Proceedings of the Institute of Acoustics

DETECTION OF SPEECH IN THE PRESENCE OF DELAYED SAME SIGNAL REINFORCEMENT

K Vivatvakin ISVR, University of Southampton, UK
IH Flindell ISVR, University of Southampton, UK

1. INTRODUCTION

In psychoacoustics, masking is defined as 'the process by which the threshold of audibility for one sound is raised by the presence of another sound' (American Standards Association, 1960). Many problems in auditorium acoustics can be explained by reference to masking phenomena, for example where direct sound is masked by reflections or echoes (Bolt, 1949; Beranek, 1986; Strong, 1992). Strong reflections from the side walls will also arrive later than direct sound from the front and can mask direct sounds to some extent, leading to possible losses in sound quality or intelligibility or both. This paper describes two experiments on speech masking by delayed same signal reinforcement.

The first experiment was carried out under free-field listening conditions over varied masker delays (15, 30, 50, and 100ms), varied masker sound level (40, 60, and 80dBA), and varyed masker angle of incidence (0, 30, 60, and 90 degrees). The data suggest that masked thresholds decrease with masker delay, increase with masker sound level, and are relatively unaffected by masker angle of incidence except for whether the masker is coming from the same or different direction to the direct sound.

The second experiment was carried out using headphones to investigate self and overlap masking phenomena over varied masker delays and using frequency band limited speech and masker signals to investigate same and different frequency masking..

2. BACKGROUND

Many auditoria rely on either passive reflections or active reinforcement via loudspeakers to enhance sound quality or intelligibility or both. Delayed same signal reinforcement has to be finely balanced because it can do more harm than good, particularly in highly reverberant spaces (Bolt, 1949; Beranek, 1986; Strong, 1992). The first reports of relevant masking effects appear to have been made by Knudsen in 1929. He theorised that masking effects arise when the energy of a preceding sound overlaps onto the following sound. This is called 'overlap masking' and can be easily demonstrated in many practical situations. Self masking was defined by Bolt and Macdonald (1949) as an internal temporal smearing of the energy within each sound. Self masking often appears together with overlap

masking in many practical situations, but it is uncertain which effect takes precedence. An additional problem arises in public address systems where the direct sound is completely replaced by same signal reinforcement and masking of the direct signal is no longer an issue and may even be beneficial. Consequently, it is accepted that the relevance of masking phenomena to speech perception in different auditorium situations is not completely understood.

3. PERCEPTUAL MASKING IN SOUND FIELD WITH SINGLE REFLECTION (ECHO)

3.0 Introduction

The main purpose of this experiment is to investigate the relationship between masking effects caused by delayed same signal reinforcement in relation to differences in the azimuth directions of the direct and masker signals.

3.1 Experimental Design

The experiment was carried out under free-field listening conditions in an anechoic room at the ISVR. The direct sound was always presented from the front (ξ = 0°) while the same signal masker was presented over a range of different time delays using a range of different azimuth angles (ξ = 0°, 30°, 60° and 90° (left-hand side only)). The anechoic chamber was 4.33m wide × 4.95m long × 2.74m high and all loudspeakers (dual concentric bookshelf type loudspeakers KEF C35) were positioned at a constant radius of 2m from the center of listener's head when seated centrally. The initial time delays Δt between the direct and reflected sounds were chosen Δt = 15, 30, 50 and 100ms. The masked threshold levels of the direct speech signals were measured at masker sound levels of 40, 60, and 80 dBA (LTRMS) measured at the subject's head position.

3.2 Stimuli and Calibration

All direct signals and same signal maskers were running male and female speech, carefully edited to eliminate glitches etc using .wav format in a personal computer. The direct signal levels were controlled by the experimenter using a Kamplex type AD-27 audiometer. Regular calibration before and after each test session was carried out at the centre of the listeners head position with the listener absent using a DAWE type D-1421D sound level meter.

3.3 Procedure

The psychoacoustic up-down or staircase method was used to measure masked thresholds. Listeners were instructed to press the audiometer response button whenever they heard or thought that they heard the speech from the direct sound speaker (at the front). The experimenter always started testing with the 15ms delay condition at 40dBA masker sound level at the side on incidence angle of 90 degree. The direct signal level was ramped up in 5dB steps from the 0 dB start point until the listener

responded positively. After each positive response, the direct signal level is ramped down by 5 dB steps until the listener signals that they can no longer distinguish the direct sound by releasing the response button. Then the signal is ramped back up agains until a further response is elicited and so on. Each threshold measurement run consisted of 10 reversals, where the first 4 reversals are ignored and the threshold taken as the average of the final 6 reversals.

3.4 Results

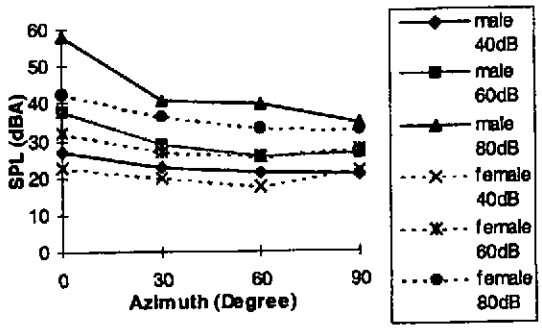

Figure 3.1 Threshold level when masker arrived from 0, 30, 60, and 90 degree of azimuth

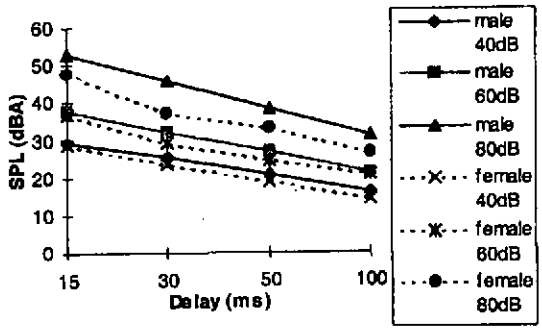

Figure 3.2 Threshold level when masker delayed 15, 30, 50, and 100ms

3.5 Discussion

3.5.1 Level

On average, the masked thresholds increased by 6dBA as the masker level increased from 40dBA to 60dBA, and then increased by a further 11dBA when the masker level increased to 80dBA. This non-linear relationship is consistent with previous findings in the literature (Henning & Zwicker, 1984). The relative signal detectibility seems to improve as the masker level increases. In the context of this experiment, there is a possibility that slightly increased loudspeaker distortion at the higher masker sound levels may have contributed to enhance detectability, but it seems more likely that the reduced detectability at the lower masker sound levels could have resulted from self masking in the auditory system itself. Interestingly, this finding of a reduced masking effect at higher masker sound levels is counter to established wisdom that intelligibility decreases at higher sound levels.

3.5.2 Type of signal

The male speech thresholds were slightly higher than the female speech thresholds. This result could have been due to differences in fundamental frequency and associated higher harmonic content for voiced sounds between the male and female speech (Zwicker & Henning, 1985), but it could also have been due to several other factors, such as the tendency for the female speaker to leave slightly longer silence gaps between utterances and thereby changing the peak to long term rms ratio of the speech.

3.5.3 Delay

The relative time delay between direct and masker signals was found to be very important in this experiment. It can be clearly seen that an increase in masker delay has led to the reduction of the threshold level. When masker delay was 15ms, the threshold were high. The speech signal were difficult to detect because of precedence effect (Aoki & Houstgast, 1992). When the time delay was increased, the direct speech signals became relatively easier to detect.

3.5.4 Azimuth Angle

There were no differences in masked thresholds for any masker angles from 30 to 90 degrees, although the direct signals were harder to detect when both signal and masker came from the same direction at 0 degrees azimuth. This result is also consistent with previous findings reported in the literature (Saberi et al, 1991; Gilkey & Good, 1995) and implies that directional hearing in spatial masking situations may be more sensitive to similarities and differences in horizontal azimuth angle than to specific azimuth angles per se.

4. SELF AND OVERLAP MASKING EFFECTS

4.0 Introduction

Signal detection in the present of delayed same speech signal is mediated by similar masking phenomema to those investigated in the first experiment. Either or both of overlap masking and self masking could be important here. They have been mentioned in literature widely. Nevertheless, these effect can be only explained in term of energy, and cannot be explained in term of our auditory system. With regarding to our hearing mechanism, masking effect much depends on the preperties of signal and masker, including delay, characteristic of signal, frequency, and duration. Consequently, two types of masking which are simultaneous and non-simultaneous masking would be investigated instead of overlap and self masking. Simultaneous masking occurs when the signal and masker are presented at the same time. It is similar to overlap masking. Non-simultaneous masking can occur when the signal and masker are not completely time-coincident, thus allowing an opportunity for signal detection either just before or just after the masker is present. Forward masking describes the situation where the masker precedes the signal. Backward masking describes the situation where the masker follows the signal. For non-simultaneous masking to occur, there must be some time delay or integration period within the auditory system and within which period some masking can occur even though the masker is not physically present at that time..

'The threshold of signal in the present of speech masker is determined by combination of simultaneous and non-simultaneous masking, rather than either factor alone' (Spiegel, 1987). Clearly, when the masker is not delayed, the threshold of the speech signal depends on simultaneous masking effect alone. However, when the speech masker is delayed, both simultaneous and non-simultaneous masking could be involved. The thresholds in this case depend on the characteristic of the signal, because the characteristis of the speech can provide detection cues of various kinds. For example, the silence gap between successive syllables in the masker might provide opportunities for direct signal

detection (Wilson & Carhart, 1971; Zwicker, 1984). The less release from masking has been found in smaller gap, the greater the release from masking has been found in the bigger gap. However, speech is unlike noise. Certain low amplitude consonants could be considered as a sort of silence gap in relation to other speech components (Spiegel, 1987). Using this cue in speech detection related to both simultaneous and non-simultaneous masking, in contrast to the silence gap which related to non-simultaneous masking effect alone. The masking effect in this case therefore more complicated. The concept of critical bands and the upward spread of masking need to be taken into account.

The major aim of this experiment was to examine the dominant type of masking on speech detection in the present of a delayed speech masker. A subsidiary aim was to examine the effect of speech masker properties on speech detection including the effect of silence gaps in the speech

4.1 Experimental Design

The experiment was separated into two parts. The first part was an investigation of signal detection in the time domain, whereas the second part was an investigation of signal detection in the frequency domain. In the first part, the effect of the silence gap on signal detection was investigated. Masked thresholds were measured using the normal running speech stimuli and compared against similar thresholds measured using continuous running speech stimuli which had been digitally edited to remove all the silence gaps between utterances.

In the second part, the effect of same and different masker and signal frequencies was investigated. The speech frequencies were broadly classified as first formant vowel frequencies (less than 800 Hz), second formant vowel frequencies (800 Hz to 2 kHz) and higher formants (including consonants) (2 kHz and above). It is known that the most important formants used to identify different vowel sounds are the first two.

Azimuth angle was not a parameter in this experiment so all signals and maskers were presented monophonically using high quality electrostatic headphones, which were calibrated using a B&K artificial ear system.

4.2 Stimuli and Calibration

Stimuli are running speech of male and female. The stimuli were edited by personal computer. In the first part of the experiment the silence gaps were removed digitally. In the second part of the experiment the stimuli were filtered digitally in to three frequency ranges (first formant, second formant, and higher formant).

4.3 Experimental Session

stimuli	Signal	Masker	delay time
Normal speech	with silence gap	with silence gap	15, 30, and 50ms
Normal speech	without silence gap	without silence gap	15, 30, and 50ms
Filtered speech	1st formant	1st formant	15, 30, and 50ms
Filtered speech	1st formant	2nd formant	15, 30, and 50ms
Filtered speech	1st formant	higher formant	15, 30, and 50ms
Filtered speech	2nd formant	1st formant	15, 30, and 50ms

Filtered speech	2nd formant	2nd formant	15, 30, and 50ms
Filtered speech	2nd formant	higher formant	15, 30, and 50ms
Filtered speech	higher formant	1st formant	15, 30, and 50ms
Filtered speech	higher formant	2nd formant	15, 30, and 50ms
Filtered speech	higher formant	higher formant	15, 30, and 50ms

Table 4.1 Experiment session

Each subject attended a total of two sessions. In each session, there was 34 conditions, including both parts of the experiment. The first session of the experiment was the male speech stimuli, whereas the second session of the experiment was the repeat of the first session with the female speech stimuli. These had been done in order to make the subjects to be familiar with the stimuli in each session, thus the subjects were be able to detect the stimuli in different formant.

4.4 Procedure

Measurements were carried out with the subject seated in a sound-attenuating booth. The absolute thresholds of both male and female speech were measured using psychoacoutics method of staircase as decribed in session 3.3.

4.5 Result

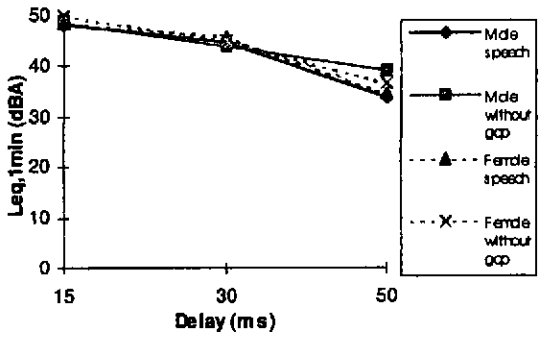

Figure 4.1 Threshold for normal speech and speech without silence gap with various masker delayed.

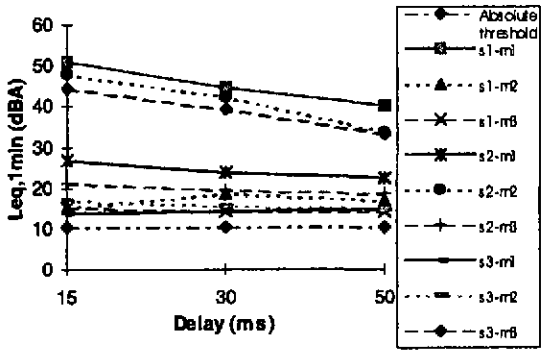

Figure 4.2 Threshold level when signal and masker were filtered into various frequency ranges.

4.6 Discussion

4.6.1 The Effect of Silence Gap

Overall masking thresholds measured for speech with and without the silence gap were the same. This finding does not preclude the possibility that vowel signals could have been detected during masker periods represented only by lower amplitude consonant sounds. However, when we take a closer look at the data, it is apparent that the masked threshold levels for speech with and without silence gaps for the masker time delay of 50ms are slightly different. This finding could be related to the typical duration

of actual speech component sounds (separate phonemes) in relation to simultaneous masking phenomena. Also, when editing the speech stimuli to remove the silence gaps, it was noted that some silence gaps were longer in duration than some of the low amplitude consonants present. To the extent that certain low amplitude consonants might be considered as equivalent silence gaps for masking purposes, the results could be explained by simultaneous masking phenomena associated with the shorter duration low amplitude consonant gaps which are still present even when the silence gaps had been removed.

4.6.2 Same and Different Frequency Masking Effects

As can be seen in Figure 4.2, there were clear differences in masked threshold between same and different frequency masking situations. When masker and signal were both in the same frequency ranges, the masked thresholds were in all cases significantly higher than for the different frequency range cases.

When signal and masker were at the same frequency ranges, the masked thresholds were similar to those measured for normal speech. The actual threshold levels appear to depend on masker delay in a way that is consistent with known non-simultaneous masking effects, with thresholds decreasing as masker delay increased toward 50ms. Even though the thresholds obtained in this groups have shown the same masking release pattern, their threshold level were slightly difference from one another. The thresholds of the third frequency range were found to be lower than that of the first and the second frequency ranges. These difference may be a result of differences in relative gap duration when the signals have been filtered into different frequency bands. It appears that the masker duration affects the threshold level mainly because of non-simultaneous masking phenomena (see Zwicker & Fastl, 1990).

The thresholds obtained when maskers were in different frequency ranges from the signal were not significantly affected by masker time delay, and suggests that only simultaneous masking phenomena were occuring under these circumstances. There was some indication of a small upwards spread of masking effect - consider the fisrt formant masker, second formant signal results which have higher thresholds than for the remaining different masker and signal frequency results.

5. CONCLUSION

The findings to date suggest that psychoacoustic masking phenomena can be important for speech detection in the presence of same signal delayed reinforcement under free field listening conditions and that both simultaneous and non-simultaneous masking phenomena can be more or less relevant for different test signal situations. For speech detection in the presence of same signal delayed reinforcement, the delay relative to the phoneme rate appears more important than any silence gaps present. The absolute magnitude of any differences in relative azimuth angle between signal and masker appear to be unimportant apart from the basic issue of whether they are same or different. A separate analysis, not reported here for reasons of space, suggests that current objective measurements intended to assist in judgements of auditorium sound quality such as the IACC (interaural cross-correlation) do not appear to be measuring the most relevant physical properties of the space in relation to the masking phenomena investigated in this study.

6. REFERENCES

1. American Standards Association. 1960. ***Acoustical terminology SI 1-1960.*** American Standard Association: New York.
2. Aoki, S. and Houtgast, T. 1992. ***A precedence effect in the perception of inter-aural cross correlation.*** Hearing Research. Vol.59, 25-30.
3. Beranek, L.L. 1986. ***Acoustics***, The American Institute of Physics: USA.
4. Bolt, R.H. and Macdonld, A.D. 1949. ***Theory of speech masking by reverberation.*** J.Acoust.Soc.Am. 21(6), 577-580.
5. Gilkey, R.H. and Good, M.D. 1995. ***Effect of frequency on free-field masking.*** Human factors. Vol.37(4), 835-843
6. Henning, G.B. and Zwicker, E. 1984. ***Effect of the bandwidth and level of noise and of the duration of the signal on binaural masking level difference. Hearing Research.*** Vol.14.
7. Knudsen, V. O. 1929. ***The hearing of speech in auditoriums.*** J.Acoust.Soc.Am. 1, 56-82.
8. Saberi, K., Dostal, L., Sadralodabai, T. Bull, V. And Perrott, D.R. 1991. ***Free field release from masking.*** J.Acoust.Soc.Am. 90(3), 1355-1370.
9. Spiegel M.F. 1987. ***Speech masking. I. Simultaneous and nonsimultaneous masking with stop /d/ and flap /f/ closures.*** J.Acoust.Soc.Am. 82(5), 1492-1502.
10. Wilson, R.H. and Carhart, R. 1971. ***Forward and backward masking: interactions and additivity.*** J.Acoust.Soc.Am. 49(4), 1254-1263.
11. Zwicker, E. and Fastl, H. 1990. ***Psychoacoustics.*** Springer Verlag Berlin Heidelberg: Germany.
12. Zwicker, E. and Henning, G. B. 1985. ***The four factors leading to binaural masking-level differences.*** Hearing Research. Vol.19, 29-47
13. Zwicker, E. and Henning, G. B. 1984. ***Binaural masking-level differences with tone masked by noise of various bandwidth and levels.*** Hearing Research. Vol.14, 179-183.