Proceedings of the Institute of Acoustics

PERCEPTUAL COMPENSATION FOR THE EFFECTS OF REVERBERATION ON AMPLITUDE-ENVELOPE CUES TO THE "SLAY"-"SPLAY" DISTINCTION

Anthony J. Watkins

Psychology Department, Reading University, Reading, UK.

1. INTRODUCTION

High levels of reverberation reduce speech intelligibility while at lower levels intelligibility is little affected even though substantial distortion of the speech signal is still present [1,2] in amounts that are sufficient to perturb speech-recognition automata [3]. This suggests that perceptual mechanisms operate to ameliorate the effects of this type of distortion if it is not too severe. One type of mechanism that does this is binaural: listening with two ears in reverberant conditions gives a 3-dB improvement in signal detection over the monaural condition [4] and speech intelligibility in reverberation is better binaurally than monaurally [5]. Compensation for reverberation might also come from monaural mechanisms in which information about the distortion of earlier sounds is compensated in the processing of later sounds. Evidence of this general kind of mechanism was found by Haggard [6] who showed that the identification of distorted words is improved if they are preceded by a carrier utterance that is distorted in a similar way. However, distortion by reverberation was not specifically studied. Reverberant carriers do enhance the perception of subsequent reverberant sounds, at least for the purposes of sound localisation. This was shown by Plenge [7] who found that a brief familiarisation with a room's reverberation improves sound localisation considerably.

One aspect of reverberation is a distortion of the spectral envelopes of the sounds that are transmitted [8]. This could impair speech intelligibility as spectral envelopes are an important determinant of the identity of speech segments and many other sounds. Spectral-envelope distortion is perceptually compensated by monaural mechanisms: compensation arising from a preceding carrier has been shown to affect a variety of subsequent vowel-distinctions as well as the consonant distinction between "slow" and "flow" [9]. This compensation mechanism appears to be central (as opposed to peripheral) and auditory (as opposed to phonetic) [10].

Another aspect of reverberation is distortion of the amplitude envelope: reverberation fills in gaps, extends offsets and smooths onsets, acting as a low-pass filter of the amplitude envelope [11]. The amplitude envelope is important for consonant distinctions. This is especially true for distinctions such as that between "slay" and "splay" where other sources of information about the presence of the /p/, notably the characteristic spectrum of the short burst that often follows the silent gap, are often weak or absent in these bilabial stops [12]. This study asks whether reverberation impairs this distinction, and whether any impairment is compensated when the same reverberation is also present in a preceding carrier.

The present experiments extend an earlier study in which reverberation was introduced by playing sounds from a speaker at one end of a room and recording from a microphone at the other end [13]. Evidence of compensation was found but there were some uncertainties about the stimuli due to the idiosyncrasies of the room, transducer placements, and transducer directionality.

2. GENERAL METHOD

Sounds were played through a computer-simulated room by convolution with the impulse response that was calculated using the method of images program [14]. These computations were performed with a resolution of 20 kHz using a Sun Sparcstation computer.

This simulation departs a little from truly natural reverberation: Transducers are omnidirectional, whereas real talkers and listeners are not; Low frequency components of echoes (below 200Hz here) cannot be simulated; Above 200Hz reflections are specular, which is uncommon but not impossible. Otherwise the method allows simulation of diverse reverberation patterns, controlled by parameters that specify physical aspects of the listening conditions. In the present experiments the Sabine energy absorption coefficient of the room's surfaces, α was varied while the room's size and the transducer placements were held constant.

The dimensions of the simulated room in feet were 16 x 32 x 16. The source and receiver were 19.6 feet apart along the room's long diagonal, equidistant from the centre. All surfaces were given the same value of α, and the values used are shown in the table below along with aspects of the reverberation patterns that these values generated in the room. The value used for the speed of sound was 1000 feet per second. The reverberation times are the -60-dB points of energy decay curves that were calculated by reverse integration [15].

α	Typical Surface at 500 Hz [16]	Reverberation Time, Seconds	Direct/Reverberant Ratio, dB
.1	Wood	1.48	-30
.2	Carpet	1.20	-24
.3	Coarse concrete block	.78	-20
.4	2.5-cm cork, airspace behind	.53	-17

A male speaker (AJW) was recorded saying "next you'll get the word","slay", and "splay". The phrase was to be used as a carrier. The other recorded words were spoken so that the phrase was a suitable preceding context. Recordings were made in an IAC 1201 booth using an Altec 681A microphone. These signals were amplified (Revox, A77), low pass filtered at 9 kHz with a 48-dB per octave cutoff slope (Kemo VBF8), digitised with 16-bit resolution at a sampling frequency of 20 kHz (Data Translation DT2823) and stored with the ILS program RDA [17] running on a Victor PC286 computer.

The test sounds differed only in their amplitude envelopes. They were all derived from the recording of "slay" by amplitude modulation. This was done with ILS programs as follows:

a) The recordings of "slay" and "splay" were time-aligned at the onset of voicing.
b) The amplitude envelopes of both words were obtained by playing the full-wave rectified waveforms through a low-pass filter that had a 50-Hz cutoff. The output was then reversed, filtered again (by the same filter) and reversed. This zero-phase filtering preserves the time alignment of the envelopes.
c) Each point in the envelope of "splay" was divided by the temporally corresponding point in the envelope of "slay" to obtain the envelope ratio.
d) A value of interpolation, k was chosen. Values between 0.0 and 1.0 in steps of 0.1 were used for a continuum from "slay" to "splay".
e) Each point in the envelope ratio was multiplied by k followed by the addition of 1 - k to obtain the modulation function.
f) The original recording of "slay" was multiplied by the modulation function.

Digital waveforms of this test-sound continuum and the carrier were transferred to the Sun Sparcstation computer, reverberated, and returned. Test sounds were reverberated in isolation as well as with the carrier abutted. The reverberant 'tails' at the ends of sounds were preserved. These tails last for the duration of the room's impulse response which was truncated at the -72-dB point of the energy decay curve.

Sounds were delivered to subjects on-line under the control of the PC286 computer. Analog signals were generated from the digital waveforms with 16-bit resolution at a conversion rate of 20 kHz (Data Translation DT2823) using the ILS program LDA. These signals were low pass filtered at 9 kHz with a 48-dB per octave cutoff slope (Kemo VBF8) and presented to subjects monaurally with Sennheiser HD424 headphones at 53-dB sound-pressure level in the IAC 1201 booth. On each trial a test sound was presented in isolation (experiments 1a and 1b) or with a preceding carrier (experiment 2). Subjects were asked to identify the test sound by pressing one of six response buttons that were labelled for a "slay" to "splay" rating scale. This was done with "definitely", "probably" and "maybe" qualifiers on the labels. Visual prompts to listen or to respond were conveyed by messages on the computer's screen. The computer waited for the subject's button press before recording the response and presenting the next trial. A minimum intertrial interval of 4 s was enforced.

In experiments 1a and 1b each of the 11 continuum steps was presented 8 times, and the continuum was reverberated with 3 different values of α. An undistorted continuum was also presented giving 11 x 8 x 4 = 352 trials per subject. In experiment 2 the two end-points of the continuum were presented 20 times each. They had an undistorted carrier or a carrier that had the same reverberation (α). 4 values of α were used giving 2 x 20 x 2 x 4 = 320 trials per subject. Different random orders of these trials were obtained for each subject and they were administered in one 45-min session without practice trials. Experiments 1a and 1b each used 5 subjects. Experiment 2 had a between-subjects factor and used two groups of 5 subjects.

3. EXPERIMENTS 1A AND 1B

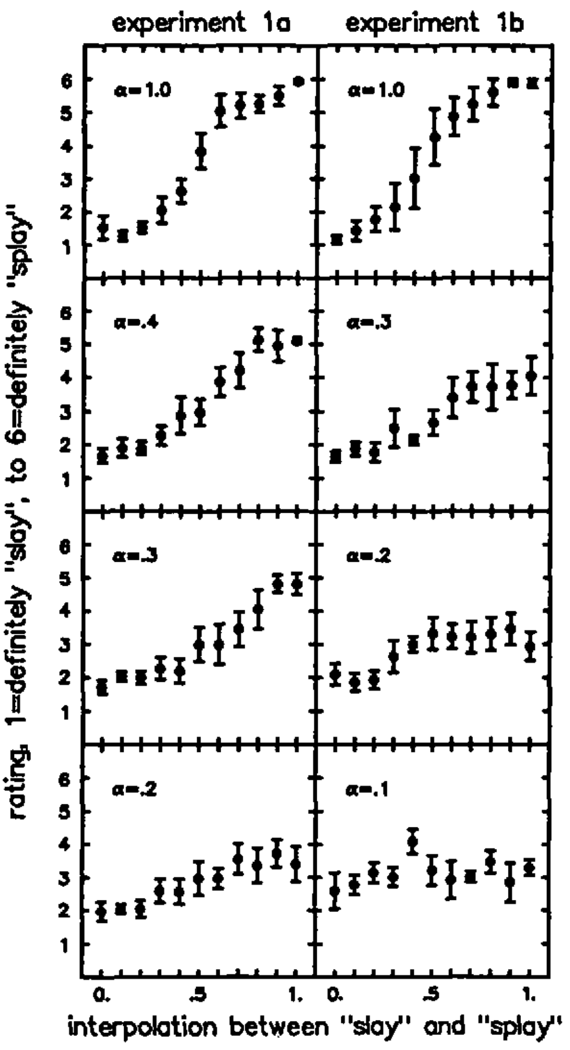
These experiments ask whether the amplitude modulation technique is sufficient for "slay" and "splay" identification with the present stimuli and whether this identification is influenced by reverberation. Some combinations of test sounds and distortion can lead to a shift in the category boundary [9] but Watkins [13] found that reverberation influenced an amplitude-envelope continuum ("dish" to "ditch") by flattening the identification function, leaving the category boundary in more or less the same position.

In experiment 1a three continua had reverberation generated with α = 0.2, 0.3 and 0.4. There was also an undistorted continuum which is effectively reverberation with α = 1.0. Experiment 1b was similar except that the values of α were 0.1, 0.2 and 0.3.

The mean rating was found for each stimulus and for each subject: A number was assigned to each response category, 1 for definitely "slay" up to 6 for definitely "splay". The probability of each category was found by dividing the number of responses in the category by 8 (the number of times each stimulus was presented). Each category number was multiplied by the corresponding category probability and these 6 values were added together to obtain the mean rating.

Mean ratings, averaged across the 5 subjects, are shown as a function of interpolation and for the different values of α in the adjoining figure. Bars are one standard error on each side of the mean. Analysis of variance was performed on the mean ratings for each experiment. This revealed main effects of interpolation: $F(10,40) = 34.19$, $p<0.0001$ in experiment 1a and $F(10,40) = 16.23$, $p<0.0001$ in experiment 1b. This reflects the increase in "splay" ratings with increasing interpolation. There were also interactions between α and interpolation: $F(30,120) = 4.99$, $p<0.0001$ in experiment 1a and $F(30,120) = 6.67$, $p<0.0001$ in experiment 1b. This reflects the flattening of these identification functions as reverberation is increased. The only other significant F-ratio is a main effect of α in experiment 1a: $F(3,12) = 4.43$, $p<0.03$. This reflects an inclination towards "slay" ratings as reverberation increases.

Results with the undistorted carrier show that amplitude-envelope cues are sufficient for "slay" and "splay" identification. These averaged identification functions reach the extreme ratings at the extreme values of interpolation indicating that all subjects heard good exemplars of these categories on virtually every presentation of stimuli that were near the ends of the continuum.



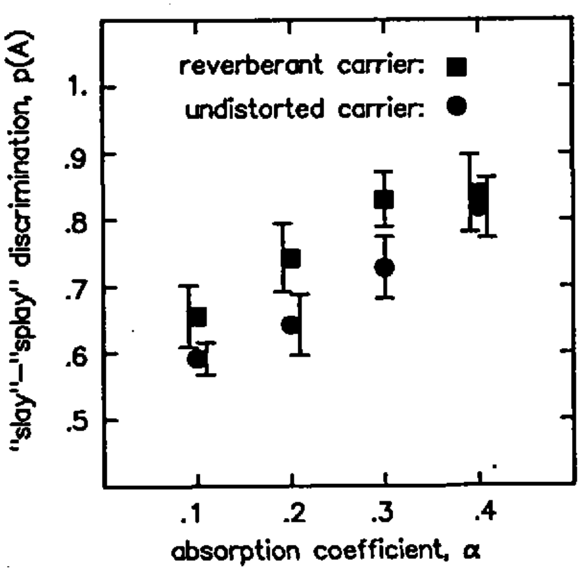
When the sounds contain reverberation the functions are flatter indicating that the extreme stimuli are poorer exemplars of the categories. This effect increases as the reduction in α increases the amount of reverberation. This flattening is reminiscent of that found with natural reverberation in a previous study [13] where the reverberation time was estimated to be in the region of 1 s. This would be generated by values of α between 0.2 and 0.3 with the simulation conditions used in this study, and flattening is correspondingly apparent here at these values. The category boundary, which is the interpolation that corresponds to a rating of 3.5, is near the middle of the undistorted continuum. It changes a little as as the reverberation is increased, in a direction that increases "slay" ratings. This could be because reverberation reduces evidence that the /p/ is present in sounds towards the "splay" end of the continuum, and listeners interpret this as the presence of "slay". However, the corresponding main effect of α is significant but small in experiment 1a, and not significant in experiment 1b.

The flattening of the identification functions is statistically robust. It keeps the category boundary at a fairly central position along the continuum and therefore preserves a fairly even balance of "slay" and "splay" ratings. This could be because listeners respond with "splay" ratings on the basis of weaker evidence of the /p/ when reverberation is also heard. This sort of criterion shift in the presence of reverberation is consistent with the increase in "slay" ratings to stimuli near the "splay" end of the continuum as well as with the attendant increase in "splay" ratings to stimuli near the "slay" end of the continuum.

In summary, reverberation impairs discrimination of "slay" and "splay" by weakening evidence of the /p/ in "splay". This is accompanied by a criterion shift towards "slay" that maintains a noisier, but appropriate separation of the phonetic categories.

4. EXPERIMENT 2

This experiment asks whether perceptual mechanisms compensate for reverberation by strengthening evidence of the /p/ in "splay", making it easier to discriminate from "slay". Compensation for other forms of distortion is evident when a preceding carrier contains the same distortion as the test sound [6,9] so reverberant and undistorted carriers were compared. An attempt was also made to reduce information in the reverberant tail of the test sound to see if compensation comes from this source as well. The rationale for this is that some compensation for spectral-envelope distortion comes from information arriving after the test sound [18].



Discrimination of the end points of the "slay" to "splay" continuum was measured in reverberation with $\alpha = 0.1$, 0.2, 0.3 and 0.4. A bias-free, non-parametric discrimination index, p(A) [19] was calculated from the 5 response criteria that partition the 6-category "slay" to "splay" rating scale to give a score between 0.5 (guessing) and 1.0 (perfect). A preceding carrier that had the same reverberation as the test sound was compared with an undistorted carrier. For half of the subjects the reverberant tail of the test sound was obscured by adding an undistorted recording of "select answer" (spoken by AJW).

Analysis of variance revealed a main effect of α, $F(3,24)=14.2$, $p<0.0001$, reflecting an increase in discrimination score as α is increased. There is also a main effect of the carrier's reverberant content, $F(1,8)=20.89$, $p<0.002$, reflecting the fact that discrimination is easier when the carrier has the same reverberation as the test sound. This indicates perceptual compensation for the effects of reverberation. There were no other significant F-ratios. The effects of α and the compensation effect are shown in the adjoining figure where bars are one standard error on each side of the mean.

The perceptual compensation seen here is reliable and indicates that the effects of degradation by reverberation are reduced when the test sounds are preceded by other sounds that have been distorted in the same way. The difference this makes to the discrimination index might appear small, but the corresponding change in α, for α between 0.2 and 0.3, indicates perceptual enhancement of the ratio of direct to reverberant sound by around 3 dB. This advantage, accrued by a monaural mechanism, is comparable to the advantage of two ears over one for signal detection in reverberation [4].

This compensation for reverberation is consistent with a mechanism that strengthens evidence of the /p/ in "splay". It might also be similar to compensation for spectral-envelope distortion [10] in that it operates at an auditory level and is not restricted to operate only for speech perception. It is therefore appropriate to ask whether non-speech carriers are as effective as speech carriers, whether non-speech distinctions are affected by speech carriers as well as non-speech carriers, whether compensation for reverberation only operates among sounds that are heard to have come through the same transmission channel, and whether peripheral or central mechanisms are responsible.

Perceptual compensation for reverberation gives a discrimination advantage with a bias-free index. It is therefore independent of factors that set systematically different positions for response criteria in different amounts of reverberation (cf. experiments 1a and 1b). However, compensation might not be independent of factors that cause irregular fluctuations in the positions of criteria: such 'criterion noise' reduces the discrimination index and is associated with stimulus uncertainty [19]. This might explain the superiority of the reverberant carrier conditions: Here the subject has time to select appropriate criteria before the test sound arrives. However, when the carrier is undistorted there may be more criterion noise as the subject is uncertain about the distortion in the test sound until after it arrives.

5. ACKNOWLEDGEMENTS

The subjects were recruited and run by Simon Makin whose work was supported by a grant from the Medical Research Council to the author. Linda Shockey suggested that I go for a /p/, avoiding a burst.

6. REFERENCES

[1] V O KNUDSEN, 'The Hearing of Speech in Auditoriums', J Acoust Soc Am, 1 p56 (1929)

[2] A K NABELEK, 'The Effects of Room Acoustics on Speech Perception Through Hearing Aids by Normal-Hearing and Hearing-Impaired Listeners', in G A Studebaker & I Hochberg (eds) Acoustical Factors Affecting Hearing Aid Performance University Park Press, Baltimore (1980)

[3] H G HIRSCH, 'Automatic Speaker- and Speech Recognition in Rooms', in W Ainsworth & J Holmes (eds) Proceedings of the 7th Symposium of the Federation of Acoustical Societies of Europe: Speech '88 Institute of Acoustics, Edinburgh (1988)

[4] A H KOENIG J B ALLEN & D A BERKLEY, 'Determination of Masking Level Differences in a Reverberant Environment', J Acoust Soc Am, 61 p1374 (1977)

[5] J P MONCUR & D DIRKS, 'Binaural and Monaural Speech Intelligibility in Reverberation', J Speech Hear Res, 10 p186 (1967)

[6] M P HAGGARD, 'Selectivity for Distortions and Words in Speech Perception', Br J Psychol, 65 p69 (1974)

[7] G PLENGE, 'On the Differences Between Localization and Lateralization', J Acoust Soc Am, 56 p944 (1974)

[8] J J JETZT, 'Critical Distance Measurement in Rooms from the Sound Energy Spectral Response', J Acoust Soc Am, 65 p1204 (1979)

[9] A J WATKINS & S J MAKIN, 'Perceptual Compensation for Spectral-Envelope Distortion: Effects on Vowels and on Fricatives', Br J Audiol, 26 p186 (1992)

[10] A J WATKINS, 'Central, Auditory Mechanisms of Perceptual Compensation for Spectral-Envelope Distortion', J Acoust Soc Am, 90 p2942 (1991)

[11] H J M STEENEKEN & T HOUTGAST, 'A Physical Method for Measuring Speech Transmission Quality', J Acoust Soc Am, 67 p318 (1980)

[12] B H REPP, 'Limits of the Power of Silence as a Stop Manner Cue', J Acoust Soc Am, 72 pS16 (1982)

[13] A J WATKINS, 'Effects of Room Reverberation on the Fricative/Affricate Distinction', Paper Given at the Second Franco-British Speech Meeting, University of Sussex (1988)

[14] J B ALLEN & D A BERKLEY, 'Image Method for Efficiently Simulating Small-Room Acoustics', J Acoust Soc Am, 65 p943 (1979)

[15] M R SCHROEDER, 'New Method of Measuring Reverberation Time', J Acoust Soc Am, 37 p409 (1965)

[16] M D EGAN, Concepts in Architectural Acoustics McGraw-Hill, New York (1972)

[17] ILS-PC Version 6.1, Signal Technology Inc., Goleta, CA (1989)

[18] A J WATKINS, 'Spectral Transitions and Perceptual Compensation for Effects of Transmission Channels', in W Ainsworth & J Holmes (eds) Proceedings of the 7th Symposium of the Federation of Acoustical Societies of Europe: Speech '88 Institute of Acoustics, Edinburgh (1988)

[19] D MCNICOL, A Primer of Signal Detection Theory, Allen & Unwin, London (1972)