our hearing system performs, which we take almost for granted, is to somehow separate the various source components of this mixed waveform and perceive them as separate auditory streams. These auditory streams are again dividable into single events originating from the same source, e.g., the individual keystrokes of the person typing. We are able to focus our attention to one of the auditory streams while ignoring the other streams (e.g., Bregman, 1990; Alain and Arnott, 2000). This illustrates that our hearing system can act as a filter, in the broad sense of the word, that enables us to sift the valuable sensory input from the not so valuable input, which is sometimes called noise. Arguably, it is one of the main tasks of the auditory system to extract only the useful information from the plethora of incoming stimuli and it is important to learn more about the nature of the information processing of the auditory system.

This introduction will create a context for this thesis by reviewing the literature on the perception of noise, auditory objects, memory, and informational masking. In addition, it will define in which way the concept of information is used in this study.

1.1 The perception of noise

Noise in itself is a very broadly used term. In normal language, when someone talks about noise, it usually refers to unwanted sensory input, e.g., sound from loud trains or airplanes that cause acoustic pollution, or certain types of music that he or she does not like. In the context of psychoacoustics, noise is often used as a masker that increases the difficulty for a listener in hearing or attending to a target signal.

However, the term noise is not always used to indicate *unwanted* signals; it can also mean that the signal is a randomly fluctuating signal with a high degree of unpredictability. In this case, it is often described by its statistical properties (Rice, 1944). Humans use this type of signal very frequently in speech communication. For example, each time a person pronounces the sound "*f*", he or she is generating noise. All the unvoiced fricatives in speech (*s*, *f*, *sh*, etc.) are noise bursts that are generated by causing turbulence in the air (Rabiner and Juang, 1993) and which are spectrally shaped by the vocal tract. These noise bursts have similar statistical properties to bandpass-filtered Gaussian-noise bursts.

According to information theory, the information content of a signal is proportional to its unpredictability, since no information is gained from the occurrence of a completely predictable signal. Therefore, an unpredictable signal, like Gaussian noise, can contain much information (Shannon, 1948). The information content of a Gaussian-noise token is proportional to the product of its bandwidth and duration (Hartley, 1928). From a perceptual point of view, however, the information content in a piece of Gaussian noise is considered to be not so high because different realizations of a noise sound very much alike. In other words, the percept of a *new* realization of the noise is predictable because it will sound much like a previous realization. So, whereas mathematically it is possible to create a large number of different noise realizations from the same statistical process, the number of perceptually different realizations will be much smaller.

Although different Gaussian-noise realizations sound alike, Hanna (1984) found that humans are able to discriminate between them. It appeared that for most conditions discrimination ability increased with bandwidth (and therefore with increasing stimulus information). The ability to discriminate also increased with duration, but only up to 25–100 ms. Above this duration, the ability to discriminate decreased with *increasing* duration. Apparently, stimulus information and perceptual information are not monotonically related to each other in the temporal dimension. Hanna concluded that this nonmonotonic relation was primarily due to processes that were not of peripheral origin, but rather of more central origin (e.g., memory and decision making).

Although listeners' performance is poor when trying to discriminate noise tokens with a duration above 400 ms, when a single noise token is presented repeatedly in a cyclic pattern, listeners can detect its periodicity for noise-token durations of up to 10 seconds (Warren *et al.*, 2001). It must be noted, though, that listeners do not always have a global percept of the segments encompassing the entire repeated segment, but they rather perceive salient features within the noise that allow them to detect the repetition. At first, such repetitive noise sounds like ordinary Gaussian noise. However, after several repetitions the brain starts to detect the reoccurrence of the same segment (or features) and the perception of this sound starts to have, as Guttman and Julesz (1963) described for the first time, a kind of "whooshing" or "motorboating" quality. Moreover, certain features start to emerge which are often described as "clanks" or "rasping". In a study of Kaernbach (1993) with such cyclic-noise stimuli, listeners were asked to tap to the rhythm of the repetition at the location of a salient feature. It was shown that a listener tends to consistently tap to the same part of the noise. Apparently, at this part the listener perceives a salient feature. Some of these features were detected by all listeners, while it was also observed that, for other parts of the noise, different listeners selected different features. Using the tapping experiment in combination with some elegant stimulus manipulations, Kaernbach (1993) found that the duration of these features was usually not more than 100 ms. Their spectral extent varied from rather narrow, approximately the width of an auditory filter, to relatively wide, in the range of several auditory-filter bandwidths. This shows that, although Gaussian noise at first might seem like a homogeneous signal, there are certainly perceptible features that can make different realizations distinguishable from each other.

1.2 Auditory objects

The office scene, described at the beginning of this introduction, is built up from a large number of acoustical events: individual keystrokes of the person typing, words (or perhaps their phonemes) spoken by your colleague, the rotating computer fan. The percept, or mental description, of such an individual event is called an auditory object, or unit in the nomenclature of Bregman (1990). Events caused by the same source, e.g., the keystrokes, may be perceptually grouped together into an auditory stream. In this section we will focus on the auditory objects of which an auditory stream consists.

In the context of selectively attending to auditory objects, Alain and Arnott (2000) adopt the definition of Bregman (1990) that an auditory object "[...] is the percept of a group of sounds as a coherent whole seeming to emanate from a single source." An important element of this definition is that the object is perceived as a whole, a perceptual unit.

Just like a visual object is limited in spatial dimensions, an auditory object is spatially limited (Kubovy and Valkenburg, 2001). In addition, an auditory object is limited in time and therefore it has a beginning and an end. It is of interest to understand the specific cues that can initiate or terminate an auditory object. An important cue is, e.g., a sudden rise in intensity (Bregman, 1990). Yost (1991) distinguishes at least seven physical parameters that contribute to the formation of auditory objects: spectral separation, intensity profile, harmonicity, spatial separation, temporal separation, common temporal onsets and offsets, and coherent slow temporal modulation. Depending on such parameters, the spectro-temporal components in a sound mixture are either fused into an auditory object or segregated into different objects.

Other properties of auditory objects are that they can exist in different timescales and that smaller auditory objects can be grouped into larger auditory objects (Bregman, 1990). In addition, auditory objects are often assumed to adhere to the principle of exclusive allocation which states that any part of the sensory input can only belong to one object at a time (Köhler, 1947; Winkler *et al.*, 2006). Furthermore, Bregman (1990) adopted the principle from Gestalt psychology that homogeneous perceptual inputs do not contain objects. Only when discontinuities appear, e.g., arising from changes in the physical parameters mentioned above, can the input be organized into objects.

In an elaborate review of neurological studies, Näätänen and Winkler (1999) searched for the physiological stage at which the neural code elicited by a sound becomes a percept. They distinguished three forms of auditory stimulus representation: (1) the *afferent activation pattern*, i.e., the neural firing in the auditory nerve. This activation pattern is transformed into (2) a number of separate *sensory feature traces* for different sorts of stimulus information such as, e.g., pitch, loudness, spatial locus of origin. Listeners are not likely to have direct access to these feature traces. In addition, these feature traces are not unified and are generated in different loci of the auditory cortex. Finally, the sensory feature traces are integrated into (3) a unitary *sensory stimulus representation*. At this stage the neural code becomes a substrate for the percept of a unitary sound object.

1.3 Auditory sensory memory

To discriminate two auditory objects (e.g., two Gaussian-noise tokens) presented sequentially, listeners need to retain detailed information about both tokens for a short period of time to be able to compare them. This is done in auditory sensory memory. Theories of auditory sensory memory often distinguish two modes of operation: *a sensory trace mode* and a *categorical mode* (e.g., Durlach and Braida, 1969). The sensory trace mode contains detailed information about the perceived sound and lasts for up to 10 seconds (Sams *et al.*, 1993). The information in the categorical mode, also called context coding mode (Durlach and Braida, 1969), is recoded into a symbolic (e.g., verbal) representation. This information is less detailed than the sensory trace information but lasts much longer.

Cowan (1984) presents evidence for the existence of two stores within the sensory trace mode, distinguished by the timespan over which they operate: a short auditory store, which decays after approximately 200 to 300 ms, and a long auditory store which retains auditory information for about 10 seconds (Cowan, 1984; Kaernbach, 2004). According to Cowan (1984), the information in the short auditory store is experienced as sensation and is continuously overwritten by subsequent sensory input. The long auditory store is experienced as memory and may be masked depending on, e.g., similarity with previous stimuli. In addition, the information in the short store is relatively unanalyzed whereas the information in the long store contains information about different kinds of stimulus features. An alternative hypothesis for the organization in the sensory trace mode is presented by Clément *et al.* (1999). They hypothesized that there are multiple sub stores, each retaining information about different perceptual attributes, e.g., pitch, loudness, timbre. They support this theory with the finding that the rate of memory decay for loudness is more rapid than the rate of memory decay for pitch, and infer from this that pitch and loudness traces are not retained in one and the same auditory store.

Although it appears that human sensory memory is rich, and perhaps unlimited in the number of stimuli that can be represented concurrently (Cowan, 2005, p.113), the access to this set is limited by other cognitive processes. In the model of Broadbent (1958), the limited access to sensory memory is modeled by the presence of a limited capacity channel between the sensory memory and the higher order processes. According to Cowan (2005), this limited access is caused by a limitation of the focus of attention to pull information from sensory memory into working memory. The focus of attention is maybe best described by one of Cowan's own examples: "Metaphorically, it is as if the spotlight of attention has to be shined on the various parts of the sensory memory field before it disintegrates." (Cowan, 2005). Information from sensory memory that falls within the focus of attention may be consciously accessible.

1.4 Information in auditory stimuli

In the previous paragraph it was suggested that there is a limitation to the amount of information that can be accessed from auditory sensory memory. This raises questions about what the concept of information actually means in the context of perception. Therefore, we will now more carefully consider how stimulus details are transformed by auditory processing. As a formal definition of information we take the base-two logarithm of the number of stimuli (e.g., noise tokens) that can be distinguished by the most optimal discrimination device (cf. Shannon, 1948). In a same-different task, one could assume that the larger the number of distinguishable stimuli, and hence the amount of information, the more likely it is that two randomly generated stimuli can be distinguished.

Note however, that on the stimulus level, without assuming any source of uncertainty, e.g., in the form of internal noise, this definition of information is not very useful. Without uncertainty, any two independent tokens of noise will be distinguishable even when only a single sample is considered, thus the amount of information in this situation would be infinite. For example, one single sample of a waveform can be distinguished from another sample that is an infinitesimally small step higher or lower in level. In real situations, of course, some uncertainty is always present. In the context of auditory stimulus discrimination, the uncertainties in the auditory periphery limit the amount of information to a specific finite amount. To avoid assumptions on the uncertainty at the stimulus level, we will use the number of degrees-of-freedom of the stimulus rather than the information as defined by Shannon to characterize the number of potentially distinguishable stimuli. In band-limited noise, the number of degrees-of-freedom is proportional to the product of bandwidth and duration (Hartley, 1928; Nyquist, 1928).

Within the auditory periphery of a listener the stimuli undergo a series of linear and nonlinear transformations, such as critical-band filtering, hair-cell transduction, auditory nerve encoding, etc., which results in an afferent activation pattern (Dau *et al.*, 1996; Näätänen and Winkler, 1999). Moreover, it is often assumed that some source of internal noise limits the fidelity of this pattern (e.g., de Boer, 1966; Buus, 1990; Dau *et al.*, 1996). The internal noise introduces uncertainty into a discrimination task which reduces the number of distinguishable stimuli. In addition, nonlinear transformations affect the number of distinguishable stimuli. In the remaining data, the number of discriminable stimuli at the level of the afferent activation pattern will be smaller than at the level of the stimuli. We will refer to the amount of information at the level of the afferent activation as *peripheral information*.

In higher stages of the auditory pathway, feature extraction from the afferent activation pattern results in the emergence of sensory feature traces containing cues about, e.g., pitch, loudness, and spatial location (Näätänen and Winkler, 1999). It is likely that also at this level of processing the number of discriminable stimuli is further decreased. This can, for example, be caused by neural processing providing robustness against pitch and duration variation (Patterson *et al.*, 2007), or by limitations of the focus of attention (Cowan, 2005). Another example is that intensity discrimination is not so good as would be expected on the basis of the information carried in the auditory nerve (Delgutte, 1987; Moore, 2003)

On a perceptual level, it is useful to speak about cues that can lead to perceptual differences. Some examples of such cues are level, spectral shape, spatial location, spatial compactness, envelope distribution (van de Par and Kohlrausch, 1998; Verhey *et al.*, 2007), and modulation spectrum (Dau *et al.*, 1999). The number of degrees-of-freedom in a stimulus may influence the type of perceptual cues the listener uses for discrimination. For instance, the shape of the mean spectrum envelope of a very short burst of noise will greatly vary from token to token and hence may provide a good cue for discrimination. However, when the duration is much larger, and hence the number of degrees-of-freedom is increased, there will be less variability of the mean spectrum of the noise burst. For longer stimuli, the mean spectrum thus provides a less salient cue, which could lead to poorer discrimination performance. Other cues representing more local stimulus properties, such as instantaneous pitch or short-term envelope fluctuations, may therefore be more salient for long-duration stimuli. The cues that are available to the listeners represent what we will refer to as *perceptual information*.

Acoustical information inherently extends over time, and therefore needs to be combined over time in order to enable decision making, e.g., loudness comparison. Often this process is modeled by power integration across the signal. However, Viemeister and Wakefield (1991) showed that threshold for detecting a pair of pulses, which were presented during two gaps in a continuous noise, were lower than for either of the pulses alone. This indicates that a kind of integration had occurred which could not be explained by a simple power integration across the stimulus. It was as if the listener had combined two samples or "looks" of the signal. In the psychoacoustic model proposed by Viemeister and Wakefield (1991), differences between stimuli at different temporal locations ("multiple looks") are combined, because the combination gives more evidence that can contribute to discrimination or detection. The model of Dau *et al.* (1996) has a similar approach, but uses optimal filtering on a template of the internal representations, i.e., a computational transformation of the acoustic input representing several stages of the auditory processes. The model then correlates the relevant portions of the IR, with the template, which effectively is a matched filter operation. Both models, thus, can use an accumulation of the information present over the whole duration of the stimuli. There is no explicit assumption that restricts the length or informational content of these internal representations or the number of looks. Therefore, with increasing stimulus duration, the information available to the models also increases, and thus, their performance in discriminating between the internal representations of two noise signals will increase or saturate. This is not what was observed in the experiments of Hanna (1984), who found a nonmonotonic duration dependence for noise discrimination (cf. section 1.1).

1.5 Informational masking

A commonly used psychoacoustical paradigm is a detection experiment. In such a paradigm, a listener is presented with a signal, e.g., a tone, the properties of which change across intervals, e.g., its frequency or it being present in one of the intervals only. A task of the listener can be to respond in which interval the tone is present. This provides knowledge about how intense the signal must be before it can be detected by the listener. The level of presentation is called the detection threshold. When the signal is presented together with another stimulus, e.g., a broadband noise or another tone, the detection threshold for the signal will often be elevated. It is said that a signal below the detection threshold is masked by the presence of the other stimulus, the masker. This type of masking is called energetic masking (Durlach *et al.*, 2003; Kidd Jr. *et al.*, 2007).

However, sometimes the masking of a signal cannot be explained by the energy of the masker. For instance, in a study of Watson (1987), the detectability of a frequency change of one of the components in a ten-tone pattern was greatly influenced by the uncertainty of the frequency of the other tones. The detectability of this frequency change was much lower when the uncertainty of the other tones was high. Such masking cannot be explained by the energy of the masker. Therefore, it is named *informational masking*. While energetic masking occurs mainly in the periphery of the hearing system, informational masking occurs at a more central level (Durlach *et al.*, 2003).

In general, it can be stated that the higher the relative variability of the *context* of a to-be-detected target, the more difficult it is to detect this target. Thus, informational masking increases with the relative variability of the context in relation to the variability of the target (Kidd and Watson, 1992; Lutfi, 1993).

1.6 Thesis outline

As previously mentioned, the ability to discriminate broadband Gaussian-noise tokens reduces with increasing duration for stimuli with durations above 100 ms, despite the fact that the peripheral information increases. Below approximately 25 ms, the ability to discriminate increases with duration. Apparently, there is a nonmonotonic relationship between the amount of information elicited by the stimulus in the auditory periphery and the amount of perceptual information for this range of durations. It is one of the central goals of this study to investigate the underlying mechanism responsible for this nonmonotonic relationship.

Chapter 2 of this thesis describes a replication of one of the experiments of Hanna (1984), in which the nonmonotonic relationship between duration and discrimination ability was first shown for Gaussian noise. The chapter describes the effect of bandwidth and duration on the ability to discriminate Gaussian-noise tokens. Furthermore, additional discrimination experiments will show a relation between the number of degrees-of-freedom of the stimulus and the ability to discriminate stochastic stimuli.

Existing psychoacoustic models based on the optimal combination of peripheral information, such as multiple looks and temporal-integration models, do not predict a decreasing discrimination ability with increasing duration because all available information is employed to the advantage of discrimination. At most, the discrimination ability saturates at a certain level of performance. In chapter 3, a model is presented for simulating the nonmonotonic duration dependency found in the chapter 2. In this chapter it is proposed to add an extra stage to the psychoacoustic model of Dau *et al.* (1996), which imposes restrictions on the amount of information allowed in the internal representation of an auditory object.

To impose restrictions on the amount of information allowed in the internal representation of an auditory object, it is necessary to know where this object starts and where it ends. This is straightforward when the sound is homogeneous and has a strong onset and offset, like Gaussian-noise bursts. The study described in chapter 4 aimed to gain knowledge about the cues that can initiate the start of a new auditory object. In particular, the potential segregation cues, temporal separation, spectral separation, bandwidth, level differences, interaural level differences, and interaural time delay are adressed. The results give insight into the relative importance of these cues for the initiation of new auditory objects.

2 On the ability to discriminate Gaussian noise tokens or random tone-burst complexes[†]

Abstract

This study investigated factors that influence a listeners' ability to discriminate Gaussian-noise stimuli in a same-different discrimination paradigm. The first experiment showed that discrimination ability increased with bandwidth for noise durations up to 100 ms. Duration had a non-monotonic influence on performance, with a decrease in discriminability for stimuli longer than 40 ms. Further experiments investigated the cause for this performance decrease. They showed that discriminability could be improved when using frozen-noise tokens and by instructing listeners to focus on the stimulus endings. A final experiment, using a stimulus consisting of 5-ms Hanning-windowed tone-bursts randomly distributed over time, investigated whether stimulus duration and amount of information differently affect the processing capacity of the auditory system. Results showed that the number of degrees-of-freedom in the stimulus, not its duration, predominantly influenced the ability to discriminate. Overall, the results suggest that the discrimination performance for acoustic stimuli depends strongly on the amount of information per critical band and the capacity to process this information. This capacity seems to be limited in the temporal dimension, while extending the signal over more auditory filters does have a positive effect on performance.

[†] This chapter is based on Goossens, T., van de Par, S., and Kohlrausch, A. "On the ability to discriminate Gaussian noise tokens or random tone-burst complexes," accepted for publication to J. Acoust. Soc. Am.

2.1 Introduction

In informal listening, two tokens of noise generated by the same statistical process generally sound very similar although their waveforms are completely independent. When asked in a formal experiment to judge whether two presented noise tokens are the same or different, human listeners can respond with above-chance performance, but usually performance will not be perfect.

The ability to perform such a noise discrimination task is a function of the statistical properties of the noise. For instance, the ability to discriminate between tokens of Gaussian noise improves with increasing noise bandwidth for durations up to 100 ms (Hanna, 1984). This improvement is in line with the increase of details in the internal spectro-temporal excitation in the auditory periphery when more auditory channels are excited by the stimulus.

In contrast, the ability to discriminate noisy stimuli does not increase monotonically with an increase of stimulus duration. For example, a 400-ms noise stimulus leads to a longer internal excitation than a 25-ms noise stimulus, and thus the internal representation of a longer stimulus contains more stimulus details than that of a shorter stimulus. One might therefore expect that the ability to discriminate 400-ms noise stimuli is higher than that for 25-ms noise stimuli. Hanna (1984), Heller and Trahiotis (1995) and Sheft and Yost (2004) have shown that initially, the ability to discriminate noisy stimuli does increase with duration, but only up to a certain duration. This duration was around 25 ms in the case of Gaussian noise (Hanna, 1984). Beyond this duration, discrimination ability decreases, even though the amount of peripherally represented stimulus details becomes larger. The access to details in the peripheral spectro-temporal excitation pattern is apparently impaired for longer stimuli, which may be related to limitations in more central processes. The nature of this impairment is not well understood at the moment.

A performance impairment for longer stimuli contrasts with observations from signal detection experiments. The temporal and spectral integration of stimulus energy that occurs in a detection task was investigated by van den Brink and Houtgast (1990) using Gaussian tone-burst targets in the presence of a continuous noise masker. They found that this integration of target energy is less efficient than pure energetic integration, specifically for broadband signals. Nevertheless, for all bandwidths, van den Brink and Houtgast (1990) found a temporal integration effect; i.e., for a stimulus with a fixed level, detectability increased with increasing duration. Apparently detection of a known target stimulus and discrimination between independent noise tokens are governed by different integration processes, leading to a different dependence on stimulus duration. For a discussion of information in relation to perception, we refer to chapter 1.4.

The present study describes a number of discrimination experiments in order to better characterize the nature of perceptual information processing. In particular, this study investigates the decrease of discrimination ability for Gaussian noise with durations above approximately 25 ms.

The nonmonotonic duration dependence observed by Hanna (1984) suggests that an increase of number of degrees-of-freedom has a negative influence on discrimination performance for noise tokens longer than 25 ms. Such an interpretation would be in line with the idea that stimulus discrimination is based on cues that reflect more global stimulus properties such as mean spectral envelope. Therefore, we will investigate whether performance increases when listeners are instructed to listen to only a short part of a long-duration stimulus, thus ignoring the rest of the stimulus. In addition, a limitation in the ability to retain the increased amount of peripheral information may be a cause for the impairment in discrimination ability. We therefore studied whether listeners are able to better retain stimuli when they are presented more often.

Since stimulus duration and degrees-of-freedom are coupled in Gaussian noise, an additional experiment using a stimulus consisting of 5-ms Hanning-windowed tone-bursts randomly distributed over time investigated the role of stimulus duration versus number of degrees-of-freedom by decoupling the two factors. This last experiment thus also addressed whether the nonmonotonic discrimination performance can be related to auditory memory phenomena like memory decay as a function of elapsed time (see, e.g., Durlach and Braida, 1969).

2.2 Experiment 1: Temporal and spectral dependence

The first experiment is a replication of one of the experiments of Hanna (1984), to verify that listeners in the present study perform similarly. In addition, extra duration conditions were included to get a better indication of the stimulus duration at which discrimination performance is maximal.

2.2.1 Method

The experimental method was a same-different procedure where in each trial, two noise tokens were presented to the listener. These noise tokens were either identical or independent. For each trial, new noise samples were generated. Subjects were given feedback about the correctness of their answer after each trial.

Three male subjects participated, including the first (S1) and second authors (S2). The experiments were divided into sessions of maximally one hour. Each experimental condition (combination of stimulus frequency band and duration) was presented in 4 blocks of 50 trials (Subjects S1 and S2) or three blocks of 100 trials (subject S3). This excludes the training trials. Within a block, half of the trials were same trials and the other half were different trials. The trials within a block were presented in random order. The blocks were also presented in random order.

For each block of trials, the responses of the listeners were transformed into the sensitivity index, d', by calculating percentages correct for the *same* and the *different* trials. These percentages correct were converted to z-scores. Finally, d' was calculated by adding the z-scores of same and different presentations. It sometimes occurred that a subject gave correct answers for all same (or different) trials within a block, resulting in an infinite d' value. In this case an extra artificial incorrect same (or different) trial was added to the block, thus providing a non-infinite d' that could be used for calculating mean d' values and standard errors. For each subject mean d' values and standard errors of the mean were obtained by pooling all d' values of the measured blocks. Similarly, across-subject mean d' values and standard errors of the mean were obtained by pooling all the d' values of the measured blocks of all subjects.

At chance performance, the d' value equals zero. Above-chance performance results in positive d' values, e.g., 69% correct for both same and different trials results in a d' value of approximately 1 and 84% correct for both same and different trials results in a d' value of approximately 2.

Because we observed some training effects, the first 2500 trials for each subject were omitted. In the remaining data, for all subjects, the mean d''s of each repeated set of all conditions were within a margin of $\pm$ 0.2 d' relative to their mean d'.

2.2.2 Stimuli

The bandpass noise stimuli had five different frequency bands and nine durations. The −3-dB bandpass ranges were 100–3300, 100–600, 225–275, 2800–3300, and 2975–3025 Hz. The specified durations before filtering were 1.6, 6.4, 10.2, 16.1, 25.6, 40.6, 64.5, 102.4, and 409.6 ms. For the two narrowband conditions including 3000 Hz, a subset of these durations was used. The spectrum level was 40 dB and the stimuli were presented diotically.

Noise tokens were produced by digitally generating broadband noises of the specified duration and spectrum level with an inter-stimulus interval of 500 ms. The inter-stimulus interval was defined as the temporal separation between the offset of the first stimulus and the onset of the second stimulus within a trial. Subsequently, the tokens were filtered with a Chebyshev Type II digital filter with slopes of 100 dB/octave for the broadband and 500-Hz wide bands and approximately 200 dB/octave for the 50-Hz wide bands. The filters were designed using the Matlab filter design and analysis toolbox. Note that, in the study of Hanna (1984), the filtering was done with analog filters. The stimuli included the ringing of the filters in order to avoid audible truncation effects. The stimuli were presented from a PC through a high-quality soundcard (RME DIGI96/8 PAD) at 16 bit, 44.1-kHz sampling resolution using headphones (Beyerdynamic DT990Pro).

2.2.3 Results

Figure 2.1 shows mean d' values (ordinates) as a function of stimulus duration (abscissas). Curves are shown for five bandwidth and center frequency combinations (symbols) of individual subjects (upper and bottom-left panels) and the means across subjects (bottom-right panel). The error bars indicate plus and minus one standard error of the mean. The results are generally in agreement with the results of Hanna (1984). However, the average d' was about 0.5 d' units lower in our data.

The curves for the bands containing low-frequency energy (100–3300 Hz, 100–600 Hz, and 225–275 Hz) show discrimination performance that initially increases with increasing duration. We found a plateau of best performance in the range 16.1 to 102.4 ms. The precise location of the plateau depended on subject and spectral range of the noise. For durations above this maximum, discrimination performance decreased with increasing duration.

Discrimination ability for high-frequency conditions (spectral center at 3 kHz, diamonds and downward triangles) was overall poorer than for low-frequency conditions (spectral center at 250 Hz, squares and triangles) and did not show so much evidence for a mid-duration peak. This is not completely in agreement with the data of Hanna (1984), who found that the conditions with low-frequency 50-Hz wide bands and high-frequency 50-Hz wide bands gave very similar results.

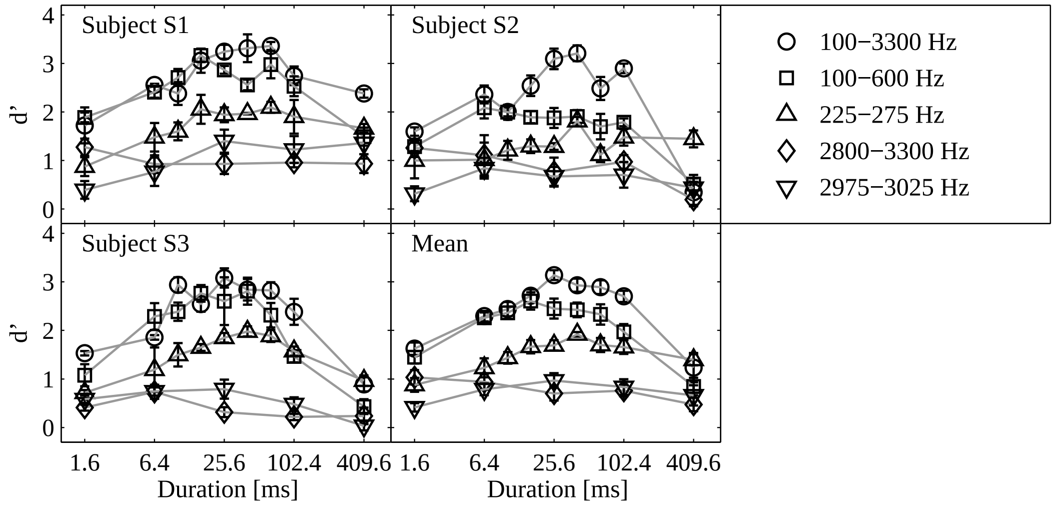

Figure 2.1: Mean d' values as a function of stimulus duration for noise with passbands of 100–3300 Hz (circles), 100–600 Hz (squares), 225–275 Hz (triangles), 2800–3300 Hz (diamonds), and 2975–3025 Hz (downward triangles) of individual subjects (upper and bottom-left panels) and across subjects (bottom-right panel). The error bars indicate plus and minus one standard error of the mean.

In general, for each duration (with the exception of 409.6 ms) an increase in the number of critical bands covered by the noise resulted in higher discrimination performance. The highest performance occurred for the noise with the greatest bandwidth (100–3300 Hz, circles).

For a duration of 409.6 ms, the mean results did not show the highest performance for broadband stimuli. In fact, subject S2 showed a higher performance for the low frequency 50-Hz wide band. The fact that, at a duration of 409.6 ms performance was better for 50-Hz wide bands than for 3200-Hz wide bands, was also the case for two of the experiments in the study of Hanna (1984). Hence, there appears to be some evidence that performance is worse for the 100–3300-Hz bands than for the 225–275 Hz bands at a duration of 409.6 ms, or at least that performance is not higher.

2.3 Experiment 2: Inter-onset interval dependence

The study of Hanna (1984) and the current study showed that the ability to discriminate broadband noises decreases for durations above 40 ms. A possible explanation for this performance decrease might lie in the increasing temporal separation between corresponding features, e.g., the onsets, in the two noise bursts within a trial. This is a direct consequence of using a fixed offset-onset interval of 500 ms. Arguably, the degradation of discrimination ability for durations above 40 ms shown in the Exp. 1 may be due to this larger temporal separation of the stimulus features. In the next experiment, the temporal separation of the stimulus features was varied while keeping the stimulus duration fixed at 40.6 ms. The results are compared with results of Exp. 1, in order to investigate if the degradation of discrimination ability can be accounted for by the temporal separation.

2.3.1 Method

The experimental method was identical to the method of Exp. 1. Five subjects, including those from Exp. 1, participated in this experiment. All subjects performed each condition in four randomized blocks of 100 trials, of which 50 were same trials and 50 were different trials. The blocks were presented in randomized order.

Because subjects S4 and S5 did not participate in Exp. 1, their missing data for variable IOI conditions from Exp. 1 were obtained in a separate session. The conditions in this separate session were presented in four blocks of 100 trials.

2.3.2 Stimuli

The 40.6-ms, 100–3300-Hz stimulus, which resulted in high discrimination performance in Exp. 1, was used but was presented with varying inter-onset intervals. The spectrum level was again 40 dB SPL. The inter-onset interval (IOI) was defined as the temporal separation between the onset of the first stimulus and the onset of the second stimulus within a trial. In Exp. 1 the pause between the two bursts was fixed at 500 ms while varying the duration of the noise bursts. Therefore, the IOIs in each condition of Exp. 1 were 500 ms plus the duration of the stimulus. In the current experiment the pause between the two bursts was varied while keeping the stimulus duration fixed at 40.6 ms. In doing so, the IOIs in the current experiment can be chosen to be identical to the IOIs of Exp. 1, but with a fixed stimulus duration of 40.6 ms. The conditions in Exp. 2 used IOIs of 540.6, 564.5, 602.4, and 909.6 ms, which are equivalent to the IOIs of the 40.6-, 64.5-, 102.4-, and 409.6-ms duration conditions of Exp. 1.

2.3.3 Results

Figure 2.2 shows the results of Exp. 2. (x symbols) and the data for the 100–3300-Hz band from Exp. 1 (circles), plotted as function of their IOI. The 540.6-ms IOI condition is the only condition in which both stimulus duration and IOI were the same in the two experiments. For all other IOIs, the overal stimulus durations are different between the curves. For IOIs above 540.6 ms, the conditions of the current experiment show consistently higher d' values than the conditions with varying durations. Although performance decreases slightly with increased IOI, the intrinsic larger temporal distance between corresponding features in Exp. 1 appears not to be a sufficient explanation for the degradation of discrimination ability for stimuli with durations larger than 40.6 ms.

2.4 Experiment 3: Gaussian-noise discrimination with selective listening

It is remarkable that listeners are unable to perform the discrimination task better for 400-ms stimuli than for 40-ms stimuli, even though there are more degrees-of-freedom in the longer stimulus. If the decrease of performance cannot be explained by the larger temporal distance between the features, as the previous experiment showed, a surplus of peripheral information might be impairing performance on the discrimination task. Possibly listeners use a suboptimal strategy by trying to retain peripheral information of the complete stimulus. If this were the case, they might be able to improve their performance by focusing on a smaller part of the stimulus when there is too much peripheral information resulting from the entire stimulus.

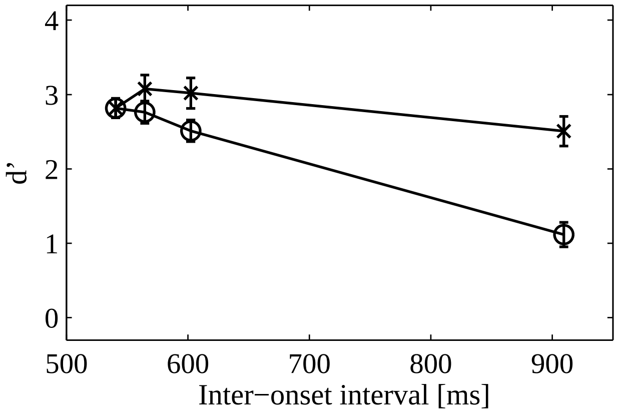

Figure 2.2: Across subjects mean d' values as a function of the Inter-Onset Interval (IOI). The *x symbols* show data for the 100–3300-Hz band with a fixed duration of 40.6 ms and IOIs of 540.6, 564.5, 602.4, and 909.6 from Exp. 2. The *circles* show data for the 100–3300-Hz band with durations 40.6, 64.5, 102.4, and 409.6 ms from Exp. 1. The IOIs for these conditions are equivalent to the IOIs for the conditions from Exp. 2. The error bars indicate plus and minus one standard error of the mean, based on the results of five subjects.

A relevant experiment in this context is that of Coble and Robinson (1992), using noise bursts that were identical on same trials. On different trials, the bursts were identical except for τ ms where the bursts were independent. They showed that discrimination performance for such partially independent Gaussian noise was dependent on the temporal location of the independent noise parts. Noise discrimination performance was better when the independent part (i.e., the part that supported discrimination) was located at the end of the stimulus than when it was located at the beginning of the stimulus. However, performance for partially independent noise was always lower than for fully independent Gaussian-noise tokens.

In the next experiment, it was investigated whether subjects were able to improve their discrimination performance by concentrating on a part of the stimulus. In contrast to Coble and Robinson (1992) the presented stimuli were either fully identical or fully independent. Thus, the stimuli were the same as in the first noise discrimination experiment. Only the instructions to the subjects were different. They were asked to listen only to either the beginning or the ending of the stimulus, which essentially required them to ignore some of the available peripheral information resulting from the stimulus.

2.4.1 Method

The experimental method was identical to the method of Exp. 1, except that the subjects were explicitly instructed to focus on the beginning (d'_{begin}, begin-focus) in one set of blocks, and to focus on the end of the stimulus (d'_{end}, end-focus) in another set. Six male subjects, including those of Exp. 1, participated in this experiment. The two conditions were each assessed in four randomized blocks of 100 trials, of which 50 were same trials and 50 were different trials. The blocks were presented in alternating order.

The results of this experiment were compared to some of the results of Exp. 1. However, subjects S4, S5, and S6 did not participate in Exp. 1. The results for the 40.6- and 409.6-ms stimuli with a bandpass range of 100–3300 Hz for subjects S4, and S5 were taken from Exp. 2. The results for the 40.6- and 409.6-ms stimuli with a bandpass range of 100–3300 Hz for subject S6 were obtained in a separate session, where they were presented in four blocks of 100 trials.

2.4.2 Stimuli

The bandpass Gaussian-noise stimuli had -3-dB cutoffs at 100 and 3300 Hz, a duration of 409.6 ms, and a spectrum level of 40 dB. As before, in same trials the two noise tokens were identical and in different trials the two noise tokens were independent. The stimuli were the same as in Exp. 1. There was no special begin- or end-section, nor was there any (visual or acoustic) indication of stimulus sections.

2.4.3 Results

The columns of Table 2.1 show the individual and across subject mean d' values and standard errors of the focus experiment. The first two rows show the results for the 40.6-ms and 409.6-ms conditions, with a frequency band of 100–3300 Hz, from Exp. 1. The other two rows show the results for the 409.6-ms conditions, with a frequency band of 100–3300 Hz, where listeners focused on the beginning or on the end of the stimuli.

The lowest performance was found for the original 409.6-ms duration condition and the begin-focus condition, both resulting in a mean d' of 1.3. The d' for the end-focus condition was 0.4 higher. The highest d' of 2.8 was achieved for the original 40.6-ms condition. An ANOVA combined with post-hoc Tukey HSD multiple comparisons

Table 2.1: Results of the selective listening experiment, where listeners were asked to focus on the beginning or the end of the stimuli. Data are expressed as mean and standard error of the mean (between brackets) d' values of subjects S1 to S6, as well as across-subjects means. Gaussian-noise stimulus −3-dB cutoffs were at 100 and 3300 Hz.

	S1	S2	S3	S4	S5	S6	Mean
40.6 ms[1]	3.3 (.3)	3.2 (.2)	2.8 (.2)	2.3 (.1)	2.4 (.3)	2.7 (.3)	2.8 (.1)
409.6 ms[1]	2.4 (.1)	0.3 (.3)	0.9 (.2)	0.9 (.1)	0.8 (.2)	2.1 (.1)	1.3 (.2)
409.6 ms begin	2.2 (.4)	0.3 (.2)	1.0 (.2)	0.8 (.1)	1.6 (.3)	1.7 (.2)	1.3 (.2)
409.6 ms end	2.6 (.4)	1.5 (.3)	1.2 (.3)	1.2 (.1)	1.5 (.2)	2.0 (.3)	1.7 (.1)

[1] From experiment 1

revealed significant differences ($F_{5,86} = 15.9$, $p < 0.05$) between all conditions, except between the original 409.6-ms condition from Exp. 1 and the condition in which listeners focused on the beginning of the stimulus. Listeners were treated as random effects, thus; differences in baseline performances for the individual listeners were taken into account.

When asked to focus their attention deliberately on the end of the stimulus, most listeners performed the discrimination task better than the original discrimination experiment. Apparently more perceptual information can be retrieved from the stimulus than is typically done by the listeners. When asked for their introspection, listeners reported that the focus experiment was more difficult to perform than the normal discrimination experiment. This could be a reason why most listeners were not inclined to develop this listening strategy independently. Only S1 and S6 may have developed a similar strategy already in the first experiment, as indicated by the small performance improvement when asked to focus. Interestingly, these two subjects also show the highest performance in the original 409.6-ms condition.

Although, on average, the listeners were able to achieve better performance by concentrating only at the end of the stimulus, the effect was relatively small and the overall level of performance was still significantly lower than for the 40.6-ms stimuli of Exp. 1. This indicates that listeners could not use the peripheral information that was conveyed by the end of the stimulus in the same way as if it had been presented in isolation in a short stimulus. Focusing on the stimulus beginning did not lead to a significant change in performance relative to the original discrimination experiment.

It is interesting to compare the results of the current experiment with those of Coble and Robinson (1992). They measured discrimination performance for noise tokens (with durations in the range from 25 to 150 ms) where only the beginning, middle, or end of the two noise tokens was independent in a *different* trial, the rest of the noise tokens was identical. In a *same* trial the two noise tokens were completely identical. Performance was highest when the stimuli differed at the end. Overall, performance in these conditions was poorer than when the entire stimulus differed. Our experiment showed, however, that performance improved slightly when listeners concentrated on the end of the stimulus. The essential difference between our experiments and those of Coble and Robinson (1992) is that our stimuli were independent across the entire duration of a *different* trial, while for Coble and Robinson (1992) only a part of the stimulus was independent. Apparently the presence of differences in the unattended part of the stimulus in our experiment influenced discrimination performance when attending to the end part of the stimulus. This finding supports our suggestion that listeners cannot selectively process only a part of the stimulus, but are always influenced by the peripheral information resulting from the entire stimulus.

2.5 Experiment 4: Frozen-Gaussian-noise discrimination

The selective listening experiment showed that subjects' performance could be somewhat increased when they were instructed to focus their attention on only part of a stimulus. Evidently, the auditory system can retrieve perceptual information more effectively from the stimulus when adopting a better listening strategy. Possibly this improvement in performance is related to a limitation in retaining the peripheral information that is present in a long duration stimulus (e.g., Cowan, 2001). By instructing the listeners to concentrate on only a part of the stimulus this limitation was partly avoided and performance increased.

In order to investigate the role of limitations in the capacity to retain perceptual information, the next experiment presented the same reference stimulus repeatedly in one block to determine whether listeners are able to form and maintain a consistent internal representation of the reference stimulus. If so, their discrimination performance should increase relative to their performance in the running noise experiment.

2.5.1 Method

The experimental method was identical to that of Exp. 1. Five male subjects participated, including the subjects from Exp. 1. The conditions were assessed in four randomized blocks of 100 trials of which 50 were same trials and 50 were different trials. The blocks were presented in random order.

2.5.2 Stimuli

The bandpass Gaussian-noise stimuli had −3-dB cutoffs at 100 and 600 Hz or at 225 and 275 Hz. These bandwidths were selected because listeners were well able to perform the task in the first experiment and their performance was not so high that it would immediately saturate at perfect performance (approximately a d' value of 4). The specified durations before filtering were 1.6, 6.4, 25.6, 102.4, 409.6, and 1638.4 ms. In the previous experiments we replicated the experiments of Hanna (1984) as closely as possible. For the frozen-noise experiment we chose to increase the spectrum level from 40 dB to 60 dB because mainly the 50 Hz wide stimuli were not so loud .

The stimuli were presented in several repetition configurations that differed in the degree to which they were reused across trails. (1) Running noise: In every trial, new noise tokens were generated. These conditions were a replication of some of the conditions from Exp. 1, but with a spectrum level of 60 dB instead of 40 dB; (2) Semi-frozen: The first token of each trial within a 100 trial block was always the same (frozen), while the second token was either identical to the first one or a newly generated token. For each block of 100 trials a new frozen-noise token was generated; (3) Frozen: The first token of each trial within a block of 100 trials was always the same, while the second token was either identical to the first one or a different frozen token which remained the same for the entire block of 100 trials. Thus, effectively, only two different noise tokens were used in a block of 100 trials. For each block of 100 trials, two new frozen-noise tokens were generated. All subjects were presented with the same frozen-noise tokens.

2.5.3 Results

Figures 2.3 and 2.4 show mean d' values (ordinates) as a function of stimulus duration (abscissas). Individual (panels one to five) and across subjects (bottom-right panel) data are shown for 500-Hz wide (Fig. 2.3) and 50-Hz wide (Fig. 2.4) Gaussian-noise bands. The symbols indicate the repetition configurations (running, semi-frozen, and frozen). The error bars indicate plus and minus one standard error of the mean.

The results show that, for both bandwidths, the ability to discriminate Gaussian noise is directly related to the repetition configurations of the stimuli. Discrimination ability is highest for frozen noise (triangles), lowest for running noise (squares), and intermediate for semi-frozen noise (diamonds). It can be concluded that the use of repeated stimuli improved performance in the discrimination task.

Again, as in Exp. 1, there was a duration at which discrimination ability was maximal. This duration was approximately 25.6 ms for both bandwidths and did not depend on the repetition configurations. This is consistent with the assumption that the duration at which the maximum occurs is dependent on the number of degrees-of-freedom per critical band in the stimulus and that the d' for this maximum can be influenced by letting the listener obtain a more accurate representation of the stimulus. Note that the maximum performance was not observed at 40.6 ms because this duration was not used in this experiment. It seems that, in combination with the results of Exp. 1, optimal discrimination performance for Gaussian-noise stimuli occurs for durations in the range 16.1 to 40.6 ms.

Interestingly, for four out of five listeners, the semi-frozen noise tokens with a frequency band of 225–275 Hz and duration of 6.4 ms led to a lower d' than the semi-frozen noise tokens with a duration of 1.6 ms in the same frequency band. Possibly, the presented frozen tokens, which were the same for all the subjects, were by coincidence more difficult to discriminate than the average of the population of possible noise tokens because, for the first stimulus, only four frozen-noise tokens were used in the semi-frozen conditions.

The discrimination maximum in Exp. 1 and 4 seems to be related to stimulus duration, with about 40 ms being the duration resulting in maximum performance. However, given that stimulus information (number of degrees-of-freedom cf. section 1.4) increases with duration, the duration at which maximum performance occurs may not be related directly to stimulus duration, but merely an effect of the correlation between the amount of stimulus information and stimulus duration. In the next experiment we wanted to obtain more insight into how the number of degrees-of-freedom of the stimulus and its duration influence discrimination performance when they were varied independently.

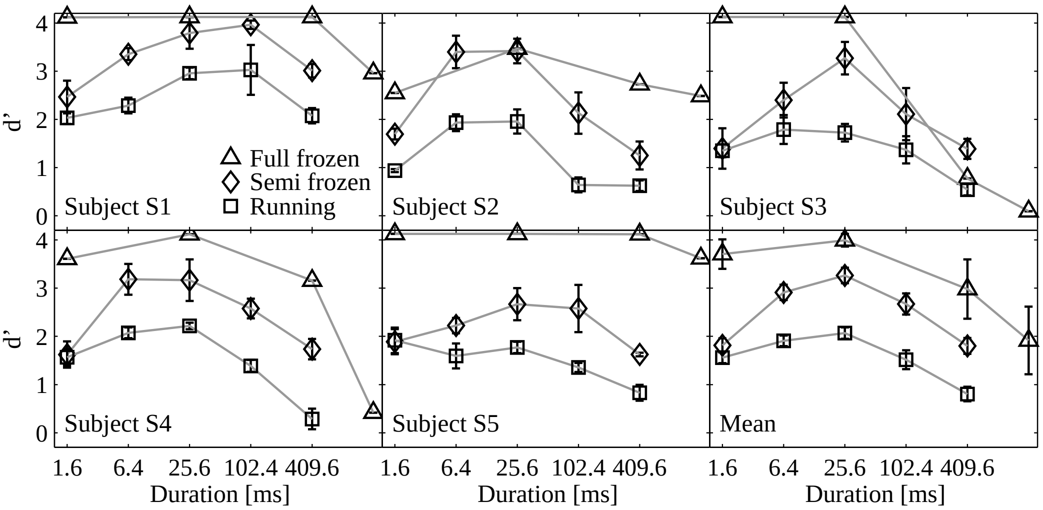

Figure 2.3: Discrimination performance for 100–600-Hz Gaussian-noise bands with three repetition configurations: Running noise (squares), semi-frozen noise (diamonds), and frozen noise (triangles). Mean data across subjects are shown in the bottom-right panel, and data for individual subjects are shown in the other panels. The error bars indicate plus and minus one standard error of the mean.

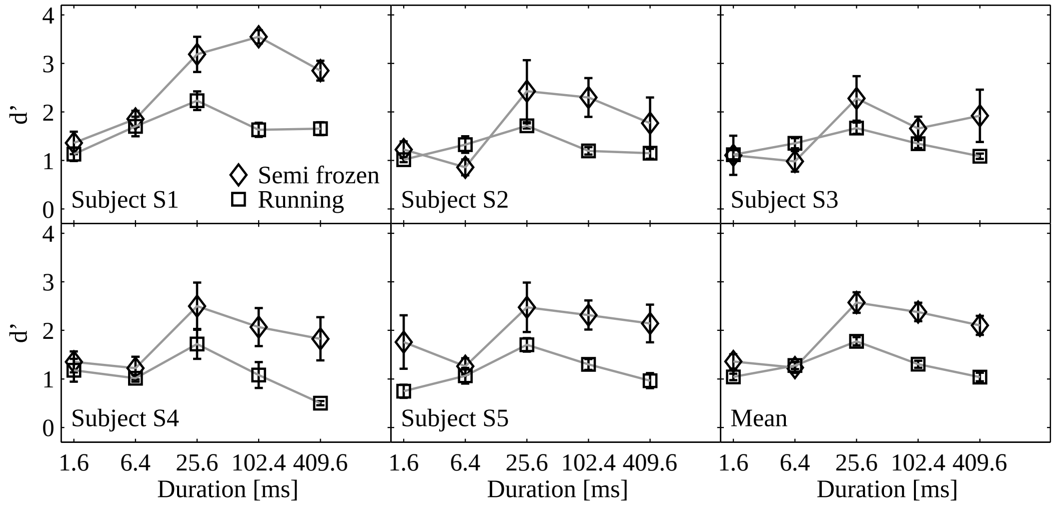

Figure 2.4: Discrimination performance for 225–275-Hz Gaussian-noise bands with two repetition configurations: Running noise (squares), and semi-frozen noise (diamonds). Mean data across subjects are shown in the bottom-right panel, and data for individual subjects are shown in the other panels. The error bars indicate plus and minus one standard error of the mean.

2.6 Experiment 5: Discrimination of random tone-burst complexes

The next experiment used a stimulus comprised of a number of 5-ms Hanning-windowed tone bursts randomly distributed over time, for which, in one set of conditions, the bursts were presented at the same frequency. In another set of conditions, seven burst frequencies were used. This stimulus will be referred to as a random tone-burst complex. The number of tone bursts and the number of tone-burst frequencies in these stimuli were varied in order to investigate the influence of number of degrees-of-freedom on discrimination ability.

Note that whereas for a Gaussian-noise token the number of degrees-of-freedom is proportional to the product of bandwidth and duration (Hartley, 1928; Nyquist, 1928), for the tone-burst complex, the number of degrees-of-freedom is proportional to the number of tone bursts and is decoupled from duration.

2.6.1 Method

The experimental method was a same/different experiment, identical to the method of Exp. 1. Five male subjects participated in this experiment including the subjects from Exp. 1. For each subject, the conditions were assessed in four randomized blocks of 100 trials of which 50 were same trials and 50 were different trials. All blocks were presented in random order.

2.6.2 Stimuli

Two types of stimuli were used: Random tone-burst complexes with tone bursts of only one frequency and random tone-burst complexes with seven frequencies. The stimulus generation is sketched in Fig. 2.5. For each frequency, tone bursts (s) were produced by multiplying a sinusoidal carrier (c) with an envelope (m). The envelope was comprised of a number of Hanning windows, each with a total duration of 5 ms. The starting points of the Hanning windows were randomly distributed within the full duration of the stimulus, which was either 51.2 ms or 409.6 ms. It was ensured that the tone bursts fell entirely within these stimulus durations. The random tone-burst complexes with seven frequencies were generated by adding seven independent tone-burst realizations, each with a different carrier frequency.

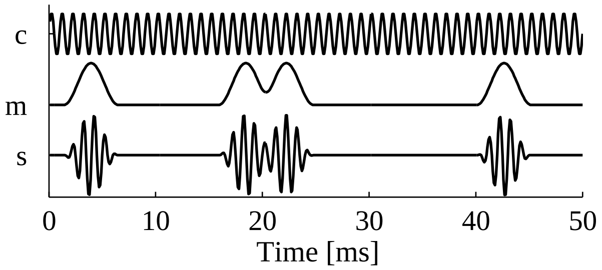

Figure 2.5: A random tone-burst signal (s) is generated by multiplying a carrier (c) with a modulation envelope (m) that consists of a number of Hanning windows additively placed at random temporal positions within the duration of the stimulus.

The peak level of the tone burst envelopes was 70 dB SPL. In the one-frequency conditions the tone bursts had a (nominal) frequency of 607 Hz (ERB$_N$ number of 12, Glasberg and Moore, 1990). In the seven-frequencies conditions the tone bursts had frequencies of 208, 314, 444, 607, 808, 1057, and 1367 Hz (ERB$_N$ numbers of 6 up to and including 18 with a spacing of 2). In both sets of conditions, the number of tone bursts was 2, 4, 8, 16, 32, 64, 128, and 256 tone bursts *per frequency*, with the exception that, for the 51.2-ms duration, the 128- and 256-tone-bursts per frequency conditions were not used. This means that, when tone bursts were distributed over seven frequencies, the total number of bursts was 14, 28, 56, 112, 224, 448, 896, and 1792.

We chose a relatively short tone-burst duration in order to limit the amount of temporal overlap. Especially for the low frequency tones, this creates some spectral overlap within the auditory filters that are centered around the different burst frequencies.

2.6.3 Results

Figure 2.6 shows the mean d' values (ordinates) for the random tone-burst complexes as a function of the *total number* of tone bursts (abscissas). This total number is the number of spectral components in the stimulus multiplied by the number of tone bursts per frequency. Data are shown for durations of 51.2 ms (dashed lines) and 409.6 ms (solid lines) with tone bursts of either one frequency (circles) or seven frequencies (x symbols). The upper panels and the bottom left panels show the individual d' means and the bottom right panel shows the mean d' across subjects. The error bars indicate plus and minus one standard error of the mean.

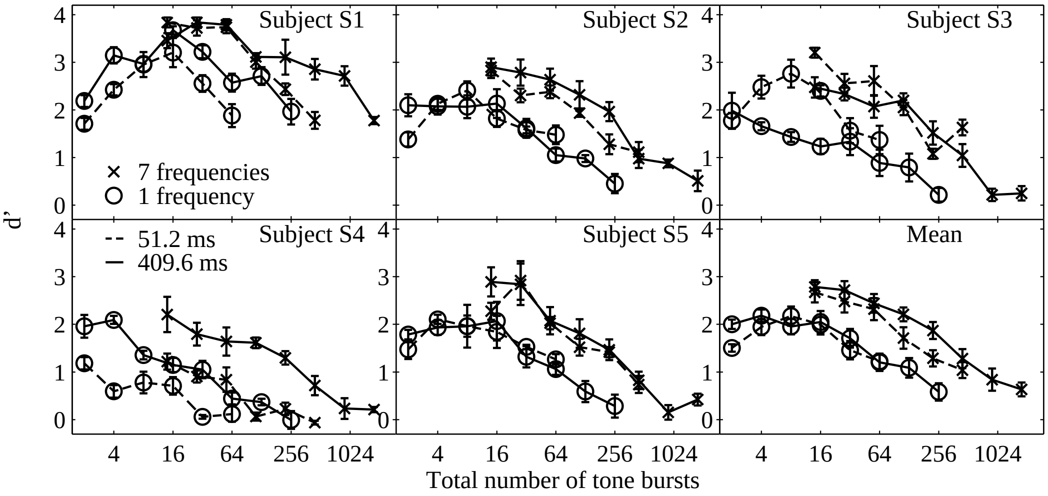

Figure 2.6: Mean d' values as a function of total number of tone bursts for random tone-burst complexes with one frequency (circles) or seven frequencies (x symbols). Stimulus durations were 51.2 ms (dashed lines) or 409.6 ms (solid lines). Means across subjects are shown in the bottom-right panel and individual results in the other panels. The error bars indicate plus and minus one standard error of the mean.

In the mean results it can be observed that, in general, discrimination performance for each frequency and duration combination decreased with increasing number of tone bursts. However, for the short-duration conditions, with only one frequency component (circles, dashed lines), discrimination performance first increased with increasing number of tone bursts. A comparison of the conditions with circle symbols and with cross symbols for the same total number of tone bursts reveals that discrimination performance was overall higher when the tone bursts were spread over seven frequencies than when they were concentrated at one frequency.

When comparing the solid lines with the dashed lines in Fig. 2.6, for equal number of frequencies (circles or x symbols), the data for short-duration stimuli showed large overlap with the data for long-duration stimuli indicating that there is little influence of duration. In the individual data, there were a few deviations from

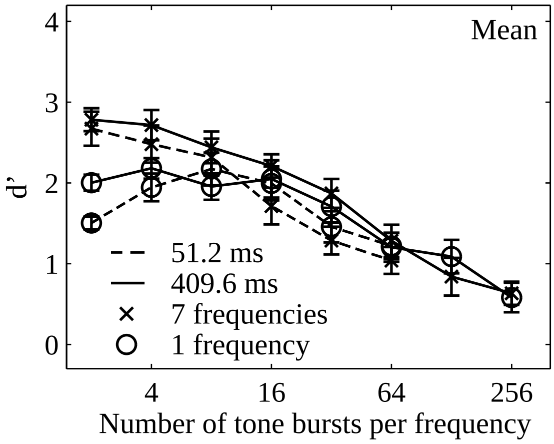

Figure 2.7: Mean d' values as a function of number of tone bursts *per frequency* for random tone-burst complexes with one frequency (circles) or seven frequencies (x symbols). Stimulus duration were 51.2 ms (dashed lines) or 409.6 ms (solid lines).

this observation. Specifically, the short-duration stimuli with 4 up to and including 16 tone bursts for subject S3 led to higher performance than for the long-duration stimuli. Furthermore, subject S4 showed generally higher performance for the longer-duration stimuli than for the short-duration stimuli. However, an analysis of variance did not reveal a significant influence of duration on discrimination ability for random tone-burst complexes. Significant effects were found for the number of tone bursts ($F_{1,2} = 94.2$, $p < 0.011$) and the number of frequencies ($F_{1,2} = 30.8$, $p < 0.031$).

These experiments show that an increase in the number of degrees-of-freedom (i.e., number of tone bursts) leads to a reduction in discrimination performance, resembling what was seen in Exp. 1 for durations in excess of 40 ms. When comparing conditions with equal bandwidth and number of tone bursts, but different durations, we see that on average discrimination performance is very similar. Apparently, discrimination performance depends primarily on the number of degrees-of-freedom and not so much on the duration per se. A similar result was found in a study of Watson *et al.* (1990, pg. 2638), in which "[...] it was shown that listeners' ability to detect *spectral-temporal changes* in randomly generated patterns of multiple non-overlapping tones is primarily a function of the number of tones in the pattern with very little effect of component duration or of total pattern duration." However, in that study, only one or two tones of each pattern were altered and the patterns were serial, so none of the tones overlapped in time. Therefore, not all tones contributed to discriminability in that study. Rather, the varied number of tones in the pattern affected the relative duration of the informative tones with respect to the total duration of the pattern. This relative duration governed performance in a large variety of tonal pattern paradigms (Kidd and Watson, 1992).

When comparing conditions with different bandwidths in Fig. 2.6, i.e., complexes with one (circles) or seven frequency components (x symbols), we see that distributing a given amount of tone bursts across more frequencies leads to better performance than presenting them all at one frequency. Figure 2.7, where the average results are plotted as a function of the number of tone bursts *per frequency*, shows that the curves overlap much more than in Fig. 2.6. By calculating the coefficient of determination[1] (R^2) it was found that the 68% of the variance of the data in Fig. 2.7 could be explained by the number of degrees-of-freedom per critical band (only using the data for 2 up to and including 64 tone bursts per frequency because these were available for all curves). The proportion of variance explained by the duration was 5% and for number of frequency components this was 8%. This shows that adding more degrees-of-freedom in the spectral dimension did not have a large effect on discrimination. The largest difference between the one- and seven-frequencies curves is seen for the 2 and 4 bursts per frequency conditions. For these conditions, there seems to be some spectral integration of peripheral information.

Interestingly, for short-duration stimuli with one frequency only, there is an initial improvement of performance for small numbers of tone bursts (2–4). This initial improvement is not seen for any of the other conditions. The relatively low performance for the short duration stimuli with one frequency component and two tone bursts may be related to difficulties in perceiving absolute timing of the tone bursts within the nominal stimulus interval. When only two tone bursts are present within both stimuli of a *different* trial, it can happen that nearly the same time intervals occur between the two tone bursts within these stimuli, but at a different timing offset relative to the nominal start time of the stimulus. In that case, discrimination

[1] Coefficient of determination was calculated with $R^2 = 1 - \sum_i (o_i - m_i)^2 / \sum_i (o_i - \overline{o})^2$, where o_i and m_i are the individual observed and model values respectively, and $\overline{o}$ is the mean over all observations. As model (m_i), the average of all observations for each value of the variable of interest (i) was taken.

purely depends on hearing the absolute timing of the tone-burst *pairs*. For short stimuli with a proportionally larger inter-stimulus interval, it may be substantially more difficult to hear this difference in timing because of a smaller relative change in timing. For the longer stimulus, absolute timing of bursts can differ more across intervals, thus providing a potentially stronger cue. For stimuli with more tone-burst frequencies, the tone bursts in other frequencies can serve as a reference to compare tone-burst timing.

When the number of bursts is relatively small, the individual bursts are resolved and the most important cue that listeners can use to distinguish different stimuli is given by the timing intervals within a stimulus. However, as the number of degrees-of-freedom is increased, other cues like temporal envelope modulation and frequency modulation (in the case of seven frequencies) may start to play a role. This is especially so when the number of bursts becomes so large that individual bursts start to overlap. This suggests that the cue a listener uses may be a function of the number of degrees-of-freedom in the stimulus.

2.7 General discussion

The first experiment of the current study reproduced the finding of Hanna (1984) that, after an initial increase of discrimination performance with increasing duration up to 40 ms, the ability to discriminate *decreases* for Gaussian-noise tokens with longer durations. As indicated by Exp. 2, this decrease could not be understood solely by the larger intrinsic temporal distance that is present between the corresponding features of two noise tokens.

Although there are more degrees-of-freedom in, e.g., a 400-ms stimulus than in a 40-ms stimulus, the ability to discriminate was higher for the shorter stimulus. This suggests that listeners are better able to retain and compare peripheral information for shorter stimuli. Such an observation is not in good agreement with temporal integration and multiple look models (e.g. Viemeister and Wakefield, 1991; Dau *et al.*, 1996), because such models predict that discrimination ability should increase with available peripheral information in the internal representation. Even if the maximal duration for the accumulation (temporal integration or multiple looks) was restricted to an interval shorter than 400 ms, one would expect that discrimination ability for 400-ms noise tokens should be at least as high as for 40-ms stimuli, but not that it was lower.

Several descriptive models have been proposed for the discrimination of Gaussian noise (Fallon, 1989; Rickert, 1998). However, these models were not aimed at explaining effects of the total stimulus duration and operated at either a fixed total stimulus duration, or over a duration range for which there was not a large effect of total duration. These studies were concerned with investigating the effect of temporal location of pieces of uncorrelated noise within a stimulus. One of the major findings was that the uncorrelated parts were more easily detected when they were placed towards the end of the stimulus.

Our experiments bear some resemblance with informational masking studies, which show that the detection of a target stimulus depends on the amount of uncertainty (information) in the masker stimulus (e.g. Watson, 1987; Durlach *et al.*, 2003). In our study there were no maskers and the complete stimulus was the target. Nevertheless, we see the same tendency that when the amount of peripheral information exceeds a certain threshold, discrimination performance decreases.

The duration effect of Exp. 1 suggests a limited access to memory for subparts of the stimulus. This is also illustrated by the third experiment, in which listeners were instructed to attend only to the beginning or to the end of a 409.6-ms stimulus. Compared to the condition where no instruction was given to focus on a part of the stimulus (Exp. 1), performance was the same when focusing on the beginning of the stimuli, and improved only slightly when focusing on the end. However, performance always remained significantly worse than for 40.6-ms duration stimuli. If listeners were able to process an arbitrary 40.6-ms part of the 409.6-ms duration stimulus independently, we would have expected similar discrimination performance as for the 40.6-ms duration stimulus. This limited access to subparts of the stimulus is in line with the idea that, within certain durations, noise bursts are stored in memory as a discrete entity as proposed by Näätänen and Winkler (1999).

In the fourth experiment, when the first noise token in each trial was frozen, discrimination performance improved. This result suggests that the repeated presentation of the “same” stimulus enabled listeners to build up a more accurate internal reference representation. Frozen noise is often used, for example, in detection experiments to investigate the relative contributions of internal and external variability (e.g., Buus, 1990). In our experiments, external variability plays a fundamentally different role than in detection experiments. Instead of being a limiting factor for performance, it is the factor that *enables* discrimination. In terms of stimulus variability in itself, one would, on average, not expect a difference in performance between random and frozen noise. The average distance in the internal representations between pairs of frozen noise is the same as between pairs of running noise. The major difference between these two types of experiments is that, for frozen noise, subjects can build up templates of the internal representations of the two noises, and also of their difference. In terms of an optimal discrimination process, having a template allows for weighting differences between the stimuli such that they are emphasized at locations where they are expected according to the template. In this way, the influence of the internal noise can be reduced.

The repeated presentation of the same noise links this study to the study of Kaernbach (1993). He investigated the perception of repeated noise, i.e., a continuous noise made by repeating a single piece of noise with a duration of, e.g., 500 ms. For such repeated noise, details of the noisy structure were perceived that would not have been perceived in a non-repeated presentation. Such details were referred to as "clanks" and "rasping", similar to what listeners reported informally in the current study after doing the experiment with frozen noise.

For Gaussian noise, duration and number of degrees-of-freedom are inseparably coupled. Therefore, in Exp. 5, a stimulus was used consisting of a specified number of tone bursts that were randomly placed within a time frame of either 51.2 ms or 409.6 ms. In this type of stimulus, the random tone-burst complex, the duration and the number of degrees-of-freedom are decoupled, allowing their effects to be studied separately. The results showed that there was not a large effect of duration on the ability to discriminate, but there was a large influence of number of degrees-of-freedom in the stimulus. In fact, the number of tone bursts per auditory filter appeared to account for the majority of the trends in the results.

An interpretation of the above findings, that fits the framework provided by Cowan (2005), is that listeners retrieve or remember the stimuli as stand-alone auditory objects from sensory memory, and that there is a fixed and limited amount of resources that can be allocated to retain and process the internal representations of these auditory objects. In working memory, Cowan attributed this limitation to the focus of attention. The limitation has also been attributed to mechanisms of storage and of processing (cf. Halford *et al.*, 1998).

It is informative to compare the number of degrees-of-freedom of the stimuli from

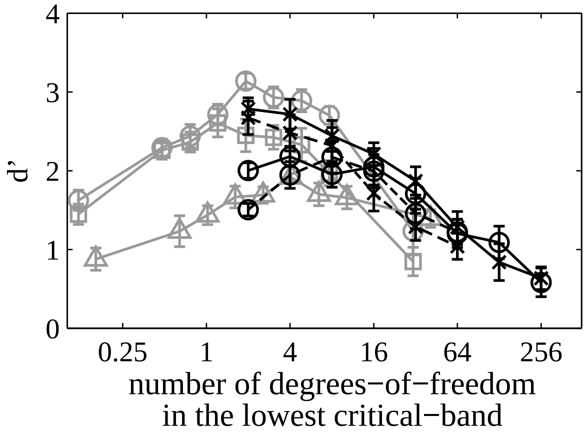

Figure 2.8: Mean discrimination performance expressed as d' values as a function of the number of degrees-of-freedom in the lowest critical band that is excited by the stimulus for Gaussian-noise tokens (from Fig. 2.1) and random one-burst complexes (from Fig. 2.7). The grey symbols indicate results for the Gaussian-noise tokens with frequency ranges of 100–3300 Hz (circles), 100–600 Hz (squares), and 225–275 Hz (triangles). The black symbols indicate results for the random tone-burst complexes with one frequency (circles) or with seven frequencies (x symbols) and a duration of 51.2 ms (dashed lines) or 409.6 ms (solid lines)

Exp. 1 to those from Exp. 5. In Fig. 2.8 we have plotted the d' values from Fig. 2.1 (in grey) and from Fig. 2.7 (in black) as a function of the number of degrees-of-freedom in the lowest critical-band that is excited by the stimulus. For the Gaussian-noise stimuli from Exp. 1, the number of degrees-of-freedom is calculated by taking two times the product of the duration and the width of the lowest excited critical band (cf. Hartley, 1928). The motivation for plotting results as a function of the number of degrees-of-freedom in the *lowest* critical band was that it had the smallest bandwidth and therefore the lowest number of degrees-of-freedom. If a large number of degrees-of-freedom is limiting performance, analyzing this auditory filter should thus lead to the best performance for long-duration stimuli. Note that the duration used in the calculation of the number of degrees-of-freedom was the duration prior to filtering, which results in a number of degrees-of-freedom smaller than one. The actual stimulus durations after filtering were longer. The bandwidth of the lowest excited auditory filter was determined using the ERB_N scale. If this width exceeded the stimulus bandwidth, then the stimulus bandwidth was used. For the random tone-burst complexes from Exp. 5, the number of degrees-of-freedom in the lowest critical band was simply the number of tone bursts per frequency.

Figure 2.8 shows that the Gaussian noise and the random tone-burst complex stimuli had a comparable range of the number of degrees-of-freedom, which enables us to compare the conditions of Exp. 1 and Exp. 5. It appears that, for both types of stimuli, there is a similar upper limit in performance when the number of degrees-of-freedom per critical band is larger than eight. In this range discrimination performance decreases with increasing number of degrees-of-freedom with a similar slope. The data may indicate that the number of degrees-of-freedom per critical band is an important measure that determines the maximum discrimination performance that can be achieved by the listeners. In addition, for both types of stimuli, while there seems to be an advantage of integrating stimulus information across frequency for a low number of degrees-of-freedom below about eight degrees-of-freedom, this advantage seems to be lost when the number of degrees-of-freedom per critical band is too high. This can, for instance, be observed when comparing results for the Gaussian-noise conditions with a bandpass range of 225–275 Hz (grey triangles) to those for the conditions with a bandpass range of 100–600 Hz (grey squares), or, when comparing the results for the one-frequency tone-burst complexes (black circles) with those for the seven-frequency tone-burst complexes (black x symbols).

When the number of degrees-of-freedom was low, discrimination performance was higher for the broadband conditions than for the narrowband conditions. When the number of degrees-of-freedom was high, discrimination performance was similar for these conditions.

It seems that discrimination performance for these stimuli depends predominantly on the amount of peripheral information of an auditory object and the capacity to process this peripheral information. This capacity seems to be limited in the temporal dimension, leading to a poor discrimination performance when there is a large amount of peripheral information. In the spectral dimension no such limitation was observed. Depending on the number of degrees-of-freedom in the lowest critical band, performance either increased with increasing number of excited auditory filters, or it remained unaffected by additional peripheral information in other critical bands.

3 A model for Gaussian-noise discrimination[†]

Abstract

The performance of human listeners in discriminating Gaussian-noise tokens depends non-monotonically on duration (Hanna [Percept. Psychophys. **36**, 409–416 (1984)]). Discriminability initially increases with duration but decreases for durations above 40 ms, suggesting a limitation in the auditory system's information-processing capacity. Current psychoacoustic models based on optimal information processing do not predict this. In the present study, an extra stage was added to the model of Dau *et al.* [J. Acoust. Soc. Am. **99**, 3615–3622 (1996)] that restricted the number of samples in the IR to a fixed amount independent of the stimulus duration which implies that a stimulus interval is treated as an undividable auditory object. Moreover, the model's decision stage was adapted to make it applicable to noise discrimination. The adapted model was able to simulate the non-monotonic duration dependence, as well as to reproduce data concerning partially correlated noises, and to predict data for noise stimuli with an added noise fringe without changing any of the model parameters. These results support the hypothesis that the non-monotonic duration dependence is caused by a limited capacity for retaining or processing information about auditory stimuli.

[†] This chapter is based on Goossens, T., van de Par, S., and Kohlrausch, A. "A model for Gaussian-noise discrimination," submitted for publication to J. Acoust. Soc. Am.

3.1 Introduction

Listeners have little problems perceiving timbre or roughness of Gaussian noise samples or to discriminate between such samples based on differences in the underlying global signal properties; signal duration, spectral envelope, and envelope modulation. However, when listening to two tokens that are generated by the same (statistical) process, determining whether the two tokens are identical or are generated independently is more difficult (Hanna, 1984; Heller and Trahiotis, 1995, and the study presented in chapter 2 of this thesis), and, in some conditions, discrimination is at chance level.

From an information theoretical viewpoint a Gaussian noise signal contains much information because it is an unpredictable signal (Shannon, 1948). Therefore, listeners' ability to discriminate Gaussian noise may tell us something about the information processing limitations of the human auditory system. Hanna (1984) investigated noise token discrimination in a same/different paradigm as a function of bandwidth and duration. He found that with increasing noise bandwidth, at least for durations up to 100 ms, discrimination performance increased. This is in line with the idea that when more critical bands are covered by the stimulus, more peripheral information is available to the listener.

Several studies (Hanna, 1984; Fallon, 1989; Heller and Trahiotis, 1995, and the study presented in chapter 2 of this thesis), have established that the ability of human listeners to discriminate Gaussian-noise tokens increases with durations up to approximately 40 ms. This is illustrated in Fig. 3.1, where a selection of noise discrimination data from the literature is shown. These data are expressed as d' values as a function of duration. It can be seen that, in the various studies shown in Fig. 3.1, maximum discrimination performance is observed for durations of 25–100 ms. Above this duration, the ability to discriminate decreases with duration. This is remarkable because, in Gaussian noise, the number of stimulus features, i.e., differences across the two noise tokens on which listeners may base their discrimination, increases with duration (cf., chapter 2 of this thesis). This indicates that the spectral dimension may have a different influence on discrimination than the temporal dimension.

In a study of Watson *et al.* (1990), where listeners were asked to detect spectral-temporal changes in randomly generated patterns of multiple non-overlapping tones, it was found that the ability to detect such changes strongly depended on the number

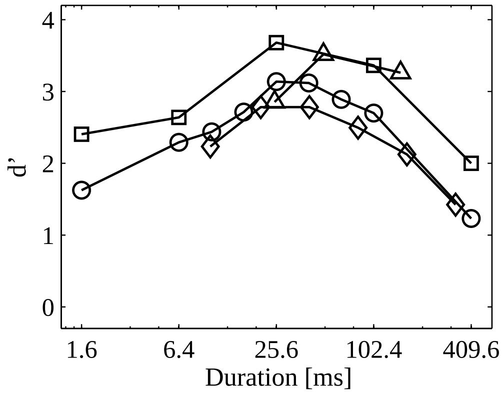

Figure 3.1: Discrimination ability of listeners for Gaussian noise in a same/different task. Expressed in d' values as a function of duration from several studies: Hanna (1984), noise bandwidth 100–3300 Hz; (squares); Fallon (1989), noise bandwidth 100–3000 Hz (triangles); Heller and Trahiotis (1995), noise bandwidth 800–1600 Hz (diamonds); and chapter 2 of this thesis, noise bandwidth 100–3300 Hz (circles).

of tonal components rather than the total duration. According to Watson *et al.* (1990) this suggests that, at least for tonal patterns "...performance may be limited by the number of acoustical components (or the amount of information) that can be held in immediate memory..."

A few models have been proposed for the discrimination of noise stimuli (Coble and Robinson, 1992; Rickert, 1998). These, however, focused mainly on modeling the influence of the temporal location of a target noise embedded within a longer noise stimulus on the ability to discriminate the target. These studies did not investigate the effect of changes in overall stimulus duration. They showed that the target noise contributes most to discrimination when it is located towards the end of the stimulus.

Other discrimination models, for example the model of Dau *et al.* (1996) using template matching, are relevant but not directly applicable to noise discrimination. The template approach assumes that the listener builds up knowledge about differences between a to-be-detected signal, the target, in the presence of another, signal, the reference. These differences are represented in the form of a template. In noise discrimination, the intervals containing the noise tokens do not show any systematic difference because they are generated from the same statistical process, that is, they have the same long-term spectrum. Therefore a template of this difference will equal zero which means that there is no a priori knowledge about the signal. A different approach is needed to use such a model for the discrimination of noise stimuli which directly compares Internal Representations (IRs) of reference and test intervals without resorting to a template that expresses what differences to expect. The closely related multiple looks model of Viemeister and Wakefield (1991) assumes that in a masking paradigm, information contributing to the detectability of the target is integrated across different temporal intervals to improve detectability of the target.

It is expected that models which combine all peripheral information over time, like template matching models (e.g., Dau *et al.*, 1996) or multiple look models (e.g., Viemeister and Wakefield, 1991), will not correctly predict the decrease of discrimination ability with increasing duration mentioned above. It is more likely that their discrimination performance will keep increasing with duration until it saturates at perfect performance.

We propose a method for predicting Gaussian noise discrimination using the preprocessing of the model of Dau *et al.* (1996) in combination with a new information limitation stage that accounts for the inability of listeners to combine all peripheral information over time. The rest of the study deals with some of the theoretical implications of the proposed model and tests some predictions that follow from these considerations.

3.2 A noise discrimination model

The modeling approach can be summarized as follows. First IRs of the two intervals in a trial are calculated using the model of Dau *et al.* (1996). Then, the size of these IRs is reduced to a fixed number of samples regardless of their initial duration. The sum of squares distance between these reduced size IRs is calculated. Using this distance, a decision is made whether the intervals were the same or different. In the next subsections these steps are explained in more detail.

3.2.1 Internal representation

The model for obtaining the IRs was originally developed by Dau *et al.* (1996). The model comprises a number of subsequent stages. The first stage is a fourth order gammatone filter bank to model basilar membrane filtering. Here, the signal is split up into 52 critical bands with center frequencies ranging from 20 Hz–10 kHz, spaced linearly on the ERB frequency scale (Glasberg and Moore, 1990).

Secondly, to model the inner-hair cell transformation, all critical bands are half-wave rectified and filtered with a fourth order low-pass filter with a cutoff frequency of 1 kHz.

Then, nonlinear adaptation is applied to each critical band using 5 subsequent adaptation feedback loops with time constants of 5, 50, 129, 253, and 500 ms. For a detailed description of the adaptation loops, see Püschel (1988); Dau *et al.* (1996); Tchorz and Kollmeier (1999). These adaptation loops are used to incorporate the adaptive properties of the auditory periphery and result in an approximately logarithmic compression of the steady state signal. Changes of the signal that are fast compared to the time constants are emphasized.

Finally, all critical bands are filtered with a first order low-pass filter with a cutoff frequency of 8 Hz and internal noise is added. The internal noise was calibrated such that the model was just able to detect intensity differences of 1 dB, as described in Dau *et al.* (1996).

As a result of these processing stages an IR is obtained that is a function of time and critical band. It will be designated with IR$[t, n]$, where t is time and n is critical-band number.

3.2.2 Information capacity limitation

If we take the two IRs resulting from the two intervals in a discrimination trial and integrate their squared differences over frequency and time we get an estimate of their perceptual distance. This difference will be larger for longer duration stimuli as there is a longer duration across which to integrate differences. A model based on this estimate will thus be expected to predict better discrimination with increasing stimulus duration. This was not shown in behavioral experiments where there was a maximum performance around 40 ms (cf. Fig. 3.1). We evaluated predictions based on comparisons between IRs using the integrated squared differences, results are shown by the triangles in Fig. 3.2, for discriminating 100–3300 Hz Gaussian noise tokens for various durations. For these predictions we replaced to 8-Hz smoothing filter that is used in Dau *et al.* (1996) by a sliding Hanning-window with a duration of 40 ms which matches the duration for which subjects have best discrimination performance.

Another way to estimate the distance between the two IRs is to integrate IRs across time, which is similar to determining the difference between the stimuli's long term spectra. With increasing stimulus duration, the variability of the long term spectrum of the IRs will decrease because they are integrated over a longer duration. Therefore, with increasing duration, discrimination ability will decrease (downward triangles in Fig. 3.2).

These two approaches lead to opposite predictions with regard to discrimination ability as a function of stimulus duration. Neither predicts the nonmonotonic dependence on duration that is seen in Fig. 3.1. Using a sliding integration window, of for instance 40 ms, to smooth the IRs reveals discrimination performance that saturates towards perfect performance for longer durations (triangles in Fig. 3.2).

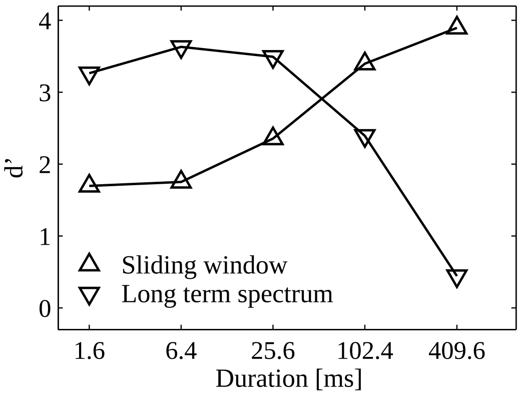

Figure 3.2: Model simulations for discrimination of Gaussian noise. Mean d' values as a function of stimulus duration for a model using a 40 ms sliding window (triangles), and a model estimating the long term spectrum (downward triangles). The spectral range of the noise was 100–3300 Hz

The non-monotonic duration dependency shown in Fig. 3.1 implies that listeners cannot make use of extra stimulus details that become available when stimulus duration is increased. Therefore, we hypothesize that the amount of internal information of a stimulus or auditory object that a listener has access to is fixed. In the following, this hypothesis is implemented in an extra stage in the model that follows the stage generating the IRs and that precedes the decision stage. This extra stage will be referred to as the Information Limitation (IL) stage. The basic idea of the IL stage is to transform the original $\mathrm{IR}[t,n]$ into a fixed-size IR of which each critical band has a fixed number of samples in the temporal dimension. The number of critical bands were left unaltered. The fixed-size IR is designated with $\widehat{\mathrm{IR}}[k,n]$, where k is the time sample number, and n is the critical band number. The fixed-size IR ($\widehat{\mathrm{IR}}$) was derived by multiplying each critical band of the IR with a number of 75% overlapping Hanning windows, see upper panel of Fig. 3.3. The length of the Hanning windows was *always* $\frac{1}{5}$ of the duration of the input stimulus' IR, including 75 ms of ringing of the auditory filters and adaptation loops. This caused the *length* of the Hanning windows to be directly dependent on the duration of the stimuli. In addition, the *number* of Hanning windows was independent of the duration with this method. The amplitude of the signal in each window interval was weighted with the window and averaged into a scalar value. The concatenation of these scalar values resulted in the fixed-size IR, see lower panel of Fig. 3.3. Effectively, this is method is a low-pass filter with a time constant that is inversely proportional to the stimulus duration.

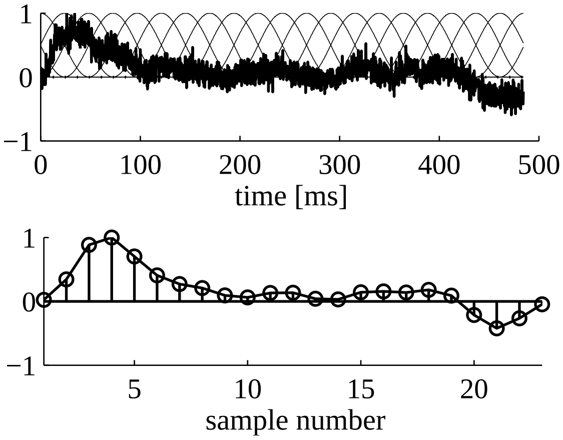

Figure 3.3: Example of windowing one critical band of an IR, IR_A, of a 409.6-ms noise stimulus (upper panel), resulting in a fixed-size IR, $\widehat{\mathrm{IR}}_A$, (lower panel)

.

The proportional window length of $\frac{1}{5}$ of the duration of the IR was the best fit to the data of Exp. 1 in a set of simulations containing the following proportional window lengths: 1, $\frac{1}{2}$,..., $\frac{1}{6}$, $\frac{1}{7}$. The fit was determined by calculating the sum of squared differences between the across subject mean of the behavioral d' data and the model simulations of all conditions in Exp. 1. Changing the length of the window to a smaller relative length with respect to the total stimulus duration will cause the number of windows to increase. Thus, also number of degrees of the fixed-size IR increases. As a result, the duration at which the discrimination performance reaches a maximum will shift to a longer duration. The amount of window overlap did not influence the performance of the model provided that it was 75% or more and was therefore chosen to be 75%, in order to limit computational complexity.

3.2.3 Distance metric

Because we are dealing with a noise discrimination task, the model's task is not the detection of a signal but determining the perceptual distance between stimuli. For *detection*, e.g., of a signal in the presence of a masker, the decision stage that is used in the model of Dau *et al.* (1996) uses prior knowledge of the stimuli. This prior knowledge is an estimation of the IR of the reference interval (often a masker) and of the target interval (often a masker + signal) which are obtained from various previous exposures to these intervals. For a given, unknown, interval, a decision variable is obtained from the correlation between the expected difference between test and reference intervals and the difference between the observed interval and the expected reference interval.

In the current noise discrimination task, such prior knowledge is not available since it is not known in what way the two intervals will differ from each other. Instead, to obtain a decision value (D), first the Sum of Squared Differences (SSD) between the fixed-size IRs of the two intervals is calculated using

$$SSD = \sum_k \sum_n (\widehat{\mathrm{IR}}_A[k,n] - \widehat{\mathrm{IR}}_B[k,n])^2, \tag{3.1}$$

where $\widehat{\mathrm{IR}}_A[k,n]$ is the fixed-size IR of interval A and $\widehat{\mathrm{IR}}_B[k,n]$ is the fixed-size IR of interval B.

Finally, to obtain a decision value D, decision noise ($N_{decision}$) is added to the SSD to adjust the overall performance of the model. This decision noise is a random scalar value from a Gaussian distributed noise-source. A higher decision noise results in a lower overall performance. $N_{decision}$ was adjusted such that the sum of the squared differences between the d' values of the model predictions and the behavioral results from Exp. 1 was minimized. This value for $N_{decision}$ was used for all simulations in this paper.

A similar distance metric was used in a number of other studies concerning audio quality assessment (e.g., Tchorz and Kollmeier, 1999; Hansen and Kollmeier, 2000; Huber and Kollmeier, 2006). These studies used similar auditory preprocessing to obtain IRs. Then, the linear cross correlation was calculated to obtain a decision variable, which is very similar to the SSD (cf., Green, 1992).

3.2.4 Decision stage

The proposed model is designed to perform the same/different discrimination experiment described in Exp. 1 of chapter 2 of this thesis, which will be explained in Sec. 3.3, and performs the experiment using the same experimentation software as used for presenting stimuli to the listeners. Thus, the model functions as an artificial listener.

The obtained decision variable D represents either the distance between two intervals in a same trial or in a different trial. D will, on average, be larger for a different trial. Therefore, the decision is made by comparing D to a criterion (C). If D is larger than criterion C, then the model gives the decision that the presented stimuli are different. Otherwise, the decision is that the stimuli are the same.

The same/different discrimination trials are presented to the model in blocks of 100 trials. A decision is given by the model after each trial. The criterion C, to which the decision value D is compared, is determined heuristically. At the start of a block of 100 trials, the criterion is set to a fixed arbitrary positive value (i.e., 100). The value of C is adjusted after each trial by storing the values of D in two separate bins. One bin is used for values that, after feedback from the experimentation software, are known to result from a same trial and the other for values that result from a different trial. The decision stage estimates the mean and variance of the values of D in both bins under the assumption that they are normally distributed. These statistics are used to determine a new criterion using maximum-likelihood estimates (Green and Swets, 1988/1966). In every subsequent trial, the model adapts to a more accurate criterion and its performance improves. Within ten trials, C on average converges to a value that is within 10% of the end value of C after 100 trials.

3.3 General method and stimuli

The experimental method was the same as the method described in Exp. 1 of chapter 2 of this thesis, which was a replication of an experiment by Hanna (1984). Listeners were presented in each trial with two noise stimuli which were identical or independent. The noise tokens were presented with an inter-stimulus interval of 500 ms. The listeners' task was to respond whether these tokens were the same or different. Feedback about the correctness of the answer was given to the listeners (or the model) after each trial.

The trials were presented in randomized blocks with an equal number of same and different trials. A sensitivity index, d', was calculated for each block of trials. Unless stated otherwise, the behavioral results were obtained using three blocks of 100 trials per experimental condition for each listener. The model predictions were obtained using 12 blocks of 100 trials per experimental condition.

The d' values were obtained by summing the Z-scores of the hit and correct rejection rates. It sometimes happened that a subject answered all same (or different) trials within a block correctly which resulted in an infinite d' value. In such cases, an extra artificial incorrect same (or different) trial was added to the block, thus providing a finite d' that could be used for calculating mean d' values and standard deviations. At chance performance, the d' value equals zero. Above-chance performance results in positive d' values, e.g., 69% correct for both same and different trials results in a d' value of approximately 1 and 84% correct for both same and different trials results in a d' value of approximately 2. The mean d' and standard error were obtained by pooling the block d'-values of each condition.

The spectrum level of the stimuli was 40 dB. The stimuli were generated from white Gaussian-noise with a specified duration that were filtered with a Chebyshev Type II digital filter with slopes of 100 dB/octave for the broadband and 500-Hz wide bands and approximately 200 dB/octave for the 50-Hz wide bands. The filters were designed using the Matlab (R14) filter design and analysis toolbox. Ringing of the filters was truncated after 150 ms, where the signal was sufficiently decayed to make truncation inaudible. The stimuli were presented diotically.

3.4 Experiment 1: Bandwidth and duration

In this experiment the bandwidth and duration of the Gaussian noise stimuli were varied in order to test their influence on discrimination. The behavioral results were taken from Exp. 1 of chapter 2 of this thesis, which was a replication of one of the experiments of Hanna (1984).

3.4.1 Method and stimuli

The noise stimuli were bandpass filtered Gaussian-noise tokens generated at a sampling frequency of 44.1 kHz with −3 dB bandpass ranges of 100–3300, 100–600, 225–275, 2800–3300, and 2975–3025 Hz and durations of 1.6, 6.4, 10.2, 16.1, 25.6, 40.6, 64.5, 102.4, and 409.6 ms prior to filtering. The behavioral results were obtained using four blocks of 50 trials (two listeners) or three blocks of 100 trials (one listener) per experimental condition.

3.4.2 Results

The left panel of Fig. 3.4 shows the results from the behavioral experiment from chapter 2 of this thesis. The best performance was achieved for the widest bandwidth (circles). Conditions with frequency ranges around 3000 Hz (diamonds and downward triangles) showed lower performance than conditions with frequency ranges around 250 Hz (squares and triangles). In general the ability to discriminate increased up to a duration of 25 to 40 ms. Above this duration discrimination ability decreased with increasing duration.

The right panel of Fig. 3.4 shows the results of the model simulations with an information-limitation stage incorporated in the model. The bandwidth ordering is correctly predicted. In addition, the model's discrimination performance now shows non-monotonicity similar to listener data showed in the left panel of Fig. 3.4: there was an initial increase of predicted d' values with increasing duration up to a duration of approximately 40 ms, and above this duration d' values decreased. The largest discrepancies with the behavioral data are observed for the 100–3300 Hz bands of short duration (1.6–16.1 ms) that are very close to with the 100–600 Hz bands of short duration in the behavioral results of but not in the predicted results.

By calculating the coefficient of determination[1] (R^2) it was found that the 64% of the variance of the data in was explained by the model simulations.

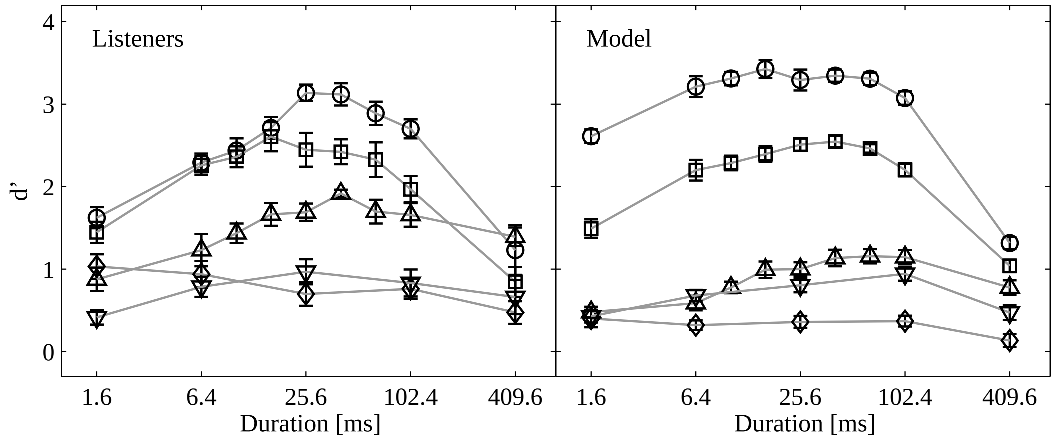

Figure 3.4: Mean d' values as function of stimulus duration for listeners (left panel) and for the model (right panel) with the information limitation stage. Spectral ranges were 100–3300 Hz (circles), 100–600 Hz (squares), 225–275 Hz (triangles), 2800–3300 Hz (diamonds), and 2975–3025 Hz (downward triangles). The error bars indicate plus and minus one standard error of the mean.

3.4.3 Discussion

The results in Fig. 3.4 show that the model predicts some of the essential characteristics of the listeners' data. The nonmonotonic dependence of performance on stimulus duration is predicted, as well as the increase in performance for low frequency noise that increases in bandwidth. Also the high frequency conditions tend to reveal the lowest d' values. A clear discrepancy between the data and the model is that there seems to be too much effect of bandwidth in the model for short duration stimuli.

The model's non-monotonic behavior as a function of duration is caused by two mechanisms with opposite effects. On the one hand, the amount of peripheral information elicited by the Gaussian-noise stimuli (chapter 2 of this thesis) increases

[1] Coefficient of determination was calculated with $R^2 = 1 - \sum_i (o_i - m_i)^2 / \sum_i (o_i - \overline{o})^2$, where o_i and m_i are the individual observed and model values, respectively, and $\overline{o}$ is the mean over all observations.

with duration (Hartley, 1928), causing the distance between the IRs, IR$_A$ and IR$_B$, in a *different* trial to increase as well. On the other hand, when duration increases and the samples in the internal representation are averaged by longer windows, the amount of variability after averaging reduces and hence the distance between the IRs of a *different* trial decreases. The trade-off between these effects produces maximum performance when the number of degrees-of-freedom of the IRs matches those of the fixed-size IRs. This also explains why the choice of number of windows in the IL stage influences the stimulus duration for which maximum performance occurs in the case of Gaussian-noise token discrimination.

Figure 3.4 shows that the model's performance increases with stimulus bandwidth. This is expected behavior because the information in the IRs is integrated over all filters. The behavioral results, to a large extent, also show such spectral integration (cf. left panel of Fig. 3.4). However, for durations below 25.6 ms, there seems to be no advantage for listeners of wider bandwidth in the 100–3300-Hz conditions over the 100–600-Hz conditions. In addition, listeners' performance for the narrowband 225–275-Hz condition with 409.6-ms duration seems to be at least as high as for the broadband 100–3300-Hz condition, which was also seen in the study of Hanna (1984). It is not well understood why listeners do not benefit from increased bandwidth for short stimulus durations like the model does. Apparently spectral integration is more complicated than summation over all critical bands.

It appears that the capacity limitation in the modified model of Dau *et al.* (1996) can predict the duration effect. However, the capacity limitation concept has a number of implications that need to be further investigated to assess the validity of the model. Central for the concept is that the stimuli are broken down into a number of parts using temporal windows (e.g., the Hanning window) and that the number of these parts is the same for each stimulus regardless of it's duration. An implicit assumption is that the model treats the noise stimuli as undividable auditory objects because the IL is always applied to the complete stimulus interval, i.e., it is not possible for the listener to distribute the windows only across a subpart of the stimulus. In the next section we test the implications of the modeling approach using listening tests.

3.5 Experiment 2: Interleaved durations

According to the model concept, the reduction of peripheral information in the IL stage depends on the total duration of the stimulus. It is an interesting question whether or not this information reduction depends on a priori knowledge of the duration of the stimulus interval. In a block-based design, the listener can learn the duration of the stimuli over the first few trials of a block which could predetermine the amount of peripheral information that will be lost *before* the presentation of a new stimulus interval. Alternatively, this duration-dependent reduction of peripheral information may happen *after* or *during* the presentation of the stimulus. In this case the length of the stimulus is (at least partially) known and decimation of peripheral information could take place accordingly.

The next experiment aimed to investigate whether listeners need to learn stimulus duration to predetermine the amount of peripheral information that will be lost or if it happens without such prior knowledge. This was done by presenting trials of short and long duration in an interleaved manner, which makes it impossible for the listener to predetermine how much peripheral information is to be lost as the duration of the subsequent trial is unknown.

3.5.1 Method and stimuli

In each block of 100 trials, noise tokens with two different durations were used. The duration of the two Gaussian-noise tokens was 25.6 ms in half of the trials and 409.6 ms in the other half. For both durations, half of the trials were same trials and half were different trials. The stimuli were bandpass filtered with -3 dB cutoffs at 100 and 3300 Hz. The trials were presented in random order to make sure that the listeners could not predict the interval duration of a trial beforehand. Three listeners participated, including the first and second authors.

3.5.2 Results

For the analysis, the interleaved trials were split up into a series of 50 responses to the 25.6-ms trials and 50 responses to the 409.6-ms trials for each block of trials. Separate d' values were calculated for each stimulus duration.

Figure 3.5 shows the results of the current experiment and results for corresponding durations of the experiment from chapter 2 of this thesis, which were replotted in the left panel of Fig. 3.4. The same subjects participated in both experiments. The boxes have lines at the lower quartile, median, and upper quartile values of the block d' values pooled over all subjects. The x's indicate the across-subject *mean* d' values. The whiskers extend from the smallest to the largest d' values. There were no outliers outside 1.5 times the interquartile range.

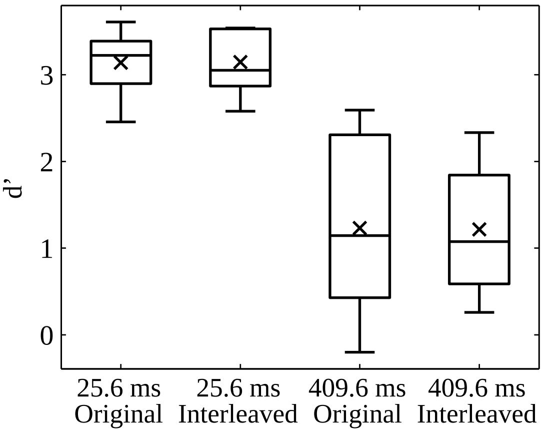

Figure 3.5: Boxplots for the original and interleaved duration experiment. The boxes have lines at the lower quartile, median, and upper quartile values and the x's indicate the mean d' values. The whiskers extend from the smallest to the largest observations in the data.

Analysis of variance with a linear model did not reveal significant main effects for presentation type ($F_{1,42} = 0.00$, $p > 0.98$), i.e. the difference between the original and the interleaved conditions. Listeners were treated as random effects. Significant main effects were found for duration ($F_{1,42} = 154.03$, $p < 0.001$) and for listener ($F_{2,42} = 12.51$, $p < 0.01$).

3.5.3 Discussion

The lack of a main effect for the type of presentation shows that it does not matter that the 25.6 ms and 409.6 ms trials were presented in an interleaved manner. Thus, we found no evidence that listeners depended on prior knowledge of stimulus duration. The significant effect of listeners was due to the fact that listeners had different baseline performances. Since listeners' performance did not depend on prior knowledge of stimulus duration, it appears that the peripheral information reduction occurs without a priori knowledge about the stimulus duration.

3.6 Experiment 3: Fringe configuration

The next experiment investigated the ability of listeners to discriminate stimuli of which one of the two stimuli in a discrimination trial had a noise fringe appended directly before or after it. This fringe, a sample of noise with the same statistical properties as the target token, was appended to the target token such that there was no audible cue at the transition of target and fringe. As a result, the two stimuli in a trial had different durations and only the ending or the beginning of the fringed stimulus was identical to the other stimulus in a *same* trial. In a *different* trial they were completely independent.

One of the properties of the model proposed in Sec. 3.2 is that the windows used to reduce the peripheral information in the IR have different lengths depending on stimulus duration. In the next experiment, where the stimulus intervals had different durations, the windows in the model's IL stage would have different lengths. Therefore, the samples in the fixed-size IRs would represent different temporal intervals for the two noise tokens in a trial. Comparing peripheral information in these fixed-size IRs across stimuli of different durations would be difficult for the model because of the low correlation between the fixed-size IRs even when the *target* noise-tokens are identical. The model will perform very poorly with these stimuli. The next experiment investigated whether listeners' discrimination performance is affected in the same way by the presence of forward and backward fringes or whether they were able to listen selectively to the target token while ignoring the fringe.

3.6.1 Method and stimuli

As in the experiment in Sec. 3.4, there were two noise tokens, which will be designated as "targets", in each trial. The listeners' task was to discriminate these target tokens. In addition, a fringe, i.e. a piece of non-informative noise with the same bandpass cutoffs as the target tokens, was present directly before or after one of the target tokens. During each block of 100 trials, the fringe properties were kept constant.

There were four kinds of fringe configurations, which are shown in Fig. 3.6. Fringes were either present directly in front of the token, i.e., a forward fringe, or directly at the end of the token, i.e., a backward fringe. The fringe was presented either in the first interval or in the second interval.

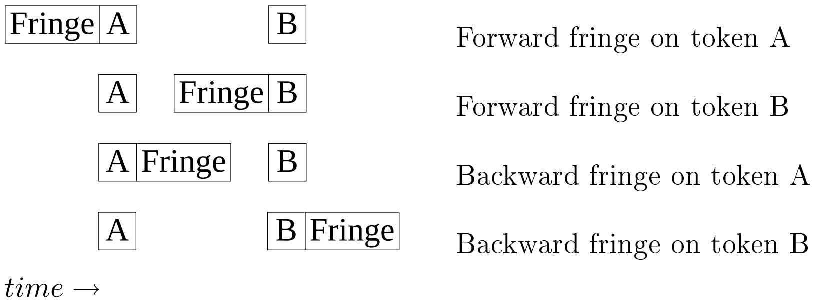
Fringe A
B
Forward fringe on token A
A
Fringe B
Forward fringe on token B
A Fringe
B
Backward fringe on token A
A
B Fringe
Backward fringe on token B
time →

Figure 3.6: Schematic timeline for the fringe conditions. Gaussian-noise target tokens *A* and *B* with duration 25.6 ms were either the same or different. The fringes were Gaussian noises with a duration of 384 ms. The duration between the onsets of target tokens A and B was always 909.6 ms.

Target tokens A and B were identical in a same trial and independent in a different trial. The task of the listener was to compare the target tokens and decide if they were the same or different. The listeners were informed if the target token was located at the beginning or ending of the fringe stimulus. Three listeners participated, including the first and second authors.

For each trial, new stimuli were made by generating two tokens of white Gaussian-noise with a duration of 409.6 ms, of which 25.6 ms served as the target and the remaining 384 ms as the fringe. In a different trial these tokens were independent. In a same trial they were identical. Subsequently, the stimuli were bandpass filtered with −3 dB cutoffs at 100 and 3300 Hz. Thus, at this point the two tokens both had a fringe. The fringe was removed from either token A or from token B by truncating the beginning or ending of the token using a 10-ms cosine ramp in order to limit spectral splatter. The target duration included the 10-ms cosine ramp.

3.6.2 Results

Table 3.1 shows the results of the fringed conditions. The first row shows the d' values for 25.6 ms, 100–3300 kHz Gaussian-noise tokens without fringe (taken from chapter 2 of this thesis). For this condition, listeners had an average d' of approximately 3.1. The d' for conditions with a forward fringe, as shown in the second and third rows, was approximately 0.03. The d' for conditions with a backward fringe, as shown in the fourth and fifth rows, was approximately 0.26.

Analysis of variance on the d' values showed significant effects of experimental condition ($F_{4,53} = 182.89$, $p < 0.001$). Tukey HSD post-hoc analysis revealed that d' for the 25.6 ms condition without fringe differed significantly from d' for the four fringe conditions. The fringe conditions did not show significant differences from each other.

Table 3.1: Results of the fringe conditions, d' mean and standard error of the original condition and conditions with four types of fringes.

	d'	std. err.
Original[1]	3.138	0.100
Forward fringe on token A	0.027	0.072
Forward fringe on token B	0.030	0.072
Backward fringe on token A	0.259	0.160
Backward fringe on token B	0.258	0.157

[1] Without fringe, from chapter 2 of this thesis

3.6.3 Discussion

For both the forward and the backward fringe conditions, the added fringe had a detrimental effect on listeners' ability to discriminate the target tokens. This effect was much larger than would be expected on the basis of the increased temporal separation between the target tokens with respect to the original conditions. The effect of increased temporal separation of this magnitude on discrimination was a reduction of approximately 0.1 d' as demonstrated in chapter 2 of this thesis.

Listeners performed poorly when a fringe was added to one of the tokens, which is in agreement with the expectations on the basis of the model and with the model hypothesis that stimulus intervals are processed as one inseparable unity with a fixed amount of information that a listener can use. Thus, listeners are not able to selectively process only the target part and ignore the fringe. This result is in line with an earlier experiment reported in chapter 2 of this thesis, where listeners were presented with 409.6-ms tokens of noise and were instructed to focus only on the beginning or end of the stimulus. Listeners were basically unable to improve performance when focusing on the beginning of the stimulus and only slightly when focusing at the end.

3.7 Experiment 4: Fringe duration

The previous experiment showed that adding a piece of non-informative noise, a fringe, to one of the target noise tokens severely decreased the ability to discriminate the target tokens. It is of interest to investigate for which fringe duration the decrease of discrimination ability starts, how steep this decrease is, and to explore the quantitative correspondence with the model predictions.

The duration of the fringe and the duration of the target tokens were varied in the next experiment to study the sensitivity of listeners to these parameters. Moreover, the results served as a critical test for the proposed model.

3.7.1 Method and stimuli

The experimental method was the same as the method described for the previous experiment (Sec. 3.6), except that all fringes were backward fringes on token B. The listeners were instructed to compare the beginning of the second interval to the first interval. The behavioral data were obtained using three listeners, including the first and second authors.

The target-token durations were 25.6 and 102.4 ms. The fringe durations were 0, 6.7, 15, 38.9, 76.8, and 384 ms when the target-token duration was 25.6 ms and 0, 26.8, 60, 153.6 and 307.2 when the target-token duration was 102.4 ms.

Construction of the stimuli was done slightly different from the fringe-configuration experiment, described in section 3.6. In the current experiment, the target of interval one as well as the target plus fringe of interval two were generated with the specified duration before bandpass filtering to a range of 100–3300 Hz. Note that in the previous experiment, the target of interval one was time limited *after* bandpass filtering. This did not lead to problems in the previous experiment where the fringe duration was longer than the ringing of the bandpass filter. In the current experiment, however, it was preferred to have the ringing of the bandpass filter treated in the same way for both the intervals, such that when the fringe duration approached zero, the two intervals could become identical.

In each trial, the duration of the second interval was longer than the duration of the first interval due to the presence of the fringe. The IL stage of the model placed the fixed number of windows over the entire duration of these tokens. Therefore, in absolute sense, these windows were larger for the interval containing the fringe.

3.7.2 Results

Figure 3.7 shows the across subject mean d' values as a function of the total duration of the second stimulus interval, i.e. target token duration plus fringe duration. The behavioral results are shown by the curves with the circles and the model results by the curves with the triangles.

The solid curves with the circles show the behavioral results with a target token duration of 25.6 ms. When the total duration of the stimulus interval was 25.6 ms, implying a fringe duration of zero, d' was 3.4. When the fringe duration was increased to 6.7 ms, d' was reduced by 1.5. Further increasing the fringe duration to 38.9 ms resulted in a further decrease of d' to a value of 0.3, which means that here it was nearly impossible for the listener to perform the task.

The dashed curves with the circles show the behavioral results with a target token duration of 102.4 ms. When the total duration of the second noise token was 102.4 ms, implying a fringe duration of zero, d' was 2.7. When the fringe duration was increased to 26.8 ms, d' was reduced by 0.9. Further increasing the fringe duration to 60 ms resulted in a further decrease of d' to a value of 0.7, which means that here it was difficult for the listener to perform the task.

The model simulations are indicated with triangles in Fig. 3.7. They show a high correspondence with the behavioral data. The standard error of the mean was below 0.23 for all conditions for both the behavioral results and the model predictions. It should be pointed out that all model parameters were identical to those derived form Exp. 1.

3.7.3 Discussion

The rapid decrease of discrimination ability predicted by the model simulations is in line with the behavioral data. Also, listeners were severely impaired in their ability to discriminate the target tokens when even a relatively small fringe was added to one of the tokens. This implies that listeners were not able to selectively compare the

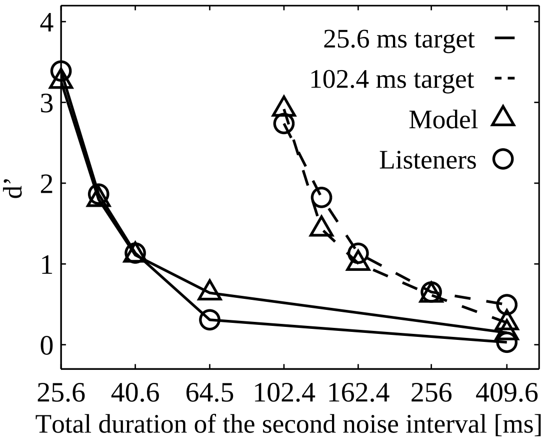

Figure 3.7: Mean d' values across listeners (circles) and of the model (triangles) as a function of the duration of the second stimulus interval (containing both the target noise and the backward noise fringe). Target-noise durations were 25.6 ms (solid lines) and 102.4 ms (dashed lines).

target of the first interval to the target that was followed by a fringe in the second interval

It was assumed in these model simulations that the fringe and target formed one auditory object because no cues were introduced that could have lead to their segregation. Therefore, the auditory object that contained both target and fringe had a longer duration than the object containing only a target. Thus, the fixed number of windows in the IL stage of the model was distributed over the entire duration of the trial intervals. This caused stimulus details for the target only interval to be represented in more samples of the fixed-size IR than for the target and fringe interval. This resulted in a mismatch of the samples in the reduced-size IRs with respect to the informative stimulus details of the target token, and hence, in poor discrimination performance of the model. The good agreement of the psychoacoustical data and the model simulations supports the underlying modeling assumptions that state that listeners process auditory objects as a unity and that they use a fixed amount of resources to retain or process auditory objects.

3.8 Experiment 5: Partially-correlated noise

In the previous sections it was shown that the model can account for the duration and bandwidth dependencies of behavioral noise-discrimination data. Moreover, the model predicted the decrease of discrimination ability when a fringe was added to one of the target tokens. To further investigate the limitations and capabilities of the model, the next experiment aimed to replicate data from the study of Fallon (1989). These data were also published in Coble and Robinson (1992), but because the data were given in percentage correct in this paper we use the data from Fallon (1989), where they were given as d' values. Fallon (1989) tested the influence of the proportional duration and temporal location of a target token in a partially correlated noise-discrimination experiment.

3.8.1 Method and stimuli

The experimental method in the current study was similar to the method of Fallon (1989). Their behavioral results were obtained using three subjects, who were each presented with four blocks of 100 trials per condition. The model results were obtained in 8 blocks of 100 trials per experimental condition.

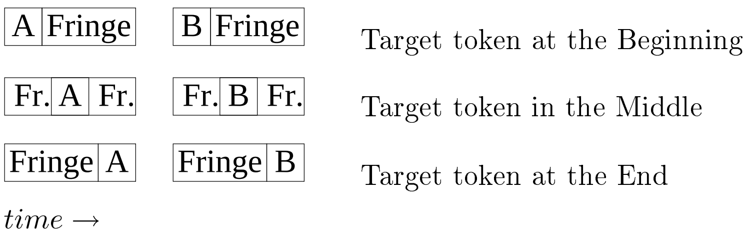

Figure 3.8: Schematic timeline for the partially correlated noise conditions. Gaussian-noise target tokens A and B were either the same or different and were positioned at the beginning, middle or end of the stimulus. In both intervals, the fringes were identical Gaussian noises, and were thus non-informative.

Target tokens A and B were positioned either at the beginning, middle or end of the stimuli, cf. Fig. 3.8. The task was to judge if the stimuli were the same or different.

The stimuli were very similar to those of Fallon (1989), with the exception that we used bandpass filtered Gaussian noise with –3 dB cutoffs at 100 and 3300 Hz for target and fringe, whereas Fallon (1989) used bandpass filtered Gaussian noise with − 3 dB cutoffs at 100 and 3000 Hz.

The relative duration of target noise with respect to the non-informative fringe was varied, while keeping the total duration constant at 25, 50, or 150 ms. In our simulations, the fringe duration, called τ, were 0, 5, 15, and 25 ms when the target duration $T = 25$ ms, 0, 5, 10, 20, 30, 40, and when $T = 50$ ms, and 0, 15, 30, 60, 90, 120, and 150 when $T = 150$ ms. The inter-stimulus interval between the first and second stimulus was 300 ms. Fringes were identical to each other within each trial. Target tokens A and B were uncorrelated in different conditions and identical in same conditions. Thus, the stimuli were completely identical in same conditions, and differed only in the target part in different conditions. New stimuli were generated for each trial. Filtering of the stimuli was done the same way as described in the previous section (Sec. 3.7).

3.8.2 Results

The top panels of Fig. 3.9 show the behavioral data, replotted from Fallon (1989) as means across subjects. The bottom panels of Fig. 3.9 show the corresponding predictions using the proposed model. From left to right the panels show results of conditions where the target tokens were located at the beginning, middle, and end of the stimuli.

Fallon (1989) noted that the ability to discriminate the target tokens was influenced by their temporal location and by the ratio of target duration to total stimulus duration. The ability to discriminate increased when the temporal location of the target token was moved from the beginning to the end of the stimuli. The ability to discriminate also increased with the proportional duration τ/T of the target token. For the stimulus durations used in her study, the proportional duration of the target tokens was more closely related to d' than the absolute duration of the target tokens. For the model, the ability to discriminate also increased along with the proportion of τ/T, and like the behavioral data, the curves overlap when plotted on as a function τ/T.

The model was not successful in reproducing the dependence of sensitivity on the temporal location of the target token. The model had the lowest d' when the target tokens were located in the middle of the stimuli.

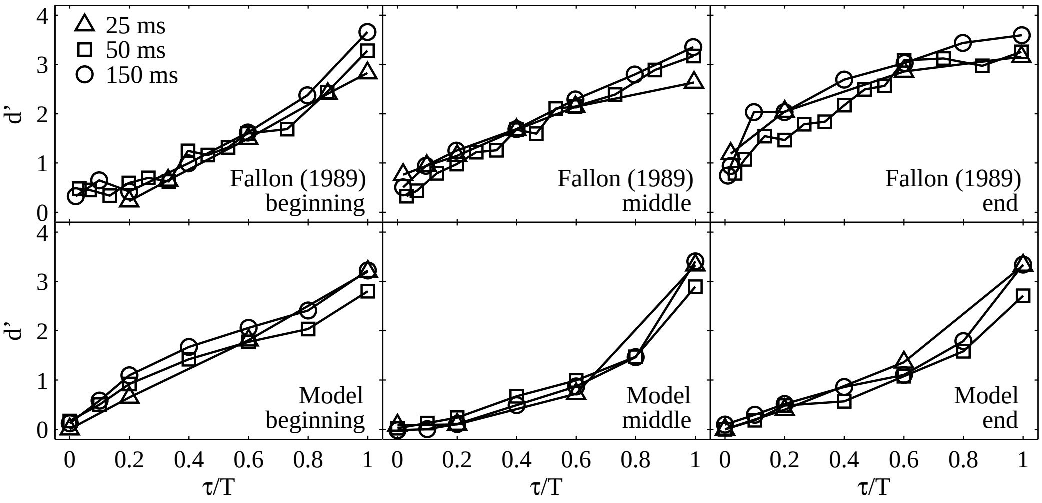

Figure 3.9: Mean d' values for partially-correlated noise conditions as a function of the duration of the target noise divided by the total duration (τ/T) of the original behavioral data replotted from Fallon (1989) (top three panels) and of the model simulations (bottom three panels). Total stimulus durations were 25 ms (triangles), 50 ms (squares), and 150 ms(circles). The target noises were located at the beginning (left panels), the middle (middle panels), or the end (right panels) of the stimulus.

3.8.3 Discussion

What enables the model to discriminate two noise tokens are differences between their IRs. Regions of these IRs where the variability across IRs is relatively higher reveal more differences, and therefore contribute more to discrimination. For the noise tokens this variability is constant over time, however, for the IRs of the noise tokens this variability is amplified at the onset and offset due to the adaptation loops. For the fixed-size IRs most of their variability was therefore located in the windows that overlapped the onset and offset of the stimuli.

Apart from the adaptation, there is no mechanism in the model to provide a higher sensitivity to stimulus differences that are located more towards the end of the stimuli. Thus, it was not expected that the dependency of d' on the temporal location of the uncorrelated noise would be correctly predicted. Fallon (1989) proposed an exponential weighting function for the IR to simulate a higher sensitivity for stimulus differences at the end of the stimulus. Although the effect of temporal location of the target was not predicted by the model, the effect that *proportional* target duration with respect to the total duration governed discrimination was correctly predicted.

3.9 Experiment 6: Two fringes

The experiment of Fallon (1989) was very similar to the experiments described in section 3.7, but with fringes on both target tokens instead of one. Therefore, the total duration was identical for the two stimulus intervals in a trial. In terms of the model, this causes the samples in the reduced size IRs to represent the same temporal intervals, which was not the case in the single-fringe experiment. Therefore it was expected that the model's ability to discriminate would increase by adding this second fringe. This counterintuitive prediction, that performance should improve by adding a non-informative fringe also to the *first* target, was tested with behavioral experiments, which also served as a test for the model.

3.9.1 Method and stimuli

The method was identical to the method of the previous experiment (section 3.8) which was a same/different experiment with fringes added to both target tokens in each trial. In this experiment, fringes were again added to both target tokens, but now the target tokens were always positioned at the beginning of the stimulus and the duration of the target tokens was always 25.6 ms. The fringe durations were 0, 6.7, 15, 38.9, 76.8, and 384 ms. The signals were time limited before bandpass filtering them to ranges of 100–3300 Hz.

3.9.2 Results

Figure 3.10 shows the results of the current experiment with dash-dotted lines. The solid lines show the results for the 25.6-ms targets with one fringe from section 3.7 for comparison. The circles indicate the behavioral results and the triangles indicate the model predictions.

When the duration of the two-fringes conditions was zero (total duration of the second interval is 25.6 ms), the obtained d', which was close to d' for the single-fringe conditions, as expected. When the duration in the two-fringes condition was increased to 6.7 ms (total duration of the second interval is 32.3 ms), the d' value was reduced by 0.25. Compared to the single-fringe condition, this is a more shallow decay of discrimination ability. Further increasing the fringe duration in the two-fringes condition to 38.9 ms (total duration of the second interval is 64.5 ms) resulted in a further decrease of d' to a value of 1, which is approximately 0.7 more than for the single fringe case. The model simulations decreased almost linearly on a logarithmic scale from a d' value of 3.2 to 0.2. The standard error of the mean was below 0.20 for all conditions for both the behavioral results and the models simulations.

3.9.3 Discussion

The prediction of the model that the ability to discriminate a double-fringe condition is increased relative to the single-fringe conditions was confirmed. The increased performance of the model for the two-fringes conditions can be understood because for this condition the target tokens are mapped to the same internal axis. These results provide further support for the underlying modeling assumptions that state that that listeners process auditory objects as a unity and that they use a fixed amount of resources to retain or process auditory objects.

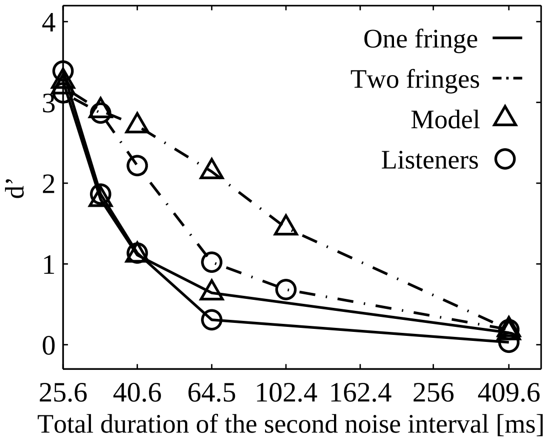

Figure 3.10: Mean d' values for listeners (circles) and for the model (triangles) as a function of the duration of the second stimulus interval (containing both the target noise and the backward noise fringe). Backward fringes were either added to only the second target noise (solid lines) or to both target noises (dash-dotted lines). Target-noise duration was 25.6 ms.

3.10 General Discussion

A number of studies have revealed a non-monotonic duration effect for discrimination of Gaussian noise (Hanna, 1984; Fallon, 1989; Heller and Trahiotis, 1995, and the study presented in chapter 2 of this thesis). In these studies, it was shown that the ability of listeners to discriminate Gaussian noise tokens increased with duration up to a duration of approximately 40 ms. Above this duration, discrimination ability decreased with duration. Such a non-monotonic duration effect would not occur if listeners could make optimal use of peripheral information because peripheral information resulting from Gaussian noise inherently increases with duration (cf., chapter 2 of this thesis). In psychoacoustical models, for example, the model of Dau *et al.* (1996) or of Viemeister and Wakefield (1991), it is often assumed that peripheral information is integrated over time and, hence, that discrimination ability increases with duration or saturates at a certain performance, but not that it decreases.

In the current study it was hypothesized that the non-monotonic duration effect was caused by a limited capacity of listeners to retain or process peripheral information represented in an auditory object. To test this hypothesis, a stage was added to the model of Dau *et al.* (1996), which limited the information in the IR. With this information limitation stage, the model could simulate the non-monotonic duration effect. This was an important step because, to our knowledge, the duration effect had not been successfully modeled before.

In additional conditions, the model also appeared to be able to reproduce data from the literature (Fallon, 1989), as well as to predict data for a new experiment without changing any of the model parameters. In the study of Fallon (1989), pieces of non-informative noise, fringes, were appended to both targets in a discrimination trial, whereas in the new experiment, a fringe was added to only one of the target tokens. In both experiments, the fringes impaired the ability of the listeners to discriminate, more so when a fringe was added to only one of the target stimuli. This can be understood with the proposed model approach, as is shown by the model simulations. It is related to the basic model assumption that stimulus intervals are processed as inseparable units.

Paramount to the proposed model approach is that the stimuli are processed as discrete auditory objects, or units in the nomenclature of Bregman (1990). The post processing of the IR, where it is decimated to a fixed-size IR, is dependent on the total duration of these auditory objects. This is in agreement with the statement of Kidd and Watson (1992), in a study concerning random tone patterns, that “For a considerable range of total durations, [...] some limited resource is being distributed across the extent of a sound, indicating that the sound is treated as a discrete entity.” Bregman (1990) noted that auditory units can be grouped into a new unit when they are sufficiently similar. It is therefore important to recognize that the Gaussian noises employed in this study were homogeneous stimuli that did not contain cues that could have led to their segregation into sub-objects, and that the interstimulus interval of 500 ms was sufficient for the two stimuli to be perceived as separate objects.

Another important aspect of the model is the assumption of the information capacity limitation of the IR that is used for the discrimination task. Several studies have found supporting evidence for such a capacity limitation (e.g., Watson, 1987; Cowan, 2005). Often, e.g., in the framework of Cowan (2001), this limit is attributed to a limitation of the focus of attention. Our model was not aimed at explaining the nature of the limitation. Rather, such a limitation was implemented to verify whether it could explain the degraded discrimination ability for stimuli with a duration longer than 40 ms.

Despite the good performance of the proposed model for noise *discrimination* tasks, it is unclear how it will perform on the psychoacoustical experiments presented to the original model in the study of Dau *et al.* (1996). However, it is not certain that the *detection* task, for which it was originally designed, would need to be dependent on the reduced size IR alone. In a detection task the listener knows on what changes in the stimulus to focus, which makes the task different from a noise discrimination task. According to Näätänen and Winkler (1999) there are indications that, in a detection experiment, listeners may benefit from direct access to sensory *feature* traces, analogous to the original IR containing all details after peripheral transuction. Whereas for other higher order tasks, such as a discrimination tasks, this direct access is not available. Such tasks would then need to use the information in the *unitary* stimulus representation, analogous to the fixed-size IR containing information about the stimulus as a unity. Moreover, in a detection task the target tone and the masker could be separate auditory objects, especially when the tone is well above threshold. It is, as yet, undefined how the proposed model should cope with such conditions.

Several models have been introduced in which the discrimination performance is a function of a target's proportion of the total stimulus duration, e.g., the proportion-of-the-total-duration (PTD) rule of Kidd and Watson (1992) and the component-relative entropy (CoRE) model of Lutfi (1993). These models are qualitative/descriptive models that predict the discrimination for stimulus components on basis of their relative variance with respect to the other stimulus components. However, these models do not take into account the total stimulus duration and therefore do not predict the effect of total duration that was observed in the first experiment. Admittedly, such effects were much less pronounced in the tonal patterns for which these models were designed. For Gaussian noise, however, there was a strong effect. Also, the approach of the current model is different. Whereas the PTD and CoRE models operate on the stimulus parameters as qualitative/descriptive models, the proposed model acts as an artificial listener on the waveforms of the stimuli.

In Sec. 3.8 the model simulated the results of an experiment of Fallon (1989) concerning the influence of temporal location of stimulus differences in the uncorrelated part of a noise token on discrimination performance. It was shown that, while the model was able to simulate the dependence of performance on the proportion of the uncorrelated stimulus part with respect to the total duration, it did not correctly simulate the dependence on temporal location. Instead of systematically better discrimination performance when the uncorrelated part was located more towards the end of the stimulus, it showed equal performance for stimulus differences located at the beginning and end of the stimulus, and slightly worse performance when the stimulus differences were located in the middle. A better dependence of the model on temporal location of the stimulus differences could be provided by the addition of temporal weighting.

In conclusion, this investigation presented corroborating evidence for the hypothesis that the inverse relationship of stimulus duration and discrimination ability for Gaussian-noise tokens with durations larger than 40 ms is caused by a limited information processing capacity for auditory stimuli. This capacity is allocated to auditory objects and these objects seem inseparable in the sense that it is not possible for listeners to selectively listen to only a part of the object.

4 Gaussian noise discrimination as a new approach for studying auditory object formation[†]

Abstract

The present study makes use of two observations. In a same/different experiment, listeners are good at discriminating 50-ms Gaussian-noise tokens with a spectral range of 350-850 Hz. However, when an identical 200-ms noise fringe, with the same statistical properties as the 50-ms target tokens, is appended to the end of the two target tokens, listeners show very poor discrimination performance. Apparently, identical uninformative fringes cannot be ignored and they impair the discrimination of the target tokens. When a perceptual cue is introduced that can lead to the segregation of the target token and noise fringe, e.g., a temporal gap between target and fringe, the ability to discriminate improves implying that the non-informative noise can be (partly) ignored when it is part of a different auditory object than the target token. It seems that a target token and the appended fringe form one auditory object and that access to subparts of these tokens is not possible. This method is used to investigate the influence of cues such as spectral range, level, interaural level difference, and interaural time delay on the formation of auditory objects. In this study, spectral separation and temporal separation were the strongest cues for auditory object formation.

[†] This chapter is based on Goossens, T., van de Par, S., and Kohlrausch, A. “Gaussian noise discrimination as a new approach for studying auditory object formation,” submitted for publication to J. Acoust. Soc. Am.

4.1 Introduction

In his work on auditory scene analysis, Bregman (1990) distinguishes between acoustic events, auditory streams and auditory units. An acoustic event is a happening in the physical world, causing vibrations that can be picked up by our hearing system. Examples are the acoustic events caused by walking in the streets, or rain falling. In everyday situations, many such acoustic events occur in rapid succession. We perceive these acoustic events with our auditory system where the two waveforms arriving at our eardrums are analyzed and divided into several separate perceptual entities, often called auditory streams, e.g., a stream containing the footsteps of somebody walking and a stream of the sound of rain. The process of auditory scene analysis is the segregation of acoustic information into separate auditory streams. The auditory streams can be a combination of several smaller auditory entities which can be named auditory objects, or units in the nomenclature of Bregman (1990), e.g., the sound of the individual footsteps of the person walking in the streets or of the drops of rain splashing into a puddle.

We can, to a certain extent, direct our attention deliberately to either one of these auditory streams (Alain and Arnott, 2000). In their study on selectively attention for auditory objects, Alain and Arnott (2000) adopt the definition of Bregman (1990) that an auditory object "[. . .] is the percept of a group of sounds as a coherent whole seeming to emanate from a single source." Much effort has been put into studying the principles that govern the grouping of auditory objects into perceptual streams, e.g., by van Noorden (1975). Reviews of the literature can be found in Bregman (1990), Yost and Sheft (1993), and Darwin and Carlyon (1995). Van Noorden (1975) found that the perceived relation of successive tones in a sequence depends on their temporal and spectral distance. He used sequences of alternating tones, in the form of an $ABAB$ pattern, for which he varied their frequency and intertone interval. When their spectral distance was sufficiently close, they were inevitably perceived as a single auditory stream. However, for some combinations of spectral and temporal distance, the tones were segregated into two separate auditory streams, an AA and a BB stream.

The As and Bs in these patterns can be considered to be the previously mentioned auditory objects, or units, of which the streams consist. Bregman (1990, pg. 644) notes about units that

> "It appears as though there is a unit forming process that is sensitive to discontinuities in the sound, particularly to sudden rises in intensity, and that creates unit boundaries when such discontinuities occur."

He further notes that these units can occur at different time scales and that smaller units can be embedded in larger ones. This last point is illustrated by the study of Royer and Robin (1986) who used cyclic patterns of tone bursts. They showed that at sufficiently high repetition rates the pattern was perceived as a repeating unit, while at lower repetition rates, the individual bursts were perceived as single auditory units.

Yost (1991) distinguishes at least seven physical parameters that contribute to the formation of auditory objects: spectral separation, intensity profile, harmonicity, spatial separation, temporal separation, common temporal onsets and offsets, and coherent slow temporal modulation. For example, several studies (e.g., Buell and Hafter, 1991; Woods and Colburn, 1992) showed that when one of the components in a harmonic tone complex had an asynchronous onset with respect to the other components, two sound objects were reported instead of one when the components had synchronous onsets. This indicates that harmonic tones are likely to be fused into a single auditory object, but it is possible that they are segregated into different objects when there is evidence, e.g., an asynchronous onset of one of the components, indicating that they were not caused by the same physical event.

In chapter 2 it was suggested that the discrimination between Gaussian-noise tokens may be related to object formation. In this study the ability of human listeners to discriminate between Gaussian-noise tokens was investigated. It was found, in agreement with the literature, that performance increases with duration up to a duration of approximately 40 ms (Hanna, 1984; Heller and Trahiotis, 1995, and the study presented in chapter 2 of this thesis). For longer noise stimuli, discrimination decreased. In chapter 3, this effect was modeled by assuming a fixed capacity for retaining or processing an auditory object, independent of the duration of the object. Therefore, more information was lost in the internal representation of the longer stimulus than in the internal representation of the shorter stimulus, which caused a maximum performance at around 40 ms duration.

In chapter 3, the Gaussian-noise target tokens were extended with a piece of uninformative Gaussian noise, a fringe. When a fringe was concatenated without fringe alteration (i.e., the fringe has similar bandwidth, level, lateralization, etc. as the target) to one of the to-be-discriminated target noises, discrimination ability dropped substantially. The interpretation was that listeners were not able to listen to a subpart of an auditory object (noise token) and that the retention or processing capacity needed to be attributed to the whole auditory object, leading to much poorer discrimination performance. The model correspondence between simulations and behavioral data in this chapter supports this interpretation.

The assumption in chapter 3 that auditory objects are processed as inseparable units raises the question as to what stimulus properties are needed to create an auditory object. More specifically one would expect that, when target stimuli have an added fringe, as described above, introducing perceptual cues in the noise fringe which enable the segregation of target and fringe, should improve the ability to discriminate the target noises compared to the situation where no such cues are present. This would provide a new method for assessing the importance of particular segregation cues in creating auditory objects consisting of Gaussian noise.

Several examples are known where segregation influences the perception of low level cues. The formation of auditory objects is sometimes investigated by presenting cyclic or continuous patterns to the listener (e.g., Royer and Robin, 1986; Crum and Bregman, 2006). Royer and Robin (1986), as mentioned above, found that the repetition rate of a sound-burst pattern influences the way the bursts are integrated into auditory objects. Crum and Bregman (2006) showed that a gradual timbre change in a continuous sound is detected earlier when silences are inserted, which cause unit boundaries, than when the sound is presented continuously without silences or when the silences are filled with loud noise bursts.

Buell and Hafter (1991) and Woods and Colburn (1992) investigated the formation of auditory objects in a noncyclic paradigm. These studies used harmonic tones of which the target tone was segregated into a separate auditory object from the remaining harmonics, by presenting it with an asynchronous onset, or by making the remaining tones inharmonic with respect to the target. They measured the influence of segregation on the detection thresholds for the targets interaural time delay, and found that, when the target tone was perceptually segregated from the maskers, interaural time delay detection thresholds were lower than when it was fused with the maskers.

In the current study we investigate a number of spectral, temporal, intensity, and spatial cues to assess their influence on object formation in a noise discrimination paradigm. In each trial, two Gaussian target noises of 50 ms duration with a spectral range of 350–850 Hz were presented that could be identical or independently generated. For this combination of duration and bandwidth discrimination performance is good. The task of the listener was to decide whether these targets were the *same* or *different*. In most conditions, an identical fringe was appended to the two targets in a trial. It was hypothesized that the ability to discriminate would be low when the fringe was perceptually fused with the target into a single auditory object and high when the fringe and target were perceptually segregated into two auditory objects.

4.2 Method

A *same/different* task was used for measuring discrimination performance. In each trial, two stimulus intervals were presented to the listener, both containing a target noise and a backward fringe, i.e. an uninformative noise. The exception was for the baseline condition in which no fringes were presented. The targets could be identical or independent. These stimulus intervals were separated by an Inter Onset Interval (IOI) of 800 ms (unless stated otherwise). The trials were presented in blocks of 100, of which half of the trials had identical (*same*) target noises, and the other half had independent (*different*) target noises. Same and different trials were presented in random order. The fringes of the two intervals were always identical, and thus, uninformative for the discrimination task. The listeners' task was to decide whether the *target* tokens were the same or different.

Conditions differed in the type of cue that was present in the fringe, which could potentially enable listeners to segregate the target from the fringe. For each experimental condition, four successive blocks of trials were presented, of which the first block was discarded. Listeners were allowed to take a break after a succession of four blocks if desired. Sensitivity indices, d', were calculated from the results for the last three blocks.

A d' value was obtained by calculating percentages correct for the *same* trials and the *different* trials. These percentages correct were converted to z-scores. Finally, d' was calculated by adding the z-scores for the same and the different trials. At chance performance, the d' value equals zero. Above-chance performance results in positive d' values, e.g., 69% correct for both same and different trials results in a d' value of approximately 1, and 84% correct results in a d' value of approximately 2.

In most figures, the d' values of the individual subjects were normalized with their individual performance in the baseline conditions measured in section 4.4. The baseline conditions comprised one condition where only the targets (no fringe, **NF**) were presented, and one condition where the targets had fringes that were appended without fringe alteration (fringe, **F**). Discrimination values for these conditions were obtained once at the beginning and once at the end of the experiment. The mean values for the session at the *beginning* of the experiment were used for normalization. The normalized d' will be called the effect and is obtained for a condition **X** using:

$$\text{effect} = \frac{d'_X - d'_F}{d'_{NF} - d'_F}. \tag{4.1}$$

This maps the performance for condition **F** to an effect of zero, and the performance for condition **NF** to an effect of one, in this way normalizing differences in baseline performances of different listeners. Note that, in the across-subject means, the individual results were first normalized using the individual baseline performances before pooling the data.

Four subjects, including the first (subject S3) and second authors (subject S2), who were all experienced in psychoacoustical experiments participated in this study. First, the baseline conditions, cf., Sec. 4.4, were repeated until stable performance was achieved before continuing with the rest of the conditions. During the training, at least 16 blocks of 100 trials were presented. The stimuli were generated on a PC and were presented, through an RME DIGI96/8PAD 24 bit PCI Digital audio card, a Tucker-Davis technologies S3 HB7 headphone driver and PA5 programmable attenuator, on Beyerdynamic DT 990 Pro headphones.

4.3 Stimuli

Each stimulus interval contained a target and a fringe (cf., Fig. 4.1). The targets were Gaussian-noise tokens with a duration of 50 ms with 10-ms raised cosine onset and offset ramps, which were applied after bandpass filtering to a bandwidth of 500 Hz with a center frequency of 600 Hz. In each trial, new noise was generated for both target and fringe tokens. The targets were presented diotically with a spectrum level of 50 dB. The target noise properties were never altered throughout the experiment.

The fringes were Gaussian-noise tokens with a duration of 200 ms with 10-ms raised cosine onset and offset ramps which were applied after bandpass filtering.

Unless stated otherwise, the fringes were bandpass filtered to a bandwidth of 500 Hz with a center frequency of 600 Hz and were presented diotically, like the target.

The filtering was done with a digital FFT filter that transformed the signal (of 44100 samples at 44100 Hz sample rate) to the frequency domain, where all frequency components lying outside the specified bandwidth were set to zero. This signal was transformed back to the time domain with the inverse FFT operation.

The fringes always followed the target and their onset and offset ramps overlapped such that the temporal envelope at the overlap was constant (unless stated otherwise). Moreover, it was made sure that during the 10-ms overlap, the fringe was identical to the target, also in a "different" condition, in order not to introduce indentation in the temporal envelope that could serve as a discrimination cue.

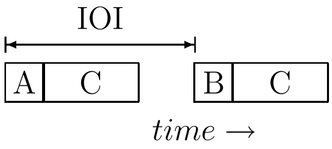

Figure 4.1: A schematic representation of a discrimination trial. A trial contained two stimulus intervals, each containing a 50-ms target (A and B) followed by a 200-ms fringe (C). The targets could be identical (same) or independent (different) across the two trials. The fringes were always identical. The Inter Onset Interval (IOI) was 800 ms (unless stated otherwise).

4.4 Experiment 1: Baseline conditions

The baseline conditions comprised one condition where the targets were presented without fringe, i.e., the condition for which the highest performance was expected, and a condition with fringe which had no audible cue that could lead to the segregation of target and fringe, i.e., the condition for which the lowest performance was expected. The two baseline conditions were tested twice; once at the beginning of the experiments, and once at the end of all the experiments, which provided information on whether the subjects were influenced by a training effect. In the remainder of this study the individual baseline conditions of the first session were used to normalize the data of each subject (cf. section 4.3) which enabled the comparison of data across subjects. The properties of the targets and fringes were as specified in Sec. 4.3.

4.4.1 Results

Figure 4.2 shows the results for the baseline conditions. In the conditions labeled **NF** (No Fringe) only the target was presented. In the conditions labeled **F** (Fringe) a fringe was added, which had the same bandwidth as the target (no fringe alteration). The conditions with a subscript b were obtained before all other conditions and those with a subscript e were obtained after all other conditions. The sensitivity index d' is shown on the ordinate. The error bars indicate the standard error of the mean.

Table 4.1 shows the results of pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. This test showed no significant difference between the conditions at the start and at the end of the experiment for either the baseline condition without a fringe (**NF**) or the baseline condition with a fringe (**F**). This shows that we do not need to take a learning effect into account. Additional comparisons showed that there was a highly significant difference between the conditions without a fringe and the condition with fringe at the start of the experiments ($d'_{NF_b} - d'_{F_b} = 2.0$) as well as at the end of the experiments ($d'_{NF_e} - d'_{F_e} = 1.7$); see Table 4.1.

Table 4.1: Pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$ for the baseline conditions. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
NF$_b$	NF$_e$	1.36	33	0.183	No
F$_b$	F$_e$	-0.12	33	0.906	No
NF$_b$	F$_b$	12.51	33	<0.001	Yes
NF$_e$	F$_e$	11.03	33	<0.001	Yes

4.4.2 Discussion

The results showed that performance for the condition where a fringe was appended to the target without fringe alteration (**F**) was significantly lower than for the condition where no fringe was present (**NF**). A explanation for the poor performance in the fringed condition with respect to the target-only condition, is that the memory

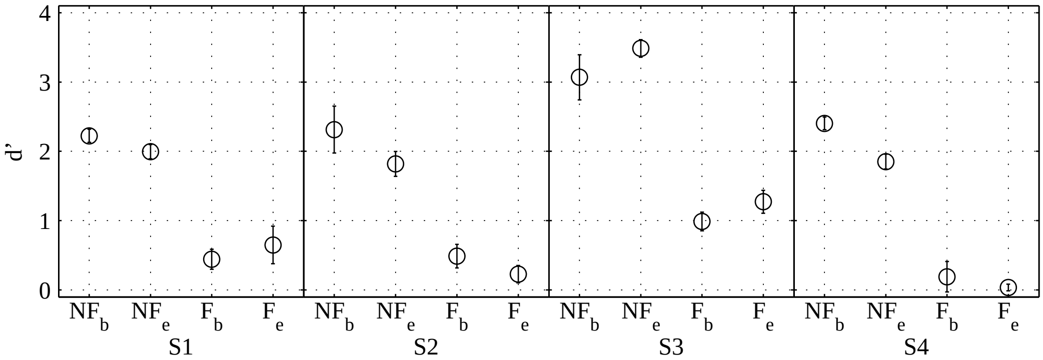

Figure 4.2: Results for the baseline conditions. Mean d' values of individual subjects are shown. In condition **NF** (No Fringe) only the target was presented. In condition **F**, both a target and fringe, without fringe alteration, were presented. The conditions with subscript b were obtained at the beginning of the experiment and those with a subscript e were obtained at the end. The error bars indicate the standard error of the mean.

for the target suffers from interference by the succeeding fringe by the presence of new stimulus information that enters the periphery. In the absence of a fringe, the target is not degraded by memory interference and performance is higher.

An alternative explanation is that, since there is no fringe alteration, the target and fringe are combined into a single auditory object and the discrimination is influenced by interference within this auditory object. For example, because the processing or retainment capacity for such an auditory object is limited and because listeners are not able to deliberately attend to a subsection of such an auditory object (cf., chapter 3). Therefore listeners allocate their resources to the whole auditory object, which are sufficient to accurately retain or process a short target but insufficient for the longer object comprising a target plus fringe.

These alternative explanations cannot be distinguished based on the data shown in Fig. 4.2. However, they make different predictions for conditions in which the fringe is presented after the target as a separate object. Therefore, in the following experiment, temporal gaps of varying duration are introduced between target and fringe.

According to the latter explanation the ability of listeners to discriminate the targets should increase in a condition where the listener is able to perceptually segregate target and fringe into separate objects. For the former explanation, the mere presence of the fringe should be sufficient to lead to a poor performance. Thus, according to this view one would not expect a strong performance increase in such a condition.

4.5 Experiment 2: Gap duration

4.5.1 Stimuli

The properties of the targets and fringes were as specified in Sec. 4.3. Except that, a temporal gap was inserted between target and fringe with a duration of 0, 20, 40, 80, 160, or 320 ms. This gap duration was defined as the temporal distance between the middle of the offset ramp of the target and the middle of the onset ramp of the fringe. The IOI was always 800 ms, except for the 320-ms gap, where it was 960 ms.

These experiments resemble one of the experiments of Hanna (1984, Exp. 2a), which was basically the same, except that he added a fringe with a temporal gap of 100 ms to only one of the targets in each trial, whereas in the current experiment a fringe was added to both targets. As an additional condition, we replicated his condition where a fringe was added to the end of the first target only, but with a temporal gap of 160 ms instead of 100 ms, to enable comparison with the two-fringes/160-ms condition.

4.5.2 Results

The normalized results for two fringes are shown by circular symbols in Fig. 4.3. The triangle symbols indicate the additional condition with only one fringe added to the end of the first target with a 160-ms temporal gap. The abscissa shows the gap duration in ms, and the ordinate shows the effect. The label **F** on the abscissa indicates performance for the condition with a temporal gap of 0 ms, which was the baseline condition with a fringe from the experiment described in section 4.4. An effect of *one* indicates that performance was the same as in the baseline condition containing only the targets. An effect of *zero* indicates that performance was the same as in the baseline condition containing a fringe without fringe alteration (cf. section 4.2). The error bars indicate the standard error of the mean.

Both the individual results, as well as the across-subject means show an increased effect with increasing gap duration. Pairwise t-tests on the across-subject data of the 0-ms gap condition (baseline condition **F**) with each of the other conditions showed that the conditions with a gap duration of 80 ms or longer were significantly different from the 0-ms condition at a 5 % significance level with Bonferroni correction. The size of the effect ranged from 0.27 at a gap duration of 80 ms to 0.68 at a gap duration of 320 ms. Listeners performed significantly worse for the extra condition where only one fringe was added with a temporal gap of 160 ms ($\mathbf{160}_{\text{one fringe}}$ in Table 4.2, or triangles in Fig. 4.3) compared to the original 160-ms gap condition (**160**). The difference in effect was 0.24 (see triangle symbols in Fig. 4.3). It must be noted though that there were large across-subject differences, e.g., subject S3 performed the same for the two conditions while subject S2 performed lower than his individual baseline condition (hence the negative effect). Condition $\mathbf{160}_{\text{one fringe}}$ did not differ significantly from baseline condition **F**. The results of this pairwise t-tests are shown in Table 4.2.

Table 4.2: Effect of gap duration: pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
20	**F**	-0.05	66	0.962	No
40	**F**	1.21	66	0.232	No
80	**F**	3.24	66	0.002	Yes
160	**F**	4.38	66	<0.001	Yes
320	**F**	8.06	66	<0.001	Yes
$\mathbf{160}_{\text{one fringe}}$	**F**	1.58	66	0.120	No
160	$\mathbf{160}_{\text{one fringe}}$	2.81	66	0.007	Yes

4.5.3 Discussion

The increased performance for target discrimination when a gap of 80 ms or more was introduced between target and fringe is not in line with the explanation of interference due to the mere presence of new stimulus information, because the amount of interference would be expected to be considerable for all temporal gap conditions.

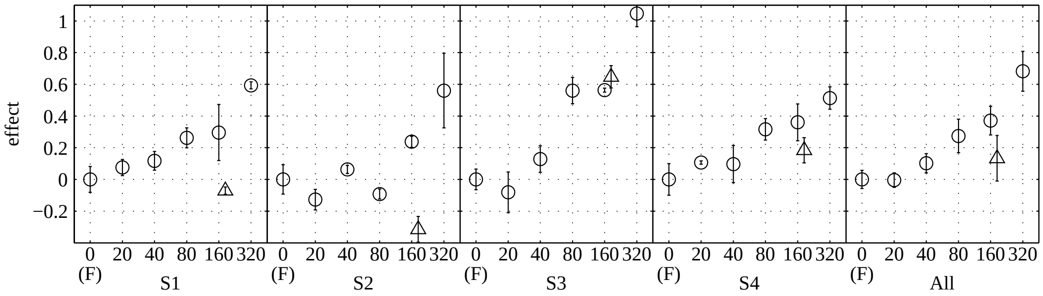

Figure 4.3: Effect of gap duration. Mean effects of individual subjects are shown in the left four panels, and panel five shows the mean effect. Label **F** on the abscissa indicates the baseline condition, i.e., without temporal gap. The other labels on the abscissa give the gap duration in ms. The circles indicate the conditions with a fringe added to both targets, the triangles indicate the conditions where only one fringe was added to the end of the first target token. The error bars indicate the standard error of the mean.

The increased performance is, however, in line with an object-based explanation in which the temporal separation enables the listener to segregate the target and fringe into separate auditory objects.

In broadband noise, a gap is detectable for gap durations as short as approximately 3 ms (Penner, 1977). For 600-Hz wide noise bands, thresholds are about 10 ms (Eddins *et al.*, 1992). Thus, the mere presence of a perceivable temporal gap was not sufficient for the segregation of target and fringe. In the case of a 20-ms gap, i.e., well above this threshold, the presence of the gap did not lead to an improvement of discrimination ability, maybe because the gap was perceived as a feature of the auditory object containing the target and fringe rather than as a cue for segregation.

Interestingly, the presence of two fringes instead of only a single fringe after the first interval (Hanna, 1984) led to an improvement in performance. One interpretation is that even with a 180-ms gap, the target and fringe are not perceived independently, adding fringes in both intervals helps to improve consistency between the first and second interval.

4.6 Experiment 3: Spectral separation

In the previous experiment the *temporal* separation between target and fringe was shown to have a large effect on the ability to discriminate the targets. Another dimension in which the target and fringe can be separated is the spectral dimension. In the next experiment a *spectral* separation between target and fringe was introduced to determine its influence on the ability to discriminate the target.

4.6.1 Stimuli

The properties of the targets and fringes were as specified in Sec. 4.3, except that, the spectral bandpass-range of the fringes was varied. In each condition the bandwidth of the fringes was constant on the ERB$_N$-number scale, (Glasberg and Moore, 1990), ERB width = 5.8, but the center frequencies were distributed such that there were five consecutive bandpass ranges. The bandpass ranges were 81–349 Hz (condition **-1**), 350–850 Hz (baseline condition **F**), 851–1783 Hz (condition **1**), 1783–3520 (condition **2**), and 3520–6757 (condition **3**). The bandpass range of the targets was always 350–850 Hz.

4.6.2 Results

Figure 4.4 shows the results. The labels on the abscissa indicate the separation between the center frequencies of target and fringe. The label **F** indicates the baseline condition with fringe (from the experiment described in section 4.4), which had no spectral separation between target and fringe. The ordinate shows the effect of the fringe manipulation with respect to the baseline conditions. Panels show the individual and mean results. The error bars indicate the standard error of the mean.

Pairwise t-tests on the across-subject data for the baseline condition **F** with each of the other conditions showed that all conditions were significantly different from condition **F** at a 5 % significance level with Bonferroni correction. The results of these pairwise t-tests are shown in Table 4.3. On average, as the spectral separation between target and fringe increased, the ability to discriminate increased. For the largest spectral separation (labeled **3**), three out of four listeners showed an effect of more than .8 which is close to performance for the basline condition without a fringe. The size of the effect for a spectral separation of one (labeled **-1** and **1**) was 0.40.

Table 4.3: Effect of spectral gap: pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
-1	**F**	5.38	44	<0.001	Yes
1	**F**	5.25	44	<0.001	Yes
2	**F**	9.08	44	<0.001	Yes
3	**F**	9.63	44	<0.001	Yes

4.6.3 Discussion

This experiment showed that spectral separation also had a large effect on the ability to discriminate the target noises. The ability to discriminate the targets increased with spectral distance. For three out of four subjects, an effect of more than 0.8 was observed, which was almost the same as performance for the baseline condition where the targets were present without fringe. Spectral separation seems to be a salient cue for enabling the listener to access information from the target.

This section addressed situations where the target and the fringe had disjunct spectral ranges. From a gestalt theory point of view, it is likely that the targets and fringes originated from separate physical events since their spectra did not adhere to the rule of good continuity. More specifically, the end of the target within one spectral range and the start of the fringe in a completely disjunct region is difficult to reconcile with good continuity. The next section addresses the situation where the spectral ranges of target and fringe fully overlap, and thus part of the stimulus adheres to the rule of good continuity.

4.7 Experiment 4: Bandwidth

In this experiment the center frequencies of the target and fringe spectra were identical. However, the *bandwidth* of the fringes was either double or half the bandwidth of the targets. Hence, the spectral ranges of the targets and the fringes were not disjunct but fully overlapping, allowing for the interpretation that the stimulus part with a narrower bandwidth continues in the part with a wider bandwidth.

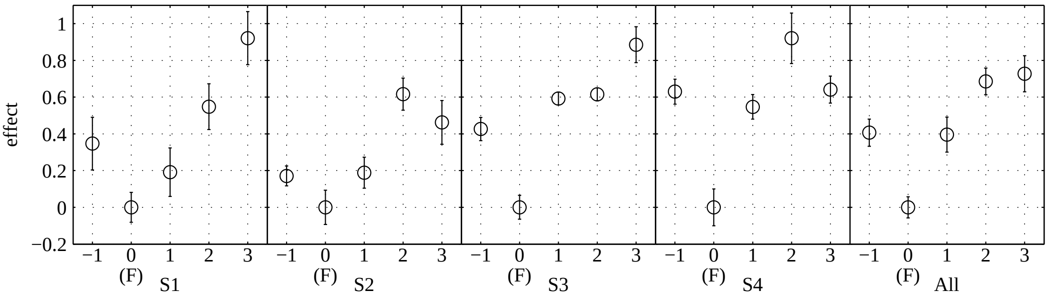

Figure 4.4: Effect of spectral separation. Mean effects of individual subjects and the across-subject mean effect are shown. Label **F** on the abscissa indicates the baseline condition, i.e., without spectral gap. The other labels on the abscissa indicate the spectral separation between the center frequencies of target and fringe in number of bandwidths (5.8 ERB_N). The error bars indicate the standard error of the mean.

4.7.1 Stimuli

The properties of the targets and fringes were as specified in Sec. 4.3. Except that the bandpass range of the fringes was either doubled to a bandpass range of 100–1100 Hz, or halved to bandpass range of 475–725 Hz. The center frequency was 600 Hz, which was the same as the center frequency of the targets. In addition, these conditions were presented with and without a temporal gap of 40 ms.

4.7.2 Results

Figure 4.5 shows the results. The ordinate shows the effect of the fringe manipulations with respect to the baseline conditions. Individual and mean results are shown. The error bars indicate the standard error of the mean.

In the conditions labeled with **2**, for which the bandwidth was doubled compared to the target bandwidth, the bandpass range was 100–1100 Hz. In the conditions labeled with **1/2**, for which the bandwidth was halved, the bandpass range was 475–725 Hz. The conditions which are labeled with subscript 40 had a temporal gap between target and fringe with a duration of 40 ms in addition to the bandwidth change. For comparison, the data for the 40-ms gap condition from the experiment described in section 4.5 were included in the figure with the label $\mathbf{1}_{40}$.

Table 4.4 shows the results of pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$ for several pairs of the bandwidth data. Comparing baseline condition **F** (including a fringe without fringe alteration) with each of the other conditions showed that the conditions where the fringe bandwidth was half the target bandwidth ($\mathbf{1/2}$ and $\mathbf{1/2}_{40}$) were significantly different from condition **F** (size of effect was 0.21 and 0.40, respectively). Other significantly different pairs were condition $\mathbf{1/2}$ compared to $\mathbf{1/2}_{40}$ (difference of effect size was 0.19) and condition $\mathbf{1/2}_{40}$ compared to $\mathbf{1}_{40}$ (difference of effect size was 0.30). The conditions where the fringe bandwidth was double the target bandwidth (**2** and $\mathbf{2}_{40}$) were not significantly different from condition **F** nor from each other. Condition $\mathbf{2}_{40}$ was also not significantly different from condition $\mathbf{1}_{40}$.

Table 4.4: Effect of bandwidth: pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
2	**F**	-1.66	55	0.102	No
$\mathbf{2}_{40}$	**F**	-0.69	55	0.494	No
$\mathbf{1/2}$	**F**	3.64	55	0.001	Yes
$\mathbf{1/2}_{40}$	**F**	6.86	55	<0.001	Yes
$\mathbf{2}_{40}$	**2**	0.98	55	0.333	No
$\mathbf{1/2}_{40}$	$\mathbf{1/2}$	3.23	55	0.002	Yes
$\mathbf{2}_{40}$	$\mathbf{1}_{40}$	-2.42	55	0.019	No
$\mathbf{1/2}_{40}$	$\mathbf{1}_{40}$	5.14	55	<0.001	Yes

4.7.3 Discussion

This experiment showed that doubling the bandwidth of the fringe with respect to the target bandwidth did not result in an improvement in discrimination of the target token, even when combining this cue with a 40-ms temporal gap. Halving the bandwidth, however, resulted in a modest but significant effect of approximately 0.2. Combined with a 40-ms temporal gap this effect was 0.4. Thus, the combination of a temporal gap with half bandwidth caused an extra improvement in effect of 0.2.

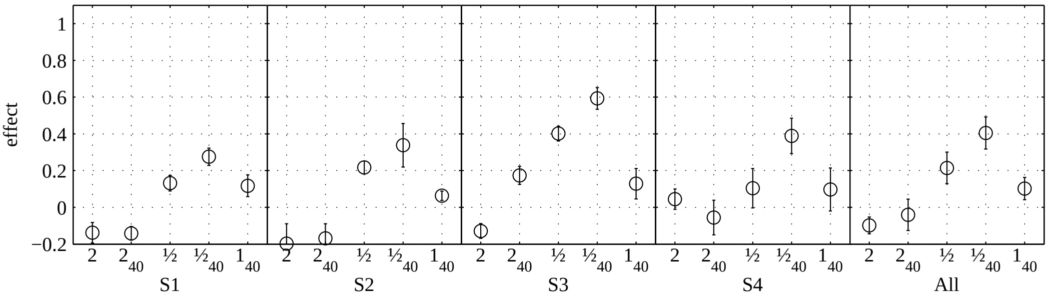

Figure 4.5: Effect of bandwidth.Mean effects of individual subjects and the across-subject mean effect are shown. Label **2** on the abscissa indicates the conditions where the fringe bandwidth was doubled to 1000 Hz. Label **1/2** on the abscissa indicates the conditions where the fringe bandwidth was halved to 250 Hz. The conditions with subscript 40 had an additional temporal gap between target and fringe. For comparison, the temporal gap condition with a duration of 40 ms and the same bandwidth as the target was included ($\mathbf{1}_{40}$). The error bars indicate the standard error of the mean.

This improvement was more than the effect of the temporal gap in isolation, which was a non-significant effect of 0.1.

4.8 Experiment 5: Level and ILD

The next experiment investigated if changing the level or the Interaural Level Difference (ILD) of the fringes leads to better discrimination of the targets.

4.8.1 Stimuli

The properties of the targets and fringes were as specified in Sec. 4.3, except that, the spectrum level of the fringes was varied. In one condition, the spectrum level of the fringe was 55 dB, i.e., 5 dB higher than for the target. In another condition, the spectrum level of the fringe was 45 dB, i.e., 5 dB lower than for the target. In a third condition the spectrum level of the fringe was 45 dB in the left ear and 55 dB in the right ear, realizing a 10-dB Interaural Level Difference (ILD). This condition was also presented with a temporal gap of 40 ms. A final condition had a monaural fringe that was presented only in the right ear with a spectrum level of 50 dB.

4.8.2 Results

The results of the level conditions are shown in Fig. 4.6. The ordinate shows the effect of the fringe manipulations with respect to the baseline conditions. The error bars indicate the standard error of the mean.

The conditions where the fringe level was 5 dB higher than the target level are labeled $+\mathbf{5}$, and the conditions where it was 5 dB lower are labeled with $-\mathbf{5}$. The condition with a 10-dB ILD is labeled $\pm\mathbf{5}$ and when this condition included a 40-ms gap it is labeled $\pm\mathbf{5}_{40}$. The condition with the monaural fringe is labeled **M**. For comparison, the 400-ms gap condition from the experiment described in section 4.5 is included in the figure with the label $\mathbf{0}_{40}$.

Table 4.5 shows the results of pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$ for several pairs of the level and ILD data. Comparing baseline condition **F** with each of the other conditions showed that only the monaural fringe condition (**M**) was significantly different from condition **F** (size of the effect was 0.35). Performance for the condition where the fringe had a combination of a 10-dB ILD with a 40-ms temporal gap ($\pm\mathbf{5}_{40}$) was not significantly different from performance for the condition where the fringe had only a 40-ms temporal gap ($\mathbf{0}_{40}$).

Table 4.5: Effect of level and ILD: pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
$+\mathbf{5}$	**F**	-1.00	66	0.319	No
-5	**F**	1.60	66	0.115	No
$\pm\mathbf{5}$	**F**	2.10	66	0.039	No
$\pm\mathbf{5}_{40}$	$\mathbf{0}_{40}$	-0.01	66	0.992	No
$\mathbf{0}_{40}$	**F**	1.46	66	0.148	No
M	**F**	5.01	66	<0.001	Yes

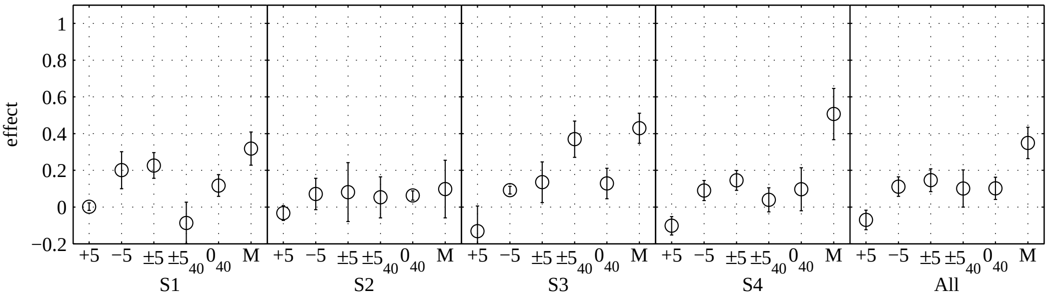

Figure 4.6: Effect of level and ILD. Mean effects of individual subjects and the across subject mean effect are shown. Label **+5** on the abscissa indicates the condition where the spectrum level of the fringe increased by 5 dB. Label **-5** indicates the condition where the spectrum level of the fringe decreased by 5 dB. Label **±5** indicates the conditions where the fringe was lateralized to the right with an ILD of 10 dB. The conditions with subscript 40 had an additional temporal gap between target and fringe. For comparison, the temporal gap condition with a duration of 40 ms and same level as the target was included with label $\mathbf{0}_{40}$. Label **M** indicates the monaural condition where the fringe was presented only to the right ear. The error bars indicate the standard error of the mean.

4.8.3 Discussion

The level and ILD conditions only showed a significant effect for the condition with a monaural fringe. Apparently the influence of level cues and ILD cues in the fringe on the ability to discriminate the targets was very small or non-existent, except for the extreme ILD case where the fringe was presented to only one ear.

In the individual data it is striking that subject S1 performed more poorly for the $\pm 5_{40}$ condition containing the combination of a 40-ms temporal gap and a 10-dB ILD than for the 40-ms temporal gap condition and the 10-dB ILD condition in isolation. On the other hand, these combined cues seemed to have an additive effect on the performance of subject S3. For the other two subjects, there was little difference between these three conditions. This may hint at individual differences in listeners' ability to use the cues in the discrimination task.

Overall, presenting the fringe monaurally led to the best performance in this experiment, possibly because it constitutes the largest ILD. An alternative explanation is that listeners may have listened to the ear without the fringe in the monaural condition.

4.9 Experiment 6: Interaural time delay

The previous experiment showed that only for the monaural condition we could find a statistically significant improvement for discrimination of the targets. In the next experiment the fringes were lateralized using an Interaural Time Delay (ITD)

4.9.1 Stimuli

The properties of the targets and fringes were as specified in Sec. 4.3. Except that a fine structure ITD of 0.5 ms was applied to the fringes before the 10-ms onset and offset ramps were applied causing a lateralization to the right. Therefore, the temporal envelopes of the fringes were not delayed (only ongoing ITDs). In additional conditions, this ITD was combined with a 5-dB increase in spectrum level, a 5-dB decrease in spectrum level, or a temporal gap of 40 ms.

4.9.2 Results

The results are shown in Fig. 4.7. The ordinate shows the effect of the fringe manipulations with respect to the baseline conditions. The conditions with an ITD of 0.5 ms are labeled **.5**. The condition with a 5 dB spectrum level increase is labeled $\mathbf{.5}_{+5}$ and the condition with a 5 dB spectrum level decrease is labeled $\mathbf{.5}_{-5}$. The condition with a temporal gap of 40 ms is labeled $\mathbf{.5}_{40}$. For comparison the 40-ms gap, the 5-dB increase, and the 5-dB decrease conditions from the experiments described in sections 4.5 and 4.8 are included in the figure using triangle symbols.

Table 4.6 shows the results of pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$ on several pairs of the ITD data. Baseline condition **F** was significantly different from the ITD-only condition **.5** (size of effect was 0.17). Comparison of the three ITD conditions which were combined with level cues or a temporal gap of 40 ms ($\mathbf{.5}_{+5}$, $\mathbf{.5}_{-5}$, and $\mathbf{.5}_{40}$) to the ITD-only condition (**.5**) showed that only the condition where the fringe had an ITD plus a 5-dB level decrease ($\mathbf{.5}_{-5}$) was significantly different from the ITD-only condition (**.5**) (difference in effect size was 0.19). In addition, performance for these three combined conditions was significantly higher than when they were presented without an ITD ($\text{effect}_{.5_{+5}} - \text{effect}_{+5} = 0.25$, $\text{effect}_{.5_{-5}} - \text{effect}_{-5} = 0.25$, $\text{effect}_{.5_{40}} - \text{effect}_{40} = 0.18$).

Table 4.6: Effect of ITD: pairwise t-tests with Bonferroni correction with a significance level of $\alpha = 0.05$. The table also denotes the value of the test statistic (t), the degrees of freedom of the test (df) and the p value (p)

Condition	Condition	t	df	p	Significantly different
.5	**F**	2.83	77	0.006	Yes
$\mathbf{.5}_{+5}$	**.5**	0.12	77	0.909	No
$\mathbf{.5}_{-5}$	**.5**	3.09	77	0.003	Yes
$\mathbf{.5}_{40}$	**.5**	1.72	77	0.090	No
$\mathbf{.5}_{+5}$	**+5**	4.07	77	<0.001	Yes
$\mathbf{.5}_{-5}$	**-5**	4.11	77	0.001	Yes
$\mathbf{.5}_{40}$	**40**	2.89	77	0.005	Yes

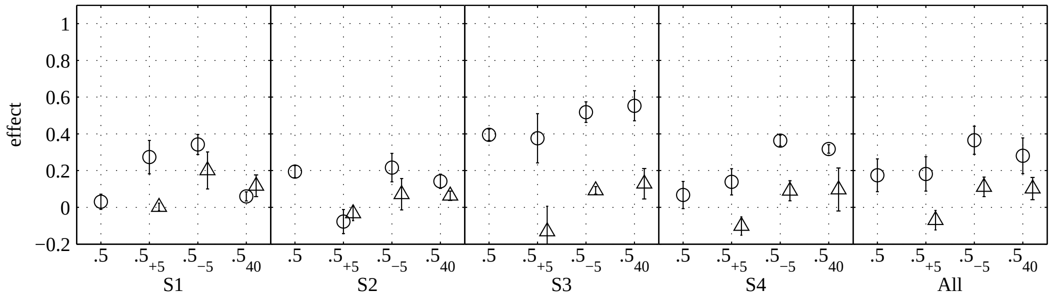

Figure 4.7: Effect of ITD. Mean effects of individual subjects and the across subject mean effect are shown. The *circles* indicate the conditions with an ITD of 0.5 ms. The *triangles* show the same conditions but *without* an ITD from the experiments in sections 4.8 and 4.5. The subscript +5 indicates the conditions with a spectrum level increase of 5 dB. The condition with subscript -5 had a spectrum level decrease of 5 dB. The condition with subscript 40 had a temporal gap between target and fringe. The error bars indicate the standard error of the mean.

4.9.3 Discussion

This experiment showed that discrimination of the target could be significantly improved by giving the fringe an ITD of .5 ms, although the size of the effect was modest (0.17). Listeners' ability to discriminate the target improved further, with an *added* effect of 0.19, by combining the ITD with a level decrease of 5 dB. Combining the ITD with a level increase of 5 dB or a temporal gap of 40 ms did not further improve the ability to discriminate the targets. Discrimination ability in these combined conditions, however, was significantly higher than in the corresponding conditions without an ITD.

4.10 General discussion

Hanna (1984) and the research presented in chapter 3 showed that discrimination performance for noise tokens is substantially impaired when a noise fringe (a piece of uninformative noise) is placed either before or after *one* of the target tokens, while the same amount of useful information was present. The current study showed that this impairment can be strongly reduced by introducing cues into the fringe that may help to segregate the fringe and the target.

The ability to discriminate Gaussian-noise target tokens in a same/different discrimination paradigm was further explored. Of particular interest was the effect of adding an uninformative backward-fringe to *both* 50-ms target tokens. This fringe had the same statistical properties as the target and its duration was 200 ms. Again, while the same amount of useful information was present, the ability to discriminate the targets decreased substantially when a fringe was added to the targets without fringe alteration. If, however, the properties of the fringes were altered, the influence of the fringes on the ability to discriminate the targets could be reduced and, hence, discrimination performance for the targets could improve. The type of fringe alteration determined whether discrimination performance improved and the size of the effect.

The largest effects were found when a spectral or temporal separation was introduced between target and fringe. In some of these conditions, some listeners improved their performance to nearly the discrimination performance that was achieved in absence of a fringe. Modest effects were found for halving the bandwidth of the fringe, for introducing a 0.5-ms ITD, or presenting the fringe to only one ear. No effect was found for doubling the fringe bandwidth, 5-dB level increases or decreases, or 10 dB ILDs. However, combining a 5-dB level decrease with a .5-ms ITD resulted in an additional modest improvement of discrimination ability. Similarly, combining a 40-ms temporal gap with halving the fringe bandwidth also resulted in an additional modest improvement of discrimination ability.

Apparently listeners were able to ignore the uninformative stimulus details in the fringe when it was sufficiently distinguishable from the target. The parameters chosen to make the fringes distinguishable from the targets were known from the literature to contribute to the formation of auditory objects (Yost, 1991). The observation that they also contribute to improved performance for target discrimination supports the hypothesis that auditory object formation leads to better ability to attend to only the noise targets and ignore the fringes. Some fringe alterations (level increase of 5 dB, level decrease of 5 dB, 10 dB ILD, and doubling the fringe bandwidth) did not lead to improved discrimination of the target tokens. This does not necessarily imply that these cues cannot lead to object formation, however, their effectiveness in creating separate auditory objects may be less than for cues that did lead to improved target discrimination in these experiments.

In this light, the decreased discrimination performance of target tokens with an added fringe without fringe alteration, can be explained by assuming that listeners perceived the target plus fringe as a single auditory object and that listeners had a fixed or limited capacity to process or retain an auditory object (Cowan (2005) and the study described in chapter 3 of this thesis). In addition, it is assumed that listeners are not able to selectively attend to a subsection of such an auditory object (chapter 3 of this thesis). In the case of a fringed target, the limited resources for retaining or processing the auditory objects need to be distributed over more peripheral information than in the case of a target in isolation. Therefore, the memory or processing for the fringed target is degraded, and discrimination performance is lower.

It has been argued that sensory memory is rich and, in contrast to working memory, perhaps of unlimited capacity (Cowan, 1988, 2005). Therefore, in the case of a discrimination trial where a fringe is appended to the target tokens, the information useful for the discrimination task should be available in sensory memory. In the framework of Cowan (2005), the inability to perform the discrimination when a fringe is appended without a differentiating cue indicates that the *access* to this information is limited. According to Cowan (2005), it is the focus of attention, enabling the listener to draw information from sensory memory into working memory, that is limited. When the fringe is differentiated from the target by introducing a cue, access to the target information is made possible. Arguably, the fringe alteration enables the listener to direct the focus of attention to the target information within the auditory stimulus containing both target and fringe. It is plausible that the segregation of target and fringe into two *separate* objects, enables the listener to direct the focus of attention to the target object.

This auditory object view therefore provides an explanation for the effect of introducing fringe alterations on discrimination. When a fringe alteration is introduced that helps to segregate target and fringe into two separate auditory objects, the information in the fringe can be better ignored, and hence, discrimination performance is affected less by the presence of the fringe. According to this view, depending on how well listeners are able to segregate the target and fringe, the ability to (partly) ignore the fringe increases, and hence, the ability to discriminate the target token also increases.

When the results of the experiments are interpreted within this framework, it can be concluded that both spectral and temporal gaps are strong cues for formation of auditory objects. For temporal gaps a similar observation was made by Bregman (1990, pg. 71). In a paragraph, describing some of the experiments by van Noorden (1975), he stated that

> "Apparently, abrupt rises in intensity tell the auditory system that a new sound has joined the mixture and that it should begin the analysis of a new unit."

In the experiments on auditory streaming by van Noorden (1975) it was shown that spectral separation was an important factor in creating segregation in repeated alternating tone-patterns.

In section 4.7, the bandwidth of the fringes was, depending on the condition, either double or half the bandwidth of the targets. Thus, their spectra overlapped. The interpretation of these overlapping spectra is ambiguous. For instance, in the condition where the fringe bandwidth doubled, an interpretation of the stimuli is that a new auditory event, the fringe, was presented exactly at the offset of the target. Another interpretation is that the target continues and two flanking noise bands are presented 50 ms after the onset of the target. When the bandwidth of the fringe is halved, a similar line of reasoning could be followed. This ambiguity is a possible explanation why there was no advantage for the bandwidth conditions like there was for the spectral separation conditions.

Throughout the experiments in this study, fringes were added to both targets, except in one the two of the baseline conditions where there were no fringes. Another exception was the extra condition in the gap-duration experiments (section 4.5), which replicated a condition of the experiments of Hanna (1984). Here, a fringe was added only to the first target with a temporal gap of 160 ms. On average, performance was worse in the situation with only one fringe compared to the situation with two fringes, although the amount of uninformative stimulus detail was less. Discrimination performance in the two-fringes condition was improved by enlarging the gap from 160 to 320 ms. Apparently the fringes were still influencing discrimination performance for the targets in the condition with a temporal gap of 160 ms . When a fringe was appended to only the first target, the perception of the target was influenced by the presence of the fringe. However, no fringe was added to the second target, and therefore, the perception of the second target was not influenced by a fringe. One interpretation is that adding a fringe to only one target may introduce an asymmetry in the perception of the target tokens, causing performance to be worse than in the two-fringes situation. Possibly, when not sufficiently segregated, the target and fringe are grouped into a larger object (Bregman, 1990, pg. 644).

In general, the results from this study suggest that segregation of auditory objects depends strongly on the presence of spectral and temporal separation between stimuli. This has been suggested many times, most notably by van Noorden (1975); Bregman (1990). Binaural cues can also lead to the segregation of auditory objects, but their importance as segregation cues seems lower than the aforementioned spectral and temporal separation. This is in line with the observation of Bregman (1990) that for the grouping of tones "Humans use spatial origin too, but do not assign such an overwhelming role to it".

In a vowel identification experiment, Drennan *et al.* (2003) found that both ITDs and ILDs can play a role in segregation, but ILDs have a larger impact on segregation (of vowels) than ITDs. This order of impact found for ITDs and ILDs is opposite to our findings. A possible explanation for this difference is that their stimuli extended up to 2 kHz whereas our stimuli for the binaural conditions extended only up to 1.1 kHz. For natural stimuli it is known that ITDs mainly contribute to localization in the low frequency region (below approximately 1500 Hz), and in the high frequency region lateralization is mainly realized through ILDs, although at low frequencies ILDs can be used when present (Grantham, 1995; Moore, 2003). The different frequency content in both experiments may have tipped the balance in favor of the ITDs in the current study.

In contrast to many of the existing studies of object formation, our Gaussian noise stimulus did not include a long succession of auditory stimuli (like, e.g., van Noorden, 1975; Royer and Robin, 1986; Crum and Bregman, 2006). As such, this provides a different approach for investigating the formation of auditory objects, with the advantage that auditory objects and auditory streams cannot be confused in the interpretation of the results. In addition, with this method, object-related discrimination performance is investigated in a way that does not rely on subjective judgments of the listeners that are not quantitatively verifiable, e.g., about the number of auditory streams the listener perceives.

5 Conclusions

Previous research (Hanna, 1984; Heller and Trahiotis, 1995), as well as the research presented in the chapter 2 of this thesis, established a nonmonotonic relationship between the ability to discriminate Gaussian-noise tokens and the duration of these tokens, with maximum performance at around 40 ms. Such a result would not be expected if listeners used all available peripheral information for the discrimination task. In this case, performance would be expected to increase monotonically (for a discussion of information in perception, see chapter 1). A set of experiments using stochastic stimuli where stimulus duration, bandwidth, and number of degrees of freedom were varied independently, suggested that:

- Discrimination performance for these stimuli depends predominantly on the amount of peripheral information of an auditory object and the capacity to process this peripheral information. This capacity seems to be limited in the temporal dimension. (chapter 2).

Current psychoacoustic models (e.g., Viemeister and Wakefield, 1991; Dau *et al.*, 1996) which optimally combine all peripheral information over time do not predict the nonmonotonic duration dependence. Instead, predicted performance increases with duration until a ceiling performance has been reached.

In chapter 3, it was shown that the nonmonotonic duration dependence can be successfully simulated by adding a new stage to the model of Dau *et al.* (1996). The new stage imposes a limit on the amount of peripheral information within each critical band of the internal representation of a stimulus. From the modeling results and the results from chapter 2 it was inferred that:

- The nonmonotonic duration dependence can be attributed to a limited capacity for retaining or processing peripheral information about auditory stimuli (chapter 3).

The model of Dau *et al.* (1996) was designed to model phenomena where the internal representation changes in a consistent, predictable way. The modified model deals with discrimination where changes in internal representation are unpredictable. Hence, useful templates cannot be built up across trials. Possibly the fact that no expectation about differences can be built up over trials imposes a severe limitation on the capacity to process or retain peripheral information for discriminating intervals within a single trial. In order to be able to perform the noise-discrimination task, the template matching and optimal filtering stage of the model of Dau *et al.* (1996) has been replaced by a new decision device in our model (chapter 3).

In the above modeling approach, auditory stimuli are interpreted holistically as auditory objects. The model's limited resources are distributed evenly over the whole object. These modeling assumptions were tested with additional listening experiments and model simulations, using stimuli consisting of a to-be-discriminated target noise and an uninformative noise fringe. The results of these listening tests are consistent with the predictions of the proposed model that is based on the assumption that:

- A limited processing capacity is allocated to auditory objects, and these objects seem inseparable in the sense that it is not possible for listeners to selectively listen to only a part of the object (chapter 3).

When temporal, spectral, or binaural cues were provided in the fringe, the ability to discriminate the targets increased. Sometimes subjects reached the same performance as was observed when no fringe was present. This is in line with the hypothesis that the processing or memory capacity for an auditory object is fixed. When the target and fringe were perceptually segregated, the fringe could be (partly) ignored. As a consequence, the processing capacity could be attributed to the target alone and discrimination performance increased. Following this line of reasoning the experiments suggest that:

- Spectral separation and temporal separation are the strongest cues for auditory-object formation of Gaussian noise, (chapter 4).

Modest effects were found for interaural time delay conditions, for halving the bandwidth of the fringe, and for monaural fringes. No significant effect was found for doubling the fringe bandwidth, level cues of ±5 dB, or 10 dB interaural level differences. This method thus may provide a new approach for investigating the formation of auditory objects.

5.1 Practical relevance

In speech and audio coding, noisy parts of the sounds are often expensive to encode in terms of bitrate when it is attempted to code the exact waveform. In *speech* coding, reproducing the exact noise waveform is of relatively less importance because a signal processing model of the vocal chords and the vocal tract is used for voiced parts. A periodic excitation signal models the vibrations of the vocal chords. This part is responsible for the harmonic speech components. For unvoiced speech, e.g., the turbulence of the air at the lips of the speaker, a noisy excitation signal is used. The noisy components are encoded using a waveform matching algorithm (e.g., using a codebook), which allows approximation of the noise at the receiver side (cf. Sluijter, 2005). The excitations are spectrally shaped by a model of the vocal tract. The waveform at the receiver side is in general not identical to the waveform at the transmitter side. This is also not necessary as long as they are perceptually equivalent.

In *audio* coding a similar technique is applied, which is known as Perceptual Noise Substitution (PNS) and is part of the Advanced Audio Coding (AAC) standard (Herre and Schulz, 1998). For audio signals there is no simple excitation model of the source. Instead, this technique makes use of a noise detection stage to find the areas in a spectro-temporal representation of the sound where the signal is noise-like. The noise in these areas can in many circumstances be replaced by parametrically defined noise. This can result in a great reduction of bitrate since only a few parameters need to be transmitted instead of a whole waveform. However, sometimes the substituted noise appears not to be perceptually equivalent to the original noise. Therefore, Skowronek and van de Par (2004) used a perceptual model to evaluate the perceptual distance between the original noise and a random substitute. The original noise was only substituted when, according to the model, the noise substitution was inaudible. However, they did not take into account the role of object formation in the evaluation of the perceptual distance between the original noise and the synthetic substitute. The automatic substitution of noise in audio or speech signals may still be improved when knowledge about the perception of noise objects, such as presented in this thesis, is applied.

For coding, one challenge is to determine what the auditory objects are in a sound mixture. Until this goal has been reached, recommendations for perceptual noise substitution are to substitute mainly noise that is of high frequency, or to make sure that the duration of the substituted noise is either sufficiently long (> 100 ms) or sufficiently short (< 20 ms) to make it indistinguishable from the original noise.

5.2 Suggestions for further investigations

In chapter 2, it was argued that the ability to discriminate auditory stimuli depends predominantly on the amount of peripheral information of an auditory object and the capacity to process this peripheral information. This was studied using Gaussian noise, and using random tone-burst patterns. In these random tone-burst patterns, the duration and the degrees of freedom were decoupled. It would be relevant to further investigate this number of degrees of freedom dependence using other types of stimuli in which the duration and the degrees of freedom are decoupled to verify to what extent this upper limit is an absolute limit. This may give insight into the limited processing or memory capacity.

Another way to investigate the limited capacity is to vary the number of different noise tokens in a block of same/different discrimination trials in a way resembling the absolute judgment experiments of Miller (1956). In chapter 2, it was found that the ability to discriminate is higher when only two (frozen) noise tokens are used throughout a block of 100 trials, than when two new tokens are generated for each trial (running noise). It would be interesting to investigate intermediate steps between the full frozen situation and the running situation. Increasing the number of different frozen noise tokens may result in a gradually decreasing ability to discriminate. Alternatively, the ability to discriminate may stay at approximately the same level for an increasing number of frozen noise tokens, as long as this number does not exceed the memory capacity of the listener. Beyond this critical point, when the maximum number of noise tokens is exceeded, the ability to discriminate could decrease more steeply. The number of tokens at which the critical point may emerge would provide insight in how many Gaussian-noise tokens can be remembered.

The new stage that was added to the model of Dau *et al.* (1996) in chapter 3, successfully simulated the nonmonotonic duration dependence found by Hanna (1984), Heller and Trahiotis (1995), and in chapter 2 of this thesis. However, it has yet to be investigated whether the adapted model is still able to account for the phenomena it was originally designed for, and if not, how the fixed information capacity strategy can be reconciled with the original purpose of the model. It is currently unclear how the modified model should cope with the experiments it was originally designed for. For instance, in a forward-masking experiment, where a to-be-detected tone burst is presented shortly after the offset of a noise masker, for what offset is the target a separate auditory object and how is the limited capacity assigned between target and masker? Especially for signal levels around the threshold level this may not be so clear. Although the adapted model proved to have explanatory value for a wide range of noise discrimination experiments, its value would be larger if it could be adapted such that it is still able to estimate thresholds in the conditions of its original purpose.

In the frozen noise experiments (chapter 2) and the repeated (cyclic) noise experiments of Kaernbach (1993), it was observed that listeners listen to certain features of the noise that become more salient when the same noise token is presented repeatedly. In another experiment, Kaernbach (1995) used a reverse correlation method (cf., de Boer, 1967) to obtain detailed spectro-temporal information about the features perceived in the noise. It is of interest to investigate whether the features that become more salient when the noise is presented more often are that same features the listeners perceive (less saliently) in the first presentation. If so, then the repeated noise literature could provide knowledge for the interpretation of the running noise discrimination experiments, and vice versa. Alternatively, a reverse correlation method could be designed for obtaining spectro-temporal information about noise that is not presented cyclically. It would provide supporting evidence for the model, presented in chapter 3, if it were shown that the temporal extent of the features perceived in a noise token appeared to be dependent on the total duration of the noise token.

Bibliography

Alain, C. and Arnott, S. R. (**2000**), "Selective attending to auditory objects," Front. Biosci. **5**, 202–212. 1, 4, 70

Bregman, A. S. (**1990**), *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT press, Cambridge). 1, 4, 66, 70, 93, 94

Broadbent, D. E. (**1958**), *Perception & Communication* (Pergamon, New York). 6

Buell, T. N. and Hafter, E. R. (**1991**), "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900. 71, 72

Buus, S. (**1990**), "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654. 7, 33

Clément, S., Demany, L., and Semal, C. (**1999**), "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811. 6

Coble, S. F. and Robinson, D. E. (**1992**), "Discriminability of bursts of reproducible noise," J. Acoust. Soc. Am. **92**, 2630–2635. 19, 22, 41, 60

Cowan, N. (**1984**), "On short and long auditory stores," Psychol. Bull. **96**, 341–370. 5

Cowan, N. (**1988**), "Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system," Psychol. Bull. **104**, 163–191. 92

Cowan, N. (**2001**), "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain. Sci. **24**, 87–185. 22, 66

Cowan, N. (**2005**), *Working Memory Capacity (Essays in Cognitive Psychology)* (Psychology Press, New York). 6, 7, 34, 66, 92

Crum, P. A. C. and Bregman, A. S. (**2006**), "Effects of unit formation on the perception of a changing sound," Q. J. Exp. Psychol. **59**, 543–556. 72, 94

Darwin, C. J. and Carlyon, R. P. (**1995**), "Auditory Grouping," in *Hearing*, edited by B. C. Moore (Academic Press, San Diego), pp. 287–424. 70

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**), "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622. 7, 8, 10, 32, 41, 42, 43, 46, 51, 65, 66, 67, 95, 96, 98, 108, 111

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**), "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760. 8

de Boer, E. (**1966**), "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552–560. 7

de Boer, E. (**1967**), "Correlation studies applied to the frequency resolution of the cochlea," J. Auditory Res. **7**, 209–217. 99

Delgutte, B. (**1987**), "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht), pp. 333–353. 7

Dennett, D. C. (**1997**), *Kinds of Minds* (Phoenix, London). 1

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**), "Perceptual segregation of competing speech sounds: The role of spatial location," The Journal of the Acoustical Society of America **114**, 2178–2189. 94

Durlach, N. I. and Braida, L. D. (**1969**), "Intensity Perception. I. Preliminary Theory of Intensity Resolution," J. Acoust. Soc. Am. **46**, 372–383. 5, 13

Durlach, N. I., Mason, C. R., Kidd Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**), "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987. 9, 33

Eddins, D. A., Joseph W. Hall, I., and Grose, J. H. (**1992**), "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. **91**, 1069–1077. 80

Fallon, S. M. (**1989**), "Discriminability of bursts of reproducible noise," Ph.D. thesis, Indiana University, Bloomington, Indiana. 33, 40, 41, 60, 61, 62, 63, 65, 66, 67, 108, 111

Glasberg, B. R. and Moore, B. C. J. (**1990**), "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138. 28, 43, 81

Grantham, D. W. (**1995**), "Spatial hearing and related phenomena," in *Hearing*, edited by B. C. Moore (Academic Press, San Diego), pp. 297–345. 94

Green, D. M. (**1992**), "On the similarity of two theories of comodulation masking release," J. Acoust. Soc. Am. **91**, 1769–1769. 47

Green, D. M. and Swets, J. A. (**1988/1966**), *Signal Detection Theory and Psychophysics* (Peninsula Publishing, Los Altos). 47

Guttman, N. and Julesz, B. (**1963**), "Lower limits of auditory periodicity analysis," J. Acoust. Soc. Am. **35**, 610. 3

Halford, G. S., Wilson, W. H., and Phillips, S. (**1998**), "Relational complexity metric is effective when assessments are based on actual cognitive processes," Behav. Brain. Sci. **21**, 848–860. 34

Hanna, T. E. (**1984**), "Discrimination of reproducible noise as a function of bandwidth and duration," Percept. Psychophys. **36**, 409–416. 3, 8, 10, 12, 13, 15, 16, 17, 23, 32, 40, 41, 48, 49, 51, 65, 71, 78, 80, 90, 93, 95, 99, 107, 108, 110, 111

Hansen, M. and Kollmeier, B. (**2000**), "Objective modeling of speech quality with a psychoacoustically validated auditory model," J. Audio Eng. Soc. **48**, 395–408. 47

Hartley, R. V. L. (**1928**), "Transmission of information," Bell Syst. Tech. J. **3**, 535–564. 2, 7, 27, 36, 51

Heller, L. M. and Trahiotis, C. (**1995**), "The discrimination of samples of noise in monotic, diotic, and dichotic conditions," J. Acoust. Soc. Am. **97**, 3775–3781. 12, 40, 41, 65, 71, 95, 99

Herre, J. and Schulz, D. (**1998**), "Extending the MPEG-4 AAC codec by perceptual noise substitution," 104th AES Convention, Amsterdam, the Netherlands , 1–14. 97

Huber, R. and Kollmeier, B. (**2006**), "PEMO-Q–A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," IEEE Trans. Audio Speech Lang. Process. **88**, 1902–1911. 47

Kaernbach, C. (**1993**), "Temporal and spectral basis of the features perceived in repeated noise," J. Acoust. Soc. Am. **94**, 91–97. 3, 34, 99

Kaernbach, C. (**1995**), "Early Auditory Feature Coding," in *Contributions to psychological acoustics: Results of the 8th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Meis, and C. Reckhardt (University of Oldenburg, Oldenburg), pp. 295–307. 99

Kaernbach, C. (**2004**), "Auditory Sensory Memory and Short-Term Memory," in *Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition*, edited by C. Kaernbach, E. Schröger, and H. Müller (Lawrence Erlbaum, Mahwah), pp. 331–348. 5

Kidd, G. R. and Watson, C. S. (**1992**), "The "proportion-of-the-total-duration rule" for the discrimination of auditory patterns," J. Acoust. Soc. Am. **92**, 3109–3118. 9, 31, 66, 67

Kidd Jr., G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2007**), "Informational Masking," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 143–190. 9

Köhler, W. (**1947**), *Gestalt Psychology* (Liveright, New York). 4

Kubovy, M. and Valkenburg, D. V. (**2001**), "Auditory and visual objects," Cognition **80**, 97–126. 4

Lutfi, R. A. (**1993**), "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758. 9, 67

Miller, G. A. (**1956**), "The magical number seven, plus or minus two: some limits on our capacity for processing information," Psychol. Rev. **63**, 81–97. 98

Moore, B. C. (**2003**), *An Introduction to the Psychology of Hearing* (Academic press, London). 7, 94

Näätänen, R. and Winkler, I. (**1999**), “The concept of auditory stimulus representation in cognitive neuroscience,” Psychol. Bull. **125**, 826–859. 5, 7, 33, 67

Nyquist, H. (**1928**), “Certain topics in telegraph transmission theory,” AIEE Trans. **47**, 617–644. 7, 27

Patterson, R., van Dinther, R., and Irino, T. (**2007**), “The robustness of bio-acoustic communication and the role of normalization,” in *19th International Congress on Acoustics Madrid*, edited by A. Calvo-Manzano, A. Pérez-López, and S. Santiago, pp. 1–6. 7

Penner, M. J. (**1977**), “Detection of temporal gaps in noise as a measure of the decay of auditory sensation,” J. Acoust. Soc. Am. **61**, 552–557. 80

Püschel, D. (**1988**), “Prinzipien der zeitlichen Analyse beim Hören,” Ph.D. thesis, University of Göttingen. 43

Rabiner, L. and Juang, B.-H. (**1993**), *Fundamentals of Speech Recognition* (Prentice Hall, Englewood Cliffs). 2

Rice, S. O. (**1944**), “Mathematical analysis of random noise,” AT&T Tech. J. **23**, 282–332. 2

Rickert, M. E. (**1998**), “Temporal and spectral effects in the auditory discrimination of Gaussian noise samples,” Ph.D. thesis, Indiana University, Ind. 33, 41

Royer, F. L. and Robin, D. A. (**1986**), “On the perceived unitization of repetitive auditory patterns,” Percept. Psychophys. **39**, 9–18. 71, 72, 94

Sams, M., Hari, R., Rif, J., and Knuutila, J. (**1993**), “The human auditory sensory memory trace persists about 10 sec: neuromagnetic evidence,” J. Cognitive Neurosci. **5**, 363–370. 5

Shannon, C. E. (**1948**), “A mathematical theory of communication,” Bell System Technical Journal **92**, 3109–3118. 2, 6, 40

Sheft, S. and Yost, W. (**2004**), "Minimum integration times for processing of amplitude modulation," in *Auditory signal processing: physiology, psychoacoustics, and models*, edited by D. Pressnitzer, A. de Chevegné, S. McAdams, and L. Collet (Springer-Verlag, New York), pp. 245–250. 12

Skowronek, J. and van de Par, S. (**2004**), "Automatic noise substitution in natural audio signals," in *Proceedings of the joint congress CFA|DAGA'04*, edited by D. Cassereau, pp. 1101–1102. 97

Sluijter, R. J. (**2005**), "The development of speech coding and the first standard coder for public mobile telephone," Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands. 97

Tchorz, J. and Kollmeier, B. (**1999**), "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050. 43, 47

van de Par, S. and Kohlrausch, A. (**1998**), "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579. 8

van den Brink, W. A. C. and Houtgast, T. (**1990**), "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711. 12

van Noorden, L. (**1975**), "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Technische Hogeschool Eindhoven. 70, 93, 94

Verhey, J. L., Rennies, J., and Ernst, S. M. (**2007**), "Influence of envelope distributions on signal detection," Acta Acoust. **93**, 115–121. 8

Viemeister, N. F. and Wakefield, G. H. (**1991**), "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865. 8, 32, 42, 65, 95

Warren, R., Bashford, J., Cooley, J., and Brubaker, B. (**2001**), "Detection of acoustic repetition for very long stochastic patterns," Percept. Psychophys. **63**, 175–182. 3

Watson, C. S. (**1987**), "Uncertainty, Informational Masking, and the Capacity of Immidiate Auditory Memory." in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277. 9, 33, 66

Watson, C. S., Foyle, D. C., and Kidd, G. R. (**1990**), “Limits of auditory pattern discrimination for patterns with various durations and numbers of components,” J. Acoust. Soc. Am. **88**, 2631–2638. 30, 40, 41

Winkler, I., van Zuijen, T. L., Sussman, E., Horváth, J., and Näänänen, R. (**2006**), “Object representation in the human auditory system,” Eur. J. Neurosci. **24**, 625–634. 4

Woods, W. S. and Colburn, H. S. (**1992**), “Test of a model of auditory object formation using intensity and interaural time difference discrimination,” The Journal of the Acoustical Society of America **91**, 2894–2902. 71, 72

Yost, W. (**1991**), “Auditory image perception and analysis: The basis for hearing,” Hear. Res. **56**, 8–18. 4, 71, 91

Yost, W. A. and Sheft, S. (**1993**), “Auditory Perception,” in *Human Psychophysics*, edited by W. A. Yost, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 193–236. 1, 70

Gaussian-noise discrimination and auditory object formation

Summary

Each day our hearing system is exposed to an enormous amount of auditory input. Arguably, it is one of the main tasks of the auditory system to extract only the useful information from the plethora of incoming stimuli. In an acoustic scene, not all sound producing events may be relevant to a listener. It is an advantage for her/him to ignore the irrelevant acoustic events and attend to only the useful ones. Also within the waveform originating from a single acoustic event, there may be stimulus features that are relevant, and others that are not relevant. For instance, for a Gaussian noise, the overall spectrum may be an important feature, whereas fine structure details may be of less importance.

Hanna (1984) has shown that the ability to discriminate broadband Gaussian-noise tokens reduces with increasing duration for stimuli longer than 100 ms, despite that the peripheral information increases, while below approximately 25 ms, the ability to discriminate increases with duration. Apparently, there is a nonmonotonic relationship between the amount of information elicited by the stimulus in the auditory periphery and the amount of perceptual information for this range of durations. One of the central goals of this study was to investigate the underlying mechanism responsible for this nonmonotonic relationship.

Chapter 2 describes the replication of one of the experiments of Hanna (1984), in which the nonmonotonic relationship between duration and discrimination ability was first shown for Gaussian noise. A similar non-monotonic duration dependency was found which had maximum performance around 40 ms. Additional experiments showed that listeners' performance could improve when the same noise tokens were used over all trials (frozen noise). However, the duration at which maximum perfor-

mance occurred did not change. In another experiment, using a stimulus consisting of 5-ms Hanning-windowed tone-bursts randomly distributed over time, it was investigated whether the roles of stimulus duration and amount of information independently affect the processing capacity of the auditory system. Results showed that the number-of-degrees-of-freedom in the stimulus, but not its duration, determined the ability to discriminate. Overall, the results presented in this chapter suggest that the ability to discriminate between acoustic stimuli depends highly on the amount of information of an auditory object, and the capacity to process this information. This capacity seems to be limited in the temporal dimension, while extending the signal over more auditory filters does have a positive effect on performance.

Models which combine all information from the auditory periphery over time will not correctly predict the nonmonotonic duration dependency. Instead, their discrimination performance will keep increasing with duration until it saturates at perfect performance. Chapter 3 describes a model, based on the existing model of Dau *et al.* (1996), that is able to predict the nonmonotonic duration dependency found by Hanna (1984) by limiting the information in the internal representations of a stimulus independent of the stimulus duration. This approach implies that stimulus intervals are treated as undividable auditory objects, and that the model has limited resources which are distributed evenly over the whole object. Therefore, this model does predictions about the inability of listeners to process only a limited part of the auditory object. These predictions were verified with behavioral experiments. In addition, the model was able to reproduce data concerning partially correlated noises from a study of Fallon (1989).

To impose restrictions on the amount of information allowed in the internal representation of an auditory object it is necessary to know where this object starts and where it ends. This is straightforward when the object is homogeneous and has a clear onset and offset, like Gaussian-noise bursts. However, when potential segregation cues such as temporal separation, spectral separation, bandwidth, level differences, interaural level differences, and interaural time delay are introduced in the stimulus, the formation of auditory objects may be influenced. Chapter 4 descibes a method to test the influence of such cues on object formation using a method inspired by the model predictions. The method makes use of the observation that listeners are good at discriminating 50-ms Gaussian-noise tokens with a spectral range of 350–850 Hz. However, when an identical 200-ms noise fringe, with the same statistical properties as the 50-ms target tokens, is appended to the end of both target tokens, listeners show very poor discrimination performance. Apparently, identical uninformative fringes cannot be ignored and they impair the discrimination of the target tokens. It seems that a target token and the appended fringe form one auditory object and that access to subparts of this object is not possible. When a perceptual cue is introduced that can lead to the segregation of the target token and noise fringe, e.g., a temporal gap between target and fringe, the ability to discriminate improves implying that the non-informative noise can be (partly) ignored when it is part of a different auditory object than the target token. It was shown that for the range of conditions used in these experiments, spectral separation and temporal separation were the strongest cues for auditory object formation.

Gaussische ruis discriminatie en auditieve object vorming

Samenvatting

Dagelijks wordt ons gehoor blootgesteld aan een enorme hoeveelheid auditieve input. Mogelijk is het een van de voornaamste taken van het auditief systeem om alleen de meest bruikbare informatie te extraheren uit deze veelheid van auditieve stimuli. In een akoestische scène zijn niet alle geluidproducerende gebeurtenissen van belang voor een luisteraar. Het is voor haar/hem voordelig om de irrelevante akoestische gebeurtenissen te negeren en de aandacht alleen te richten op de bruikbare. Ook in de golfvorm ten gevolge van een enkele akoestische gebeurtenis kunnen er kenmerken zijn die relevant zijn, en anderen die niet relevant zijn. Bijvoorbeeld, voor een Gaussische ruis zou het totale spectrum belangrijker kunnen zijn dan details in de fijne structuur ervan.

Hanna (1984) heeft laten zien dat de onderscheidbaarheid van breedbandige Gaussische stimuli vermindert met de toename van de tijdsduur voor stimuli langer dan 100 ms, ondanks dat perifere informatie toeneemt, terwijl voor stimuli korter dan ongeveer 25 ms de onderscheidbaarheid toeneemt met de toename van de tijdsduur. Kennelijk is er een niet-monotone relatie tussen hoeveelheid informatie veroorzaakt door de stimulus in de auditieve periferie en de hoeveelheid perceptuele informatie voor deze tijdsduren. Een van de centrale doeleinden van deze studie was het onderzoeken van het onderliggende mechanisme dat verantwoordelijk is voor deze niet-monotone relatie.

Hoofdstuk 2 beschrijft de replicatie van een van de experimenten van Hanna (1984), welke deze niet-monotone relatie voor het eerst heeft aangetoond voor Gaussische ruis. Hier werd een vergelijkbare niet-monotone afhankelijkheid gevonden met een maximale prestatie rond 40 ms. Additionele experimenten toonden aan dat de prestaties

van de luisteraar konden verbeteren wanneer dezelfde ruis stimulus werd gebruikt in alle tests (bevroren ruis). Echter, de tijdsduur waar het maximum optrad veranderde niet. In een ander experiment, gebruikmakend van een stimulus die bestond uit reeksen 5-ms lange tonen met een Hanning omhullende die willekeurig gedistribueerd waren in de tijd, werd onderzocht of de rollen van stimulusduur en informatiehoeveelheid de verwerkingscapaciteit van het auditieve systeem onafhankelijk beïnvloeden. De resultaten lieten zien dat de hoeveelheid vrijheidsgraden in de stimulus, maar niet zijn tijdsduur, de onderscheidbaarheid bepaalden. Over het algemeen suggereren de resultaten in dit hoofdstuk dat de onderscheidbaarheid van akoestische stimuli in hoge mate afhangt van de hoeveelheid informatie in een auditief object en van de capaciteit om deze informatie te verwerken. Deze capaciteit lijkt beperkt in de temporele dimensie, terwijl uitbreiden van het signaal over meerdere auditieve filters een positief effect heeft op de prestatie.

Modellen die alle informatie van de auditieve periferie combineren over tijd zullen de niet-monotone tijdsduur afhankelijkheid niet voorspellen. In plaats daarvan zal hun prestatie blijven toenemen met de tijdsduur tot deze satureert bij perfecte discriminatie. Hoofdstuk 3 beschrijft een model, gebaseerd op een bestaand model van Dau *et al.* (1996), dat in staat is om de niet-monotone tijdsduurafhankelijkheid die door Hanna (1984) was gevonden te voorspellen door de hoeveelheid informatie in de interne representatie van een stimulus onafhankelijk van de tijdsduur van de stimulus te limiteren. Deze benadering impliceert dat stimulus intervallen als niet deelbare auditieve objecten worden behandeld, en dat het model gelimiteerde middelen heeft die gelijkmatig verdeeld worden over het gehele object. Als gevolg doet het model voorspellingen over de onmogelijkheid voor luisteraars om een slechts gedeelte van een auditief object te verwerken. Deze voorspellingen werden geverifiëerd met luister experimenten. Ook was het model in staat om data met betrekking tot partiëel gecorreleerde te ruis uit een studie van Fallon (1989) te reproduceren.

Voor het beperken van de toegestane hoeveelheid informatie in de interne representatie van een auditief object is het nodig om te weten waar dit object start en waar het eindigt. Dit is eenduidig wanneer het object homogeen is en een duidelijk start en eindpunt heeft, zoals onze Gaussische ruis stimuli. Echter, wanneer potentiële segregatie cues zoals temporele separatie, spectrale separatie, bandbreedte, niveau verschillen, interaurale niveauverschillen en interaurale tijdvertragingen in de stimulus worden geïntrocueerd, dan zou de vorming van auditieve objecten beïnvloed kunnen worden. Hoofdstuk 4 beschrijft een methode om de invloed van zulke cues op de vorming van objecten te testen die geïnspireerd was op de voorspellingen van het model. De methode maakte gebruik van de observatie dat luisteraars goed zijn in het onderscheiden van Gaussische ruis stimuli met een tijdsduur van 50 ms en een spectraal bereik van 350–850 Hz. Echter, wanneer een identieke ruis stimulus van 200 ms, met dezelfde statistische eigenschappen als de doel stimulus, wordt toegevoegd achter beide doel stimuli, laten de luisteraars een zeer slechte prestatie zien. Kennelijk kunnen de identieke stimulus delen niet genegeerd worden en hebben zij een nadelige invloed op de discriminatie van de doel stimuli. Het lijkt alsof de doel stimulus en de toegevoegde niet informatieve stimulus samen een enkel object vormen, en dat toegang tot enkel een deel van dit object niet mogelijk is. Wanneer een perceptieve cue wordt geïntroduceerd die kan leiden tot segregatie van de doel stimulus en de niet informatieve stimulus, bijvoorbeeld een temporele ruimte tussen de twee stimuli, dan verbeterd de discriminatie, hetgeen impliceert dat het niet informatieve gedeelte (gedeeltelijk) genegeerd kan worden wanneer het onderdeel is van een ander auditief object is dan de doel stimulus. Het werd aangetoond dat voor de condities in deze experimenten spectrale en temporele separatie de sterkste cues waren voor de vorming van auditieve objecten.

Acknowledgement

It feels good to be able to write these final words. Writing these words means that a project of four years is coming to an end and that we have succeeded to end it successfully. I say *we* because, of course, I did not do this only by myself. I am greatly indebted to a number of people.

First, I would like to thank Steven and Armin for giving me the opportunity to start this project. I know I have been very lucky with such outstanding support, advice, guidance, and above all, having such pleasant supervisors. Furthermore, I am grateful to Ronald Aarts, Don Bouwhuis, Torsten Dau, Tammo Houtgast, and Brian Moore for participating in my committee.

For the non-Gaussian noise in our office at the Philips High Tech Campus, the daily cappuccino at the strip, and the Bob-Marley-friday-afternoons I would like to acknowledge my colleagues Michael Bruderer, Nicolas LeGoff, Tobias May, Alberto Novello, Othmar Schimmel, and Janto Skowronek. Without you guys I would not have had half as much fun as I did. Thanks.

I dedicate this thesis to my parents who have supported and stimulated me *all* the way, this means a lot to me. For the necessary diversion in the evenings and the weekends I thank my brother Rob and my friends. And finally, I am grateful to my girl Marloes for being there.

Curriculum Vitae

1977			September 19th, born in Uden, The Netherlands
1989	–	1993	LBO Electrical engineering, Ter Linde Uden
1993	–	1997	MBO Electrical engineering, De Leijgraaf Veghel
1997	–	2001	HBO Electrical engineering, Digital System Design, B.Sc. Fontys Hogescholen Eindhoven
2001	–	2004	WO Electrical engineering, Signal Processing Systems, M.Sc. Eindhoven University of Technology
2004	–	2008	Ph.D. student at the J.F. Schouten School for User-System Interaction Eindhoven University of Technology