COLORATION AND SPEECH PERCEPTION

AJ Watkins The University of Reading, Department of Psychology, Reading.

1. SUMMARY

When a sound travels through the transmission channel between the source and the listener its spectral envelope is often distorted because of the channel's uneven frequency-response characteristic. This brings about 'coloration' of the sound that the listener receives. Transmission channels that color sounds in this way can be 'natural', such as a room with its reflecting surfaces, or artificial, such as a telephone line. The consequences of coloration for the perception of speech are considered in the light of perceptual experiments on the identification of speech sounds that are distinguished primarily by characteristics of their spectral envelopes.

2. THE HAAS EFFECT AND THE PRECEDENCE EFFECT

Auditory space may be distinguished from visual space in several ways. One way in which auditory space is distinctive concerns the perception of sound that is reflected from surfaces in the listening space. The auditory perception of such reflections seems to be quite unlike the visual perception of reflected light. Consider some of the perceptual properties of visual reflections, as for example, when a house is seen from across Lake Windermere on a still day. The house's reflection in the lake is seen as separate from the house itself, and in a different direction. Also, if the reflection is obscured while the house remains visible, the house looks the same. The auditory perception of reflected sound exhibits none of these properties. This was shown in part by Haas [1], who studied the perception of speech that contained a single echo. He found that this reflection became fused with the direct sound so that only one sound was heard. This fusion is obtained for echo delays up to about 20 ms [2]. Haas also observed that the reflection contributes substantially to the loudness of the speech, and that with suitably short delays it can even improve intelligibility a little. This 'Haas effect' is often accompanied by a qualitatively different binaural effect, whereby the directional information in the first arriving, 'direct' part of the sound dominates the direction heard for the overall sound. This 'precedence effect' was demonstrated by Wallach et al. [3] who presented a brief click over headphones, along with an echo of the first click that was delayed by 2 ms. The experiment varied the interaural time difference of either the echo click, or the first click, and It was found that varying only the echo in this way had hardly any effect on the direction heard for the whole sound. On the other hand, large changes in the overall sound's direction were heard when the interaural time difference of the first click was altered. The Haas effect and the precedence effect can often occur together in various sounds, and they have sometimes been considered as different aspects of the same basic phenomenon. However, the Haas effect can occur at relatively long time-intervals between the direct sound and echo, where the precedence effect is less apparent. So it is probably best to consider the two effects separately [4, 5].

3. COLORATION EFFECTS

A further property of the perception of a sound with reflections is the 'coloration' that the reflections contribute. Again, this property is not shared by visual perception of the house and its reflection in the still lake. If a breeze should ruffle the surface of the lake, the reflection is distorted, but perception of the actual house is unchanged by this. However, when reflections of sound are distorted versions of the direct sound, the distortion does seem to affect the quality of the overall sound. This was shown in a lecture room demonstration referred to by Gardner [6]. The demonstrator mingled with the audience while holding a miniature speaker. This poor quality speaker could only deliver the higher frequencies properly so it would normally give a tinny sound. However, this little speaker was wired up so that it played the same signal as a large high quality loudspeaker that was placed in a fixed position at the front of the lecture room. It seemed to listeners that all the sound in the room was coming from the little speaker, as long as it was closer than the large speaker was. At the same time, the sound that this little speaker seemed to be producing had the rich and full qualities of the large loudspeaker that was more distant. Thus, the Haas effect occurs so that only one overall sound is heard, the precedence effect occurs so that the overall sound's direction is determined by the first sound that arrives, while the later-arriving 'reflections' contribute to the overall sound's 'colour'. Consequently, it was only when the demonstrator switched off the high quality speaker that the tinny quality of the miniature speaker could be heard.

4. BINAURAL HEARING

In certain circumstances binaural hearing mechanisms can reduce coloration, as was shown by Zurek [7]. He considered the dichotic situation that arises when a flat-spectrum noise-source is located straight ahead of the listener, and a single reflection comes to the listener from one side. The reflection considered was specular, which means that all frequencies are reflected by an equal amount. Adding such a reflection brings about spectral-envelope perturbations in the signals at each ear. These 'combfilter' patterns have a peak at a frequency that is the inverse of the reflection's latency as well as peaks at integer multiples of this frequency. The valley floors (tips of the comb's teeth) are at frequencies half way between the peaks (on a linear frequency scale). The reflection's latency will be different at each ear because its angle of arrival adds an interaural delay to the reflection's latency at one ear. This reduces the interaural correlation of the sound's spectral envelopes in a way that depends on the reflection's latency and its angle of arrival. Zurek [7] suggested that this might give rise to suppression of coloration if the monaural representations of spectral magnitude are summed binaurally. When such a summation is applied to combfilter patterns that arise from reflection latencies that differ by an interaural delay, it will give rise to a smoother spectral-envelope than that of either of the monaural patterns alone. This idea was supported by Zurek's [7] experiments where the reflection's amplitude was varied, and listeners were required to detect its presence. The reflection was much harder to detect in dichotic conditions than it was in diotic control conditions where each ear received the same combfilter pattern. This was true for reflection latencies shorter than about 5 or 10 ms when the interaural delay was 0.5 ms.

The sort of binaural mechanism that Zurek [7] points to here is limited in its ability to compensate for the spectral-envelope distortion that arises from reflections. One problem is that spectral smoothing by binaural summation can never completely eliminate the distortion. Perfect smoothing would arise if the spectra at each ear were the exact inverse of each other, but interaural decorrelation of spectra that arises through the introduction of an interaural delay is very unlikely to have this precise effect. Furthermore, the decorrelation that interaural delays do introduce depends on the angle of the reflections' arrival, while the distortion itself is brought about by the reflections' delay pattern. As these two factors are largely independent in everyday reverberation patterns it follows that there will commonly be combinations of angle of arrival and reflection pattern that give little or no reduction in coloration after binaural summation.

A further limitation of binaural summation is that it does not reduce coloration effects that arise from reflections that are non-specular, i.e., when a sound's frequencies are reflected by unequal amounts. Spectral perturbations that arise in this way will be interaurally correlated, so that the resulting coloration effects will persist when there is binaural summation.

Nevertheless, some authors have suggested that there is some binaural suppression of coloration, or 'de-coloration', that occurs even with reflection patterns that give rise to interaurally correlated distortions of the spectral envelope [8,5]. These authors observe that monaural listening in everyday room conditions can sometimes make coloration seem more apparent than it is when listening binaurally. If there really is some suppression here then it might happen when the precedence effect occurs [9]. Another possibility is that de-coloration goes hand in hand with the Haas effect, so that it happens when there is perceptual integration of early reflections with the direct sound. Watkins [10] tested these ideas in an experiment that measured the effects of coloration on vowel quality. A continuum between two vowels was produced and then a 'two-part' filter distorted each step. The beginning of this filter's unit-sample response simulated a direct sound with no distortion of the spectral-envelope. The second part simulated a reflection pattern that distorted the spectral envelope. A delay between the direct part and the reflections' part simulated the travel-time difference between direct and reflected sound. The reflections' frequency response was designed to give the spectral envelope of one of the continuum's end-points to the other end-point. Filtered sounds were presented over headphones for listeners to identify. It was found that the reflections in two-part filters had a substantial influence because sounds tended to be identified as the positive vowel of the reflection pattern. This effect was not reduced when the interaural delays of the reflections and the direct sound were substantially different. Also, when the reflections were caused to precede the direct sound, the effects were much the same. By contrast, in measurements of lateralization of these sounds the precedence effect was obtained. Here, the lateral position of the whole sound was largely governed by the interaural delay of the direct sound, and was hardly affected by the interaural delay of the reflections.

The dissociation between coloration effects and spatial effects seen here would seem to arise because of the way that different parts of the signal are processed by the different perceptual mechanisms that are involved. Perception of the vowel's quality involves processing relatively long portions of the signal, and this type of processing is characterised by binaural summation across different interaural delays. Thus, early reflections contribute to this attribute of the sound. On the other hand, the lateral position of these sounds seems largely to be governed by the characteristics of their onsets, as long as these are sufficiently sharp. Thus, the sort of processing involved in establishing the lateral position of the sound uses only short parts of the signal, and this processing is characterised by dominance of the onset's interaural delay over the subsequent interaural delays in early reflections [11].

It might be considered advantageous for binaural 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 suppress this information, while favouring the generally accurate directional information in the sound's direct part. The precedence effect might therefore be taken as an indication that binaural mechanisms are good at compensating for spatial distortion that might arise from early reflections. At the same time, binaural mechanisms appear to be limited in their ability to compensate for the spectral envelope perturbations that arises from these reflections and that give rise to coloration. The perceptual experiments described above indicate that early reflections, arriving with a spread of different interaural delays, will all be perceptually incorporated with the direct sound for the purpose of determining the sound's spectral envelope. However, early reflections do contain other information about the sound source. Indeed, without the reflected sound in rooms many sound sources would be completely inaudible. Coloration from these reflections might also provide information about the acoustic properties of the environment [12, 13]. Thus, it makes some sense to incorporate non-spatial information from reflected sound [14].

A disadvantage of hearing coloration is that it arises from distorting effects on the sounds' spectral envelope, but this envelope also carries information about the sound's source. For example, the source's resonances transmit information about the shape and size of the structures that produce them and impose characteristic envelopes upon the acoustic spectra of short segments of the auditory signals that arise. For this reason, descriptions of spectral envelopes, among other features, are widely held to be prereqisites of auditory identification. Thus, the distortion that gives rise to coloration might affect characteristics of musical instruments [15] as well as characteristics of vowel and consonant sounds in speech [16,17]. We have seen that this distortion is not overcome by binaural mechanisms, and that it is not overcome by mechanisms that bring about the perceptual integration of early reflections seen in the Haas effect.

5. INFLUENCES OF NEIGHBOURING SOUNDS

There do appear to be other kinds of perceptual mechanism that can compensate for distortions of the spectral envelope. These mechanisms appear to extract information about the distortion from neighbouring sounds in order to compensate for the distortion's effects on the spectral envelopes of subsequent and preceding sounds. The presence of such mechanisms could account for the apparent robustness of speech perception in the presence of the prominent spectral-envelope distortions that arise in everyday transmission channels such as telephone lines and reverberant rooms [18]. Much of the evidence of these compensation mechanisms comes from experiments in which a test vowel is preceded by a filtered precursor-phrase, with the result that the test vowel is heard as one that has been filtered by the inverse of the precursor's filter [19].

5.1 Central Compensation Mechanisms

In the experiments with filtered precursor phrases, a small part of the effect on the test vowel was found to be similar to peripheral, adaptation-like phenomena, such as the negative auditory after-image [20, 21]. However, effects with precursors that contained speech-like spectro-temporal variations were much larger than effects in control conditions where noise precursors were used. These noise precursors had time-stationary spectral characteristics, but their effects should have been much the same as those of speech precursors if peripheral mechanisms were the sole source of the compensation effects.

The bulk of the effect with speech precursors appears to arise from mechanisms that are more central than those responsible for effects with noise precursors. Comparing results from two sorts of experiment led to this conclusion. In one of these experiments the precursors were either presented to the same ear as the test sounds, or the test sounds were presented to the opposite ear. In the other sort of experiment sounds were presented binaurally. In one binaural condition precursors were given the same interaural delay as the test sound, so that they were heard to come from the same direction. In the other binaural condition precursors were given a different interaural delay to the test sounds, so that they were heard to come from a different direction. The results with noise precursors typified a peripheral mechanism. This is because their effects were abolished in different-ear conditions, but were uninfluenced by differences in interaural delay in the binaural conditions. However, effects with speech precursors seemed to typify a more central mechanism. This is because their effects were reduced, but not abolished, in the different-ear condition, while there was a similar reduction across binaural conditions when the interaural delays of the precursor and test sound were changed from same to different. Thus, it would seem that this central mechanism is 'smart', in that it reduces its compensation effects when direction differences indicate that the precursor and test sounds have arisen from different transmission channels.

Other results also indicate that a central mechanism is involved. They come from experiments where the identification of test sounds is influenced by the filtering of subsequent sounds when there are no precursors. For example, Watkins [22] found that an /ɪt/ to /ɛt/ continuum of test sounds was influenced by filtering the /t/ and a subsequent phrase, "is the next word". The direction of this influence did indicate perceptual compensation for the effects of the filter, although this influence was much smaller than that of filtered precursors.

5.2 The Inverse-Filter Model

In the experiments that measure compensation, the effect on the test-sound is like the effect of applying the inverse of the filter that the neighbouring sounds are played through. Such 'inverse filtering' resembles techniques used to remove transmission channel characteristics in voice recognition automata [23, 24]. One way to get an appropriate frequency response for the inverse filter is to compute the whole signal's long-term average spectrum and then invert it. Watkins and Makin [25] used this method in a simulation of the perceptual compensation mechanism. They found that test sounds played through the simulation were perceptually altered in ways that resemble the alterations heard in test sounds played after filtered precursors. So it would seem that this sort of simulation has characteristics that are similar to those of the perceptual compensation mechanism.

A perceptual mechanism based on inverse filtering would provide the resistance to spectral-envelope distortion exhibited in speech perception, but such resistance is not necessarily achieved in this way. It might be that speech is perceptually coded in terms of features, such as peaks in the spectral envelope, and that these retain their essential properties better than others do over the range of ecologically likely conditions of distortion [26]. Peaks might plausibly function in this way because alterations to them have much larger effects on phonetic quality than alterations to other parts of the spectral envelope [27, 28]. Furthermore, the frequency location of peaks is one of the major determinants of perceived similarity between sounds such as vowels [29]. However, simply extracting the features is not sufficient to overcome all distortions, as spurious features are likely to arise from the transmission channel [30]. Features imposed by the channel need somehow to be separated from those of the sound source. One basis for such a separation might be the presence or absence of similar features in neighbouring sounds. In experiments with precursors it may be that features of the filter that are present in both the precursor and the test sound come to be separated from other features in the test sound. Such a separation might come about in ways reminiscent of a perceptual grouping perhaps [31] or of the selective adaptation of auditory features [32].

Altering the spectral contrast of sounds can be used to test whether features in the spectral envelope are perceptually significant. Contrast is changed when a positive number, other than one, is used to multiply decibel values of the spectral envelope. This varies the difference in level between peaks and valleys, but features such as peaks stay at the same frequencies. If contrast-invariant features are more important than other parts of the spectral envelope, then contrast changes will have relatively little effect on perception.

Watkins and Makin [33] varied contrast to ask whether compensation for spectral envelope distortion involves the extraction of features that are contrast invariant. The experiments used a filtered precursor-phrase followed by a word containing a vowel test-sound, which was drawn from a continuum between /ɑpt/ and /ɔpt/. The contrast of the precursor's filters was altered by multiplying their frequency response by a positive number, other than one, while the contrast of test-sounds was altered by a similar multiplication that was applied to the spectral envelope. Compensation was measured when the test-sound's contrast was the same or above that of the precursor's filter, as well as when the test sound's contrast was reduced to a value below the precursor's contrast. These manipulations should have little effect on compensation if it is only the contrast-invariant features that are involved.

However, it was found that these contrast manipulations had substantial effects on perceptual compensation for spectral-envelope distortion. The different contrasts gave rise to compensation that generally increased and decreased with the ratio of precursor contrast to test-sound contrast. This was the case as long as the precursor's contrast was not too high. The effects were generally larger than those to be expected from peripheral mechanisms and appeared to be caused by the more central, auditory mechanism that was responsible for compensation effects in earlier studies.

This pattern of results seems to rule out the possibility that the compensation mechanism involves feature extraction. Compensation for spectral-envelope distortion appears then to precede any extraction of features in the spectral envelope rather than to occur at a subsequent stage. Either of these arrangements could in principle avoid the influence of features that are added by distortion [34,30]. However, an advantage of a compensation mechanism that precedes feature extraction is that correction can be made for distortions of existing features, such as when a spectral-envelope peak is displaced to a different frequency. This can occur when the distortion does not introduce any new features, as happens when sounds are high-pass or low-pass filtered. Compensation can correct for this kind of distortion as it varies with contrast, so that its effects are appropriate for the degree of displacement of the original features.