Proceedings of the Institute of Acoustics

VOWEL PERCEPTION AND COLOURATION FROM EARLY REFLECTIONS

A J Watkins

Department of Psychology, Reading University

1. INTRODUCTION

When sounds are played in rooms the direct sound is often accompanied by reflections from the room's surfaces. However, early reflections, arriving up to 50 ms after the direct sound, are not generally heard separately. They seem to become perceptually integrated with the direct sound to give the impression of a single overall sound [1]. In addition, the perceived direction of this overall sound is largely determined by the direction of its first arriving part, giving the well-known precedence effect [2]. Thus, the directional influences of the early reflections seem to be suppressed [3].

Early reflections can also give rise to filtering effects that alter the spectral envelope of sounds. The perceptual consequences of this might be changes of timbre or 'colouration'. This could serve to provide information about the acoustic properties of the environment [4,5]. Conversely, there may be distorting effects, which are detrimental when the sound's spectral envelope contains characteristics that are important for identification. Characteristics that might be vulnerable in this way include those of musical instruments [6] as well as those of vowel and consonant sounds in speech [7,8]. Some authors have suggested that there is some perceptual suppression of colouration, or 'de-colouration', which accompanies the perceptual integration of early reflections with the direct sound [9,10,11]. The present experiment was designed to test this idea. It measures the perceptual consequences of colouration brought about by simulated early-reflections that follow the direct sound. It asks whether there is more colouration when the temporal order of the direct sound and reflections is reversed. If the perceptual integration of early reflections is accompanied by some de-colouration, then there will be more colouration in the condition where the reflection-pattern arrives first.

Colouration effects are measured here by listeners' identifications of a vowel to vowel continuum between "itch" and "etch". These sounds are played through a two-part filter that simulates a direct sound followed by a reflection pattern. The reflection pattern in this simulation is the impulse response of one of two 'spectral difference' filters [12]. Spectral difference filters are formed by subtracting the spectral envelope of one vowel from that of another. Here, these two vowels are those at the ends of the itch to etch continuum. The two filters have opposing effects on the spectral envelopes of the continuum's members, making them more like one or other end-point vowel. In the listening test, the total number of itch responses is found for each continuum. The 'phoneme boundary' is calculated from this total, indicating the continuum step (or point between steps) where listeners' responses switch from itch to etch. The filters' perceptual effects are judged by comparing the phoneme boundaries for the differently filtered continua.

2. METHOD

2.1 Stimuli

2.11 Continuum. A continuum was formed between "itch" /ɪtʃ/ and "etch" /ɛtʃ/. Recordings of tokens of these end-point words and of /ʌtʃ/ were obtained from a male speaker (AJW). Recordings were made in an IAC 1201 booth using a Sennheiser MKH 40 P48 microphone. These signals were amplified (Revox A77), low-pass filtered at 9 kHz with a 48-dB per octave cut-off 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 (Signal Technology, 1989) running on a Victor PC386 computer.

The recording of /ʌtʃ/ was divided into a 170-ms vowel, and a closure with affricate /tʃ/ lasting 340 ms. Digital waveforms of these sounds and of the tokens of end-point vowels were transferred to a Sun Sparcstation computer for processing.

The signal processing methods that are used here to form the continuum are similar to those described in Watkins (1991) [12], except that the filtering operations used the zero-phase methods described in Watkins and Makin (1994) [13] to preserve the time-alignment of a filter's input and output.

Spectral envelopes of the end-point vowels were obtained from an LPC lattice analysis of 16-ms segments of the recorded vowels, with segments centred at 28 ms from the vowels' nominal onset.

The test-vowel continuum was derived from the recording of /ʌ/ in /ʌtʃ/. It was spectrally flattened before being passed through a 'reshaping' filter. The vowel of the "itch" end-point was generated by using the spectral envelope from /ɪ/ for this reshaping filter's response. Similarly, using the spectral envelope from /ɛ/, the vowel for the "etch" end point was generated.

Intermediate continuum steps were obtained with responses for the reshaping filter that were found by interpolating between the spectral envelopes of the end points, as follows. A difference of spectral envelopes was obtained by subtracting the dB values of the /ɛ/ envelope from the dB values of the /ɪ/ envelope before multiplying by the interpolation value. This difference of spectral envelopes was then added to the /ɪ/ spectral envelope (in dB) to obtain the interpolated spectral envelope, which was then used for the reshaping filter's response. Interpolation values between 0.1 and 0.9 in steps of 0.1 were used together with the end points (interpolations = 0.0 and 1.0) to give an 11-step continuum. The overall level (in dB) of each of these test-vowels was scaled to match the level of the /ʌ/ in /ʌtʃ/. The /tʃ/ from /ʌtʃ/ was appended to each of these test-vowels to form the words of the continuum.

2.12 Two-part filters. In the impulse responses of these filters, a unit sample effects the 'direct' sound. This is the first part of the filter's impulse response in a forwards condition, where the second part of the impulse response follows after a brief delay. Delays of 0.5, 1.0 2.0 ms were used. The impulse response of one of these filters is shown in figure 1.

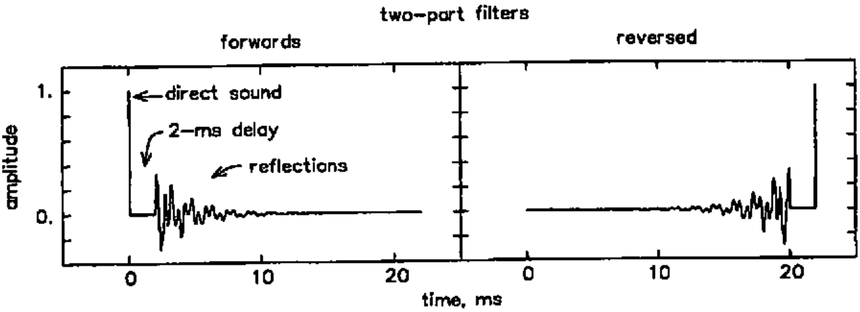

Figure 1. The impulse response of a two-part filter. Here, the delay between the direct sound and the reflection pattern is 2 ms, and the reflection pattern is the /ɛ/ minus /ɪ/ spectral-difference filter. This two-part filter's impulse response in the reversed condition is also shown.

The second part of the filter is one of two 20-ms reflection patterns, which are the impulse responses of 'spectral difference' filters. One of these was designed so that it changed the spectral envelope of the vowel at the /ɪ/ end of the continuum into the spectral envelope of the vowel at the /ɛ/ end of the continuum. The other filter was the opposite. These patterns are shown in figure 2. The Fourier transforms of these reflection patterns are therefore the difference between the spectral envelopes of the vowels /ɪ/ and /ɛ/, so the two patterns are designated /ɪ/ minus /ɛ/, and /ɛ/ minus /ɪ/.

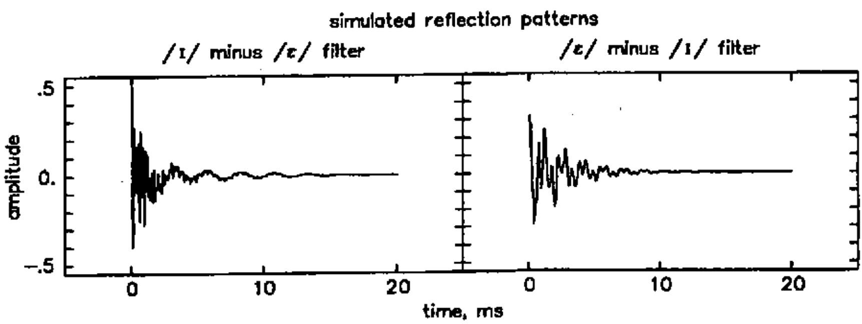

Figure 2. The two simulated reflection-patterns. These are the impulse responses of the spectral-difference filters that are indicated.

The test words were filtered by a single, linear convolution with the impulse response of a two-part filter. Thus, the filtering effects arise as they do from room reflections. In a reversed condition, the entire impulse response of the two-part filter was time reversed before the convolution (figure 1).

Informal listening-tests confirmed that the reflections were not heard separately from the direct sound when the test words were played through the two-part filters.

2.2 Procedure

Sounds were delivered diotically to listeners on-line, under the control of the PC386 computer. Analogue signals were generated from the digital waveforms with 16-bit resolution at a conversion rate of 20 kHz per channel (Data Translation DT2823) using the ILS program LDA. These signals were low-pass filtered at 9 kHz with a 48-dB per octave cut-off slope (Kemo VBF8) and presented to listeners with Sennheiser HD480 headphones at 53-dB sound pressure level in an IAC 1201 booth.

A word from a continuum was presented on each trial and listeners identified the word by pressing either a button labelled itch or one labelled etch. 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. Thus, a response was obtained for each trial. A minimum inter-trial interval of 4 seconds was enforced.

There were 12 untrained listeners who reported no hearing problems. The delay between the direct sound and the reflection pattern was 0.5 ms for one group of 4 listeners, 1.0 ms for the second group of 4 listeners, and 2.0 ms for the third group. Each listener heard the test words from 4 continua, with 2 played through two-part filters having the /ɪ/ minus /ɛ/ reflection pattern and 2 with the /ɛ/ minus /ɪ/ reflection pattern. Each of the 2 reflection patterns was used once in a forwards two-part filter and once in a reversed two-part filter. Each of these sounds was presented once giving 2 filter directions (forwards or reversed) x 2 reflection patterns x 11 continuum steps = 44 trials for each of the listeners. These trials were administered without feedback or practice trials in one session that lasted about 10 minutes. Different random orders of trials were used for each listener.

Phoneme boundaries for each listener were calculated from the total number of itch responses to members of the continuum. This total is between 0 and 11, and is converted to a test-vowel interpolation by subtracting 0.5 before dividing by 10. This then gives the test-vowel interpolation where the switch from itch responses to etch responses occurs, which is the phoneme boundary. These boundaries therefore lie between -0.05 (etch response on all trials) and 1.05 (itch response on all trials).

3. RESULTS

Figure 3 is an example of listeners' identifications of two of the continua, showing the phoneme boundaries.

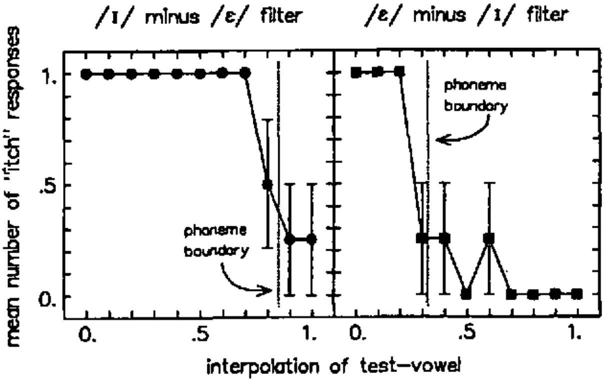

Figure 3. The number of "itch" responses to the continuum's words when they are filtered by the two different reflection-patterns. In this example a forwards filter is used, and the delay between the direct sound and reflections is 1 ms. Data points are means from the 4 listeners, and bars are one standard error on each side of the mean.

Figure 4 shows the phoneme boundaries for all of the continua, averaged across the 4 listeners in each of the 3 groups.

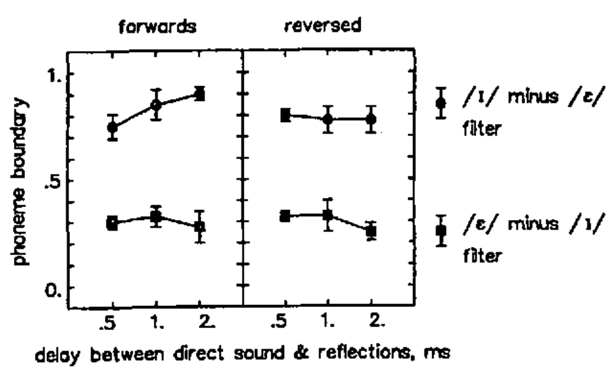

Figure 4. Phoneme boundaries from all of the continua used in the experiment. Data points are means from the 4 listeners in each of the 3 delay-conditions. Each listener heard sounds filtered by forwards and reversed two-part filters, and they heard both types of reflection pattern in these filters. Bars are one standard error on each side of the mean.

A 3-way analysis of variance was performed on the phoneme boundaries. There is one between-listener factor (delay of reflection pattern, 3 levels) and two within-listener factors (spectral-difference filter, 2 levels; forwards versus reversed, 2 levels). This analysis revealed a main effect of spectral difference filter, $F(1,9)=603.41$, $p<0.0001$. This indicates that the vowels in the test words are heard to be more like the positive vowel of the spectral-difference filter that is applied to them. There were no other significant F-ratios.

Listeners' identifications of these sounds indicate that all the two-part filters influence the vowel that is played through them. More "itch" sounds are reported for one spectral-difference filter, giving a relatively high phoneme boundary that lies towards the etch end of the continuum. The other spectral-difference filter gives much lower phoneme boundaries. This difference indicates that there is substantial colouration from the reflection patterns used here.

There is little sign of any suppression of colouration that is greater in the forwards condition. If this were the case then the difference between phoneme boundaries for the two spectral-difference filters should be larger in the reversed condition. However, this phoneme-boundary difference is about the same when the entire impulse response of the two-part filters is reversed. Thus, there is no evidence here that colouration from early reflections is perceptually suppressed when the reflections follow after an 'uncoloured' direct sound.

4. DISCUSSION

Reflections of sound become mixed with the direct sound and they influence perception in several ways. The various effects seem to arise from somewhat different perceptual mechanisms. Reflected sound can cause masking, spatial influences, colouration, loudness influences and other effects, while different aspects of the signal can separately affect these perceptual factors. For example, with two-part filters similar to those used in the present experiment it has been found that the *spatial* influences of the reflection pattern are minimal. In those experiments the overall sound is heard to come from the direction cued by the direct sound's interaural delay, regardless of the interaural delay of the reflections [14]. Thus, the precedence effect happens with these stimuli. Despite this, the colouration effects of such reflections are quite marked, as we have seen. Similarly, early reflections seem to have a substantial influence on sounds' loudness [15]. In that experiment listeners could detect intensity changes in a direct sound as easily as they could detect intensity changes in an early reflection, and this ability was attributed to the detection of loudness changes in the overall sound.

It might be considered advantageous for hearing to behave in the way described here. Early reflections contain misleading information about the sound's direction, so it makes sense for the system to inhibit this information and to favour the generally accurate directional information in the sound's direct part. At the same time, early reflections might contain other, more useful information about the sound source and maybe about the acoustic environment. For these reasons some authors have argued that it is entirely advantageous to incorporate non-directional information from echoes [15]. There are some disadvantages though, as the colouration that carries this sort of information is likely to have distorting effects on sounds' characteristics, and this may well hamper perception. We have seen here that this type of distortion is not overcome by mechanisms that bring about the perceptual integration of early reflections, and it does not seem to be overcome by binaural mechanisms either [16]. However, there is evidence of other perceptual mechanisms that can overcome such distortion, and these seem to extract information from neighbouring sounds to bring about a perceptual compensation. Earlier arriving sounds are particularly influential in this respect [12].

Compensation effects from neighbouring sounds might have influenced the phoneme boundaries in the present experiment. This is because filtering was applied not only to the vowel of the test words, but also to the neighbouring /tʃ/. However, the compensation effected by these sort of later-arriving sounds is much smaller than that from preceding sounds. Measurements of this sort of compensation indicate that it is unlikely to have moved phoneme boundaries in the present experiment by more than about 0.1 [8].

The spectral difference filters used here are not as effective when they are combined with a direct sound in the two-part filters. Without the direct sound they would shift the phoneme boundaries beyond the ends of the continuum. Adding the direct sound reduces the 'spectral contrast' of the filters, so that the dB-range of their frequency responses is compressed. This probably accounts for most of the filters' reduced effectiveness.

Adding the direct sound to the spectral difference filters also introduces a 'comb-filtering' effect, so that there are additional peaks and notches in the filters' frequency responses. This might also have a small influence on the phoneme boundaries measured here. The frequencies at these peaks and notches vary with the delay between the direct sound and the reflections' onset. However, the results show that this variation had little influence on the corresponding phoneme boundaries.

5. CONCLUSION

Despite the perceptual integration of early reflections with the direct sound there remain substantial colouration effects from the reflections. These effects are just as prominent when the reflection pattern is caused to precede the direct sound, indicating an absence of 'decolouration' when the reflections follow after an 'uncoloured' direct -sound.

6. REFERENCES

[1] H Haas, 'Uber den Einfluss des Einfachechoes auf die Horsamkeit von Sprache', Acustica, 1 pp49-58 (1951)

[2] H Wallach, E B Newman & M R Rosenzweig, 'The Precedence Effect in Sound Localization', Am J Psych, 62 pp315-337 (1949)

[3] P M Zurek, 'The Precedence Effect', in *Directional Hearing* edited by W A Yost & G Gourevitch (Springer, New York) pp85-106 (1987)

[4] A H Benade, *Fundamentals of Musical Acoustics* (Oxford, New York, 1976)

[5] R K Clifton & R L Freyman, 'The Precedence Effect: Beyond Echo Suppression' in *Binaural and Spatial Hearing in Real and Virtual Environments* edited by R H Gilkey & T A Anderson (Lawrence Erlbaum, New Jersey) pp233-255 (1997)

[6] J-C Risset & D L Wessel, 'Exploration of Timbre by Analysis and Synthesis', in *The Psychology of Music* edited by D Deutsch (Academic, New York) pp26-58 (1972)

[7] M P Haggard, 'Selectivity for Distortions and Words in Speech Perception', Br J Psychol, 65 pp69-83 (1974)

[8] A J Watkins & S J Makin, 'Some Effects of Filtered Contexts on the Perception of Vowels and Fricatives', J Acoust Soc Am, 99 pp588-594 (1996)

[9] J Blauert, *Spatial Hearing* (MIT, London, 1997)

[10] W M Hartmann, 'Listening in a Room and the Precedence Effect', in *Binaural and Spatial Hearing in Real and Virtual Environments* edited by R H Gilkey & T A Anderson (Lawrence Erlbaum, New Jersey) pp191-210 (1997)

[11] B C J Moore, *Introduction to the Psychology of Hearing* (Academic, London, 1997)

[12] A J Watkins, 'Central, Auditory Mechanisms of Perceptual Compensation for Spectral-Envelope Distortion', J Acoust Soc Am 90, pp2942-2955 (1991)

[13] A J Watkins and S J Makin, 'Perceptual Compensation for Speaker Differences and for Spectral-Envelope Distortion', J Acoust Soc Am 96, pp1263-1282 (1994)

[14] A J Watkins, 'Identification and lateralisation of a vowel played through binaural spectral difference filters', Br J Audiol 32, pp101-102 (1998)

[15] R L Freyman, D D McCall, & R K Clifton, 'Intensity Discrimination for Precedence Effect Stimuli', J Acoust Soc Am 103, 2031-2041 (1998)

[16] A J Watkins, 'The Precedence Effect and Perceptual Compensation for Spectral-Envelope Distortion' in *Psychophysical and Physiological Advances in Hearing* edited by A R Palmer, A Rees, A Q Summerfield & R Meddis (Whurr, London) pp336-343 (1998)