ABSTRACT

The ability to discriminate two noise auditory stimuli increases with bandwidth. This ability also increases with duration, but only up to a duration of about 25 to 40 ms. Beyond this duration the discriminability decreases. In template-matching and multiple-look models [e.g. Dau et al., J. Acoust. Soc. Am. 99, 3615-3622 (1996)] this decrease of discriminability with increasing duration is not modelled correctly. At most, discriminability remains constant for durations above 25 to 40 ms. This suggests there is a limitation in the information processing capacity for long duration stimuli. The current study proposes a model which restricts the amount of stimulus information across time that is allowed in the internal representation of each noise token to a fixed amount. This approach implies that stimuli with different durations have incompatible internal representations. Accordingly, the model predicts that the ability to discriminate degrades strongly when a fringe is added to one of the stimuli. A listening test confirmed the model prediction that adding a fringe to one of the stimuli severely impairs the discriminability. Apparently stimulus fringes cannot be ignored, suggesting that these stimuli are processed as inseparable units which is in line with the basic assumptions of the model.

1 INTRODUCTION

At first glance, it may be expected that two auditory Gaussian noise stimuli with e.g. a 40-ms duration are more difficult to discriminate from each other by a human listener, than two Gaussian noise stimuli with a 400-ms duration. Because the longer duration stimuli carry more acoustical information, they should be more easy to discriminate. Several studies (e.g., Goossens et al., 2007, Hanna, 1984, Heller and Trahiotis, 1995) showed that this is not the case. In fact, the ability to discriminate is lower for the longer duration stimuli. Figure 3a shows our replication of a study by Hanna (1984) which measured the ability to discriminate Gaussian-noise tokens as a function of bandwidth and frequency. One of the main observations from these data is that the ability to discriminate Gaussian noise initially increases with duration up to a duration of about 25 to 40 ms. Above this duration, the discrimination ability does not increase anymore. In fact, it decreases with increasing duration despite the fact that there is more information in the stimuli on which listeners could determine perceptual differences. It seems that listeners cannot put all the stimulus information to the advantage of the discrimination task. Psychoacoustical template matching or multiple look models based on optimal processing of information (e.g., Dau et al., 1996, Viemeister and Wakefield, 1991) have the characteristic that internal representations of stimuli with a longer duration contain more information than internal representations of short duration stimuli. These internal representations are used to calculate perceptual dissimilarities and therefore, such models predict that the ability to discriminate should increase with duration (see Fig. 3b), or at most that the ability to discriminate stays constant beyond a certain duration. This is not wat was observed in the behavioral data from Goossens et al. (2007) in Fig. 3a or in Hanna (1984) and Heller and Trahiotis (1995). The observation that not all stimulus information contributes to the ability to discriminate these stimuli has led to the current model approach. In this paper we present a model, based on the one by Dau et al. (1996), which implements a processing-capacity limitation by restricting the amount of the perceptual information that is allowed in the internal representation.
2 A MODEL WITH A PROCESSING-CAPACITY LIMITATION

The decrease of discrimination ability for long duration stimuli which was described in the introduction may be a consequence of a surplus of stimulus information. When the information in the stimulus exceeds the amount of information that can be effectively used by the human auditory system for the discrimination task, the excess shows a detrimental effect on discrimination ability. Current template matching and multiple look models like the models by Dau et al. (1996) and Viemeister and Wakefield (1991) will not predict discrimination ability to decrease with increasing durations of above 40 ms because they combine all information that is available in the internal representation. This section describes a model approach based on an existing psychoacoustical model (Dau et al., 1996) with an extra stage that fixes the amount of information allowed in the internal representation. This way, the information of longer duration stimuli will need to be summarized more than for short duration stimuli which should cause the discrimination ability for longer duration stimuli to decrease.

2.1 Method and stimuli

The model that is presented in this section was embedded in a psychoacoustical procedure and acted as an artificial observer, giving the same type of answers as a human listener. The method was the same as described in Hanna (1984) and Goossens et al. (2007), a same-different procedure in which blocks of 100 trials were presented to the model. A trial consisted of two noise tokens that were independent in 50% of the trials and identical in the other 50%. For each condition the model was presented with 8 of these blocks with 100 trials. A $d'$ value was calculated for each individual block. The mean $d'$ and standard error were calculated by pooling the individual block $d'$ values per condition.

The stimuli consisted of Gaussian-noise tokens of which the bandwidth and the duration was varied over the conditions. The durations before filtering were: 1.6, 6.4, 10.2, 16.1, 25.6, 40.6, 64.5, 102.4, and 409.6 ms. The $-3$-dB filtering bandwidths were: 100–3300, 100–600, 225–275, 2800–3300, and 2975–3025 Hz. The spectrum level was 40 dB. Stimuli were presented diotically.

2.2 Model description

The presented model comprised four processing stages (see Fig. 1). In the first stage, the input signals $x_A(t)$ and $x_B(t)$, i.e. the two intervals in a discrimination trial, were both transformed to an internal representation, $R_A(c,t)$ and $R_B(c,t)$, using the model by Dau et al. (1996). Here, $c$ represents auditory channel number and $t$ time. This model consisted of the following subsequent parts: basilar-membrane filtering, inner hair-cell simulation, adaptation, temporal smoothing, and addition of internal noise. Outer and middle-ear filtering was not used in our simulations.

The processing-capacity limitation was implemented in the second stage. Here, the internal representations $R_A(c,t)$ and $R_B(c,t)$ were mapped to fixed-size internal representations $\hat{R}_A[c,n]$ and $\hat{R}_B[c,n]$, in which $n$ is the sample number, by sampling the internal representation with a fixed number of samples in the temporal dimension. Note that the spectral dimension already contained a fixed number of N (equals 52) basilar-membrane filters. For each auditory channel, the temporal sampling was achieved by applying 15 Hanning windows with 75% overlap to the internal representation (cf. Fig. 2). The length of the Hanning windows was always one third of the duration of the input stimulus’ internal representation including 150 ms of ringing of the filters after the stimulus offset. The signal in each window was averaged to a scalar value. The concatenation of the scalars from each window resulted in a fixed-size internal representation of 15 samples per auditory channel.

In the third stage, a scalar decision variable $D$ was calculated by summing the squared difference between the two fixed-size internal representations over time and auditory channels using

$$D = \sum_c \sum_n (\hat{R}_A[c,n] - \hat{R}_B[c,n])^2.$$  \hspace{1cm} (1)

The final stage calculated a decision about whether the input signals $x_A(t)$ and $x_B(t)$ were the same or different. First, an observation noise $N_{\text{decision}}$ was added to the decision variable $D$ to account for the inaccurateness of the listener in making the decision. This resulted in $\hat{D}$ which represented the perceptual distance between $x_A(t)$ and $x_B(t)$. When $\hat{D}$ was larger than a criterion $C$, the final output “$\hat{y}$” of the model was different otherwise it was same.
A \( R_A(t) \) \( R_A(1,t) \) \( R_A(N,t) \) Generate IR Reduce IR \( R_A[1,n] \) \( R_A[N,n] \) Calculate decision variable \( D \) \( \hat{D} \) Estimate maximum likelihood \( y \) \( y \) Input signals Internal representations Fixed size internal representations Decision variable Decision (same or different)

Figure 1: Block diagram of the model. Input stimuli (time continuous signals \( x_A(t) \) and \( x_B(t) \)) are transformed into internal representations (time continuous signals \( R_A(c,t), R_B(c,t) \)) with \( N \) auditory channels. Subsequently, these are transformed into fixed-size internal representations (time discrete signals \( \hat{R}_A[c,n] \) and \( \hat{R}_B[c,n] \)). Their difference is summed over frequency and time into a scalar value (\( D \)) and decision noise (\( N_{\text{decision}} \)) is added, resulting in a decision variable (\( \hat{D} \)). Finally maximum likelihood estimation is applied to \( \hat{D} \) to calculate a decision (‘\( y \)’ = same or different), i.e., an estimation of the true answer from the experiment (\( y = \) same or different).

Figure 2: Example for the information reduction in one auditory channel, \( R_A(c,t) \) (left panel), of a 409.6-ms noise stimulus. 15 overlapping Hanning windows are applied to the internal representation resulting in a fixed-size internal representation, \( \hat{R}_A[c,n] \) (right panel).

The criterion \( C \) was determined heuristically in the following way. At the start of a block of 100 trials, the criterion was set to a fixed arbitrary value larger than zero. However, \( C \) was adjusted after each trial by storing the values of \( \hat{D} \) in two separate storages. One storage for values of \( \hat{D} \) that represented the perceptual distance of same trials and the other for values that represented the perceptual distance of different trials. Note that the listeners could also build up such a storage because feedback \( y \) was presented in the experiment after each trial. Using Gaussian fits of the data in both storages and maximum-likelihood estimates (Green and Swets, 1988/1966), the decision device calculated a new criterion. Thus, in every subsequent trial, the model adapted to a more accurate criterion and its performance improved. After about six to ten trials the criterion had converged to a reasonable value.

The internal noise added to the internal representations \( R_A(c,t) \) and \( R_B(c,t) \) was calibrated as described in Dau et al. (1996). The variance of \( N_{\text{decision}} \) was calibrated such that the mean squared error relative to the mean behavioral data, see Fig. 3a, was minimal.

### 2.3 Results

Figure 3b shows the results of the simulations with the unmodified model by Dau et al. (1996) (without the \textit{Reduce IR} stage). Comparing curves of the model results with the behavioral results in Fig 3a, shows that the results of the model have a correct bandwidth dependency. The duration dependency, however, is incorrect because the ability to discriminate does not decrease with increasing durations above 40 ms.
Figure 3: The ability to discriminate Gaussian-noise tokens in $d'$ values (ordinates) as a function of stimulus duration (abscissas) for (a) human listeners, (b) the model by Dau et al. (1996), (c) and the model by Dau et al. (1996) with fixed information. Gaussian-noise bandwidth was 100–3300 (circles), 100–600 (squares, dashed lines), 225–275 (triangles, dashed lines), 2800–3300 (squares, dash-dotted lines), and 2975–3025 Hz (triangles, dash-dotted lines). The error bars indicate the standard error of the mean.

Figure 3c shows the results of the simulations with the model by Dau et al. (1996) including the Reduce IR stage. The correct bandwidth dependency is preserved. In addition, the duration dependency is similar to that shown in the behavioral results in Fig 3a. The largest discrepancies are observed in the 100–3300 Hz conditions of short durations (1.6 – 16.1 ms) that overlap with the 100–600 Hz in the behavioral data but not in the simulated data.

3 DISCRIMINATION OF FRINGED GAUSSIAN-NOISE

An implication of the proposed model approach is that stimuli with different durations have incompatible fixed-size internal representations. Due to the addition of the fringe and the longer window size, corresponding windows cover a different part of the input signals. Accordingly, the model predicts that the ability to discriminate should degrade strongly when a fringe is added to one of the stimuli. Alternatively, listeners might be able to focus on the corresponding signal parts ignoring the fringes. In this experiment we investigated if human listeners are able to perform discrimination when a fringe is added to one of the two tokens in each trial.

3.1 Method and stimuli

The method was the same as described in Hanna (1984) and in Goossens et al. (2007), which was a same different procedure in which blocks of 100 trials were presented to the listeners. Three listeners participated. A trial consisted of two noise tokens that were independent in 50% of the trials, and partially identical in the other 50%. All three listeners were presented with 4 of these blocks with 100 trials for each condition. The order of the blocks was randomized. A $d'$ value was calculated of each individual block of each listener. The mean $d'$ and standard error were calculated by pooling the individual block $d'$ values per condition.

The −3-dB bandwidth of the Gaussian noise was 100–3300 Hz with a spectrum level of 40 dB. Figure 4 shows a schematic time-line of the four fringe conditions. The duration of the target tokens A and B were 25.6 ms. These tokens were identical in 50% of the trials and independent in the other 50%. Depending on the experimental condition, a backward- or forward-fringe with a 384 ms duration was added to token A or token B, see Fig. 4. The listeners were notified about the location of the fringe prior to each block of 100 trials via a written instruction on the computer screen. The duration between the onsets of token A and B was 909.6 ms in each condition.

For each trial, new stimuli were made by generating two tokens of broadband Gaussian-noise with a duration of 409.6 ms. In a same trial these were identical, in a different trial these were independent. Subsequently, both tokens were filtered to the appropriate bandwidth, introducing on- and offset effects which were left intact. One of these two tokens was the token with noise fringe (target duration of 25.6 ms plus fringe duration of 384 ms). The other token was truncated, from the beginning (cf. Fig 4, III and IV) or the end...
(I and II), to a duration of 25.6 ms using a 10 ms raised cosine on- or offset ramp in order not to introduce 
spectral splatter.

<table>
<thead>
<tr>
<th></th>
<th>Fringe</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Forward fringe on token A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Forward fringe on token B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Backward fringe on token A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Backward fringe on token B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Schematic time line for the fringe conditions. Gaussian-noise target-tokens A and B with duration 
25.6 ms are either the same or different. The fringe duration was 384 ms. The time interval between the 
onsets of token A and B was 909.6 ms in each condition.

3.2 Results

Table 1 shows the results of the conditions where fringes were added to 25.6 ms noise tokens as well as of 
the original 25.6-ms, 100–3300-Hz condition from Goossens et al. (2007). All the conditions with a fringe 
show close to chance level performance indicating that adding a fringe to one of the stimuli severely impairs 
the ability to discriminate these stimuli. Apparently listeners cannot ignore stimulus fringes, suggesting that 
these stimuli are processed as inseparable units, which is in line with the basic assumptions of the model.

Table 1: Results of the fringe conditions, \( d' \) mean and standard error of the original condition and conditions 
with four types of fringes.

<table>
<thead>
<tr>
<th></th>
<th>original (^1)</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d' )</td>
<td>3.138</td>
<td>0.027</td>
<td>0.030</td>
<td>0.259</td>
<td>0.258</td>
</tr>
<tr>
<td>std. err.</td>
<td>0.100</td>
<td>0.072</td>
<td>0.072</td>
<td>0.160</td>
<td>0.157</td>
</tr>
</tbody>
</table>

4 FRINGE DURATION EXPERIMENT

The next experiment was designed to investigate to which extent the model can quantitatively predict the 
dependence of fringe duration on the ability to discriminate Gaussian noise.

The method was the same as for the experiment in Sec. 3. The same three listeners participated. The 
−3-dB bandwidth of the Gaussian noise was 100–3300 Hz with a spectrum level of 40 dB. In this experiment 
only the backward fringe on token B (condition IV in Fig. 4) was used. Target token durations were 25.6 and 
102.4 ms. Fringe durations were 0, 6.7, 15, 38.9, 76.8, and 384 ms for the 25.6-ms target tokens, and 0, 
26.8, 60, 153.6, and 307.2 ms for the 102.4-ms target tokens. In each condition, the duration between the 
onsets of token A and token B was 500 ms plus the duration of the fringe.

The stimuli were generated in the same way as in Sec. 3, except that now the stimulus without fringe 
was truncated from the stimulus with fringe, not using a ramp. Afterwards, the stimuli were filtered to the 
appropriate bandwidth, introducing on- and offset effects which were left intact.

4.1 Results

Figure 5 shows the ability to discriminate Gaussian-noise tokens when a backward fringe was added to the 
second noise token for human listeners (circles) and the presented model (triangles). The results are shown 
as a function of the total duration of the second noise token; i.e., the noise token with the added fringe. The 
model was not recalibrated and all its parameters were the same as for the simulations in Fig. 3c. The 
standard error of the mean was in the range of 0–0.23.

The model simulations and the behavioral data both show a rapid decrease of discrimination ability with in-
creasing fringe duration for both target token durations and are highly similar. The largest difference between

\(^1\) Without fringe, cf Goossens et al. (2007)
the data and the model is that performance of the model decreased to a level of 0.14 when the target token was 102.4 ms, whereas the performance of the listeners decreased to 0.49. This may indicate that for this duration there is still some limited ability for the listeners to focus on the relevant stimulus part.

5 CONCLUSIONS

The decreased ability of human listeners to discriminate Gaussian noise tokens with durations above 25 to 40 ms suggests that only a limited amount of stimulus information can contribute to the discrimination of the stimuli. Exceeding this amount of information apparently has a detrimental effect on discrimination ability. This duration dependency is incorrectly predicted by current psychoacoustical models that optimally combine the information over time; e.g., template matching and multiple look models like the models by Dau et al. (1996) and Viemeister and Wakefield (1991). These models show a discrimination ability that always increases with duration. We were able to improve these models by incorporating a stage into an existing model by Dau et al. (1996). This extra stage limited the amount of information in such way that the information in the internal representation of a stimulus was independent of the stimulus duration.

This approach, however, implies that the stimuli are treated holistically and that auditory stimuli with different lengths have incompatible internal representations. Moreover, the modeling approach implies that the stimuli undergo post processing in the auditory pathway because the processing depends on the stimulus duration and for reasons of causality this duration is unknown before the stimulus has been presented entirely.

Model simulations of discrimination experiments (see Sec. 4) where a noise fringe is added to one of the two stimuli within a trial predict a strong decrease of discrimination ability with increasing duration of the noise fringe. A similar decrease of discrimination ability was observed for human listeners in Sec. 3 and 4 which corroborates the model approach.

On the basis of the model simulations and the listening tests we suggest that in a discrimination task of stochastic stimuli, stimuli are processed as non-separable auditory objects or entities. The amount of information in the auditory objects that can be used by listeners is limited to a fixed amount.

References


