ment); 23% by drift (21.8% frequent drift, 1.3% rare drift or rafting), and 4% by wind.

His figures and mine are close for overall dispersal by birds, but they are reversed for wind dispersal and drift. Although floating vegetation originating on the mainland has been reported from the vicinity of the archipelago[21–23], or even cast up on its beaches[18,24], drift has had a minor role in plant colonisation of the Galapagos. Those species that have arrived in this way are all members of the mangrove, beach or salt flat associations. I can find no evidence for any vascular plant species having arrived by rafting.

The small number of species that have been derived through oceanic drift is not surprising. Although drift-dispersed species have relatively high immigration rates in comparison with those which are dispersed by birds and wind, very few species are adapted to this mode of introduction. In addition, their higher immigration rates and the small numbers of habitats available to them in the Galapagos also reduce the opportunities for endemism in this group[25]. The role of man in the introduction of exotic species into the islands is surprising (Table 1). Even Hooker[15] remarked on the number of alien species that man had added to the flora by the time of Darwin's visit in 1835. Man has introduced 124 weeds by the latest count, and 57 of his cultivated exotics have escaped and now reproduce themselves in the wild. Today *Homo sapiens* has replaced birds as the most important factor in the dissemination of plants to the Galapagos. The deleterious effects on the native flora of some of the species introduced by man are just beginning to be appreciated[6,19,26–28].

Recent potassium–argon studies[29] indicate that the archipelago has a probable maximum age of 3 Myr. This means that the arrival and establishment of one successful disseminule about every 7,900 yr would account for the present indigenous flora. Introduction and subsequent extinction cannot be estimated, but they certainly have occurred as well. This is a higher rate of introduction than Fosberg's[30] estimate of one successful disseminule every 20–30,000 yr to account for the derivation of the indigenous vascular flora of the Hawaiian Islands, but the latter are much further from their source areas than are the Galapagos.

Such a relatively high rate of introduction is not surprising. These volcanic islands are a jumble of open pioneer habitats inhibited by a weedy flora. As others have pointed out, dispersal is only half the battle; establishment is the other half. In the Galapagos, those plants that have become established are almost all weedy, a phenomenon first recognised by Darwin[1]. The chance of immigrants surviving in any specific locality is greater initially than later when more closed communities have evolved. Weedy plants, being adapted to open habitats, have been at an advantage when their disseminules reached the Galapagos Islands. Their offspring have given us the present flora.

C. S. Adkisson, S. Carlquist, P. R. Grant, F. James, A. E. Linkins, R. Mayes, R. A. Paterson, S. H. Porter, D. A. West and I. L. Wiggins all made helpful comments on an earlier version of this communication.

DUNCAN M. PORTER

Department of Biology,
Virginia Polytechnic Institute and State University,
Blacksburg, Virginia 24061

Received September 20; accepted November 9, 1976.

[1] Darwin, C., *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle, Between the Years 1826 and 1836, Describing Their Examination of the Southern Shores of South America, and the Beagle's Circumnavigation of the Globe*, **3**, *Journal and Remarks, 1832–1836* (Colburn, London, 1839).
[2] Svenson, H. K., *Am. J. Bot.*, **22**, 208–277 (1935).
[3] Wiggins, I. L., and Porter, D. M., *Flora of the Galapagos Islands* (Stanford University Press, Stanford, 1971).
[4] Eliasson, U., *Bot. Notiser*, **123**, 346–357 (1970).
[5] Eliasson, U., *Bot. Notiser*, **125**, 320–322 (1972).
[6] Eliasson, U., *Op. bot. Soc. bot. Lund.*, **36**, 1–117 (1974).
[7] Schofield, E. K., *Biol. Conserv.*, **5**, 48–51 (1973).
[8] Weber, D., *Bull. Soc. neuchâtel. Sci. nat.*, **96**, 17–30 (1973).
[9] Hamann, O., *Bot. Notiser*, **127**, 245–251 (1974).
[10] Hamann, O., *Bot. Notiser.*, **127**, 252–255 (1974).
[11] Hamann, O., *Bot. Notiser.*, **127**, 309–316 (1974).
[12] Wiggins, I. L., *Madroño*, **23**, 62–64 (1975).
[13] Carlquist, S., *Bull. Torrey bot. Club.*, **94**, 129–162 (1967).
[14] Carlquist, S., *Island Biology* (Columbia University Press, New York, 1974).
[15] Hooker, J. D., *Trans. Linn. Soc., Lond.*, **20**, 235–262 (1847).
[16] Robinson, B. L., *Proc. Am. Acad. Arts Sci.*, **38**, 77–269 (1902).
[17] Stewart, A., *Proc. Calif. Acad. Sci.*, Ser. 4, **1**, 7–288 (1911).
[18] Stewart, A., *Pl. Wld.*, **18**, 192–200 (1915).
[19] Thornton, I., *Darwin's Islands, A Natural History of the Galapagos* (Natural History Press, New York, 1971).
[20] Harris, M., *A Field to the Birds of Galapagos* (Collins, London, 1974).
[21] Agassiz, A., *Bull. Mus. comp. Zool. Harv.*, **23**, 1–89 (1892).
[22] Beebe, W., *The Arcturus Adventure* (Putnam, New York, 1926).
[23] Townsend, C. H., *Bull. N. Y. zool. Soc.*, **31**, 148–168 (1928).
[24] Fitz-Roy, R., *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle*, **2**, *Proceedings of the Second Expedition, 1831–1836, Under the Command of Captain Robert Fitz-Roy, R. N.* (Colburn, London, 1839).
[25] Johnson, M. P., and Raven, P. H., *Science*, **179**, 893–895 (1972).
[26] Eckhardt, R. C., *BioScience*, **22**, 585–590 (1972).
[27] Porter, D. M., *BioScience*, **23**, 276 (1973).
[28] Hamann, O., *Biol. Conserv.*, **7**, 37–59 (1975).
[29] Bailey, K., *Science*, **192**, 465–467 (1976).
[30] Fosberg, F. R., in *Insects of Hawaii*, **1** (edit. by Zimmerman, E. C.) (University of Hawaii Press, Honolulu, 1948).

Hearing lips and seeing voices

MOST verbal communication occurs in contexts where the listener can see the speaker as well as hear him. However, speech perception is normally regarded as a purely auditory process. The study reported here demonstrates a previously unrecognised influence of vision upon speech perception. It stems from an observation that, on being shown a film of a young woman's talking head, in which repeated utterances of the syllable [ba] had been dubbed on to lip movements for [ga], normal adults reported hearing [da]. With the reverse dubbing process, a majority reported hearing [bagba] or [gaba]. When these subjects listened to the soundtrack from the film, without visual input, or when they watched untreated film, they reported the syllables accurately as repetitions of [ba] or [ga]. Subsequent replications confirm the reliability of these findings; they have important implications for the understanding of speech perception.

To further confirm and generalise the original observation, new materials were prepared. A woman was filmed while she fixated a television camera lens and repeated ba-ba, ga-ga, pa-pa or ka-ka. Each utterance was repeated once per second, with an interval of approximately 0.5 s between repetitions. From this master recording four dubbed video-records were prepared in which the original vocalisations and lip movements were combined as follows: (1) ba-voice/ga-lips; (2) ga-voice/ba-lips; (3) pa-voice/ka-lips; (4) ka-voice/pa-lips. Dubbing was carried out so as to ensure, within the temporal constraints of telerecording equipment, that there was auditory–visual coincidence of the release of the consonant in the first syllable of each utterance. Each recording comprised three replications of its auditory–visual composite. Four different counterbalanced sequences of recordings (1)–(4) were prepared, each with a ten-second gap of blank film between successive segments. The recordings were suitable for relay via a 19-inch television monitor; audio–visual reproduction was of good quality.

Twenty-one pre-school children (3–4 y), 28 primary school children (7–8 yr) and 54 adults (18–40 yr) were tested. The adult sample was predominantly male; there were approximately equal numbers of boys and girls in the younger samples. Subjects were individually tested under two conditions: (1) auditory–visual, where they were instructed to watch the film and repeat what they heard the model saying, and (2) auditory only, where they faced away from the screen and again had to repeat the model's utterances. Every subject responded to all four recordings ((1)–(4) above) under each condition, each time in a different sequence; sequence of presentation was counterbalanced across subjects.

For the purpose of analysis, a correct response was defined as an accurate repetition of the auditory component of each recording. Under the auditory-only condition accuracy was high, with averages of 91, 97 and 99% for pre-school, school age and adult subjects respectively; errors were unsystematic. Under the auditory–visual condition, where subjects heard the original soundtrack, errors were substantial. For pre-school subjects

Table 1 Stimulus conditions and definition of response categories from auditory–visual condition

Stimuli		Response Categories				
Auditory component	Visual component	Auditory	Visual	Fused	Combination	Other
ba-ba	ga-ga	ba-ba	ga-ga	da-da	—	—
ga-ga	ba-ba	ga-ga	ba-ba	da-da	gabga bagba baga gaba	dabda gagla *etc.*
pa-pa	ka-ka	pa-pa	ka-ka	ta-ta	-	tapa pta kafta *etc.*
ka-ka	pa-pa	ka-ka	pa-pa		kapka pakpa paka kapa	kat kafa kakpat *etc.*

average error rate was 59%, for school children 52% and for adults 92%.

Subsequent analysis was confined to detailed consideration of responses to the auditory–visual presentations. Responses were first categorised according to the operational definations illustrated in Table 1.

The meaning of 'auditory' and 'visual' categories is self-evident. A 'fused' response is one where information from the two modalities is transformed into something new with an element not presented in either modality, whereas a 'combination' response represents a composite comprising relatively unmodified elements from each modality. Responses which could not be unambiguously assigned to one of these four categories were allocated to a small, heterogeneous 'other' category. Table 2 presents the percentage of responses in each category.

Table 2 shows that the original observation of the effect of [ba]/[ga] presentations on adult responses is highly replicable; 98% of adult subjects gave fused responses to the ba-voice/ga-lips presentation and 59% gave combination responses to its complement. The effect is also generalisable, at least to other stop consonants; 81% of adults gave a fused response to pa-lips/ka-voice and 44% gave combination responses to its complement. The effects, however, are more pronounced with [ba]/[ga] than with [pa]/[ka] combinations; the latter comment applies to all ages.

The data in Table 2 also illustrate that the auditory perception of adult subjects is more influenced by visual input than is that of subjects in the two younger groups; the latter do not differ consistently from each other. It is interesting to note that where responses are dominated by a single modality, this tends to be the auditory for children and the visual for adults. However, it should also be noted that the frequency of fused responses to ba-voice/ga-lips, and pa-voice/ka-lips presentations is at a substantial level for pre-school and school children alike. These auditory–visual illusions, therefore, are observable across a wide age span, although there clearly are age-related changes in susceptibility to them, particularly between middle childhood and adulthood.

Appropriate analyses confirm that the various effects reported for the auditory–visual condition are statistically significant. Alone, however, the data fail to testify to the powerful nature of the illusions. We ourselves have experienced these effects on many hundreds of trials; they do not habituate over time, despite objective knowledge of the illusion involved. By merely closing the eyes, a previously heard [da] becomes [ba] only to revert to [da] when the eyes are open again.

Contemporary, auditory-based theories of speech perception are inadequate to accommodate these new observations; a role for vision (that is, perceived lip movements) in the perception of speech by normally hearing people is clearly illustrated. Our own observations and those of others[1] indicate that, in the absence of auditory input, lip movements for [ga] are frequently misread as [da], while those for [ka] are sometimes misread as [ta]; [pa] and [ba] are often confused with each other but are never misread as [ga, da, ka or ta]. It is also known that, in auditory terms, vowels carry information for the consonants which immeduately preceed them[2]. If we speculate that the acoustic waveform for [ba] contains features in common with that for [da] but not with [ga], then a tentative explanation for one set of the above illusions is suggested. Thus, in a ba-voice/ga-lips presentation, there is visual information for [ga] and [da] and auditory information with features common to [da] and [ba]. By responding to the common information in both modalities, a subject would arrive at the unifying percept [da]. Similar reasoning would account for the [ta] response under pa-voice/ka-lips presentations.

By the same token, it could be argued that with ga-voice/ba-lips

Table 2 Percentage of responses in each category in the auditory visual condition

Stimuli			Responses				
Auditory	Visual	Subjects	Auditory	Visual	Fused	Combination	Other
ba-ba	ga-ga	3–5 yr (*n*=21)	19	0	81	0	0
		7–8 yr (*n*=28)	36	0	64	0	0
		18–40 yr (*n*=54)	2	0	98	0	0
ga-ga	ba-ba	3–5 yr (*n*=21)	57	10	0	19	14
		7–8 yr (*n*=28)	36	21	11	32	0
		18–40 yr (*n*=54)	11	31	0	54	4
pa-pa	ka-ka	3–5 yr (*n*=21)	24	0	52	0	24
		7–8 yr (*n*=28)	50	0	50	0	0
		18–40 yr (*n*=54)	6	7	81	0	6
ka-ka	pa-pa	3–5 yr (*n*=21)	62	9	0	5	24
		7–8 yr (*n*=28)	68	0	0	32	0
		18–40 yr (*n*=54)	13	37	0	44	6

and ka-voice/pa-lips combinations the modalities are in conflict, having no shared features. In the absence of domination of one modality by the other, the listener has no way of deciding between the two sources of information and therefore oscillates between them, variously hearing [babga], [papka], and so on.

The *post facto* nature of this interpretation is acknowledged. More refined experimentation is required to clarify the nature and ontogenetic development of the illusions and their generality needs further investigation. We are at present working on these issues but believe that the finding now reported are of some interest in their own right.

Full details of the dubbing procedure are available from us. We thank the staff of the University of Surrey AVA Unit for their technical assistance in preparing materials and also Susan Ballantyne whose lip movements we filmed.

HARRY MCGURK
JOHN MACDONALD

Department of Psychology,
University of Surrey, Guildford,
Surrey GU2 5XH, UK

Received July 15; accepted November 11, 1976.

1 Binnie, C. A., Montgomery, A. A., and Jackman, P. L., *J. Speech Hearing Res.*, **17**, 619–630 (1974).
2 Liberman, A. M., Delattre, P. C., and Cooper, F. S., *Am. J. Psych.*, **65**, 497–516 (1952).

Squeezing speech into the deaf ear

DEAFNESS caused by attenuation at the middle ear—conduction deafness—can be ameliorated by hearing aids and surgical procedures, but damage to the cochlea presents more severe problems[1]. We are altering the amplitudes of different parts of the speech wave relative to one another, so that the features normally determining intelligibility are selectively amplified. Tests using speech modified in this way will show whether major aspects of speech processing are unchanged by sensorineural deafness, and, if so, whether the improvement is large enough to justify the use of such processing in hearing aids.

The principal characteristics of sensorineural deafness due to cochlear damage are: (1) loss of sensitivity to high frequencies; (2) tinnitus—whistling or hissing sounds; (3) recruitment—reduction of dynamic range for intensity, the threshold for detection being raised while the maximum acceptable intensity is near normal. Both conduction and sensorineural deafness can be simulated in the normal ear[2,3]; conduction deafness by ear plugs and sensorineural deafness by adding masking white noise (that is, noise containing equal power in all audible frequencies). Linear amplification will compensate for the signal attenuation of conduction deafness. It cannot, however, overcome loss of hearing associated with the limited dynamic range of the damaged cochlea. Standard hearing aids applied in the latter case can amplify the peak sound intensities above acceptable limits, overloading the middle ear to produce distortion, discomfort and sometimes pain without giving adequate intelligibility to normal speech. They may also produce further damage to the hair cells[4,5].

Fortunately, the energy peaks are not important as carriers of speech information[6]. Most of the speech information is carried by the zero-energy crossover points of the wave. This implies that the form of the speech wave is not inherently important, and that liberties can be taken with the speech wave for a hearing aid. Gregory and Wallace[2] concluded that it should be possible to prevent overloading while retaining the necessary information, by using suitable peak clipping in hearing aids. Amplitude clipping was, however, found (by ourselves and others) to produce unpleasantly distorted sounds, and the Fourier components generated by sharp clipping give masking tones which reduced intelligibility. "Rounding" the flat tops of the clipped waves with a high cut filter gives some improvement, but removes information since some of the generated components are in the speech frequency band. Alternatives, such as frequency transposition, were also attempted but were abandoned for various reasons[3]. Amplitude compression (AGC) is standard practice for broadcasting and recording. It is only suitable for hearing aids if the "attack time" is made short enough to limit individual energy peaks (of the order of hundreds of microseconds) and so minimise peak overloading. The "release time" must also be made short enough to preserve information of consonants following vowels without causing sound level fluctuations ("breathing")[7].

Villchur used a 2-channel amplitude compressor with an attack time of "less than 1 ms" and a release time of 2000 dB s^{-1}. He found improvements of syllable recognition of from 22% to 160% in quiet and 10% to 230% in a "cocktail party" situation. "Improvement was defined as

$$100\left(\frac{\text{score with processing}-\text{score without}}{\text{score without processing}}\right)$$

An alternative and possibly complementary approach is instantaneous amplitude clipping, but without the squaring and generation of cross modulation products in the speech band associated with peak clipping. This has for a long time seemed impossible; but it can now be achieved.

The solution we have adopted is to generate the inevitable cross-modulation products outside the speech band. This is done by amplitude modulating a high frequency carrier with the speech, and clipping the carrier[8,9]. The cross-modulation products are now multiples of the chosen carrier frequency. By choosing a suitable carrier frequency (50 kHz–100 kHz) the harmonics and cross-modulation products generated by the clipping are widely separated from the speech band, and so can be removed. The result is instantaneous limitation of the audio peaks without flattening or appreciable intermodulation product distortion.

Considering first normal audio frequency peak clipping, any wave exceeding the clipping level will be flattened. For two

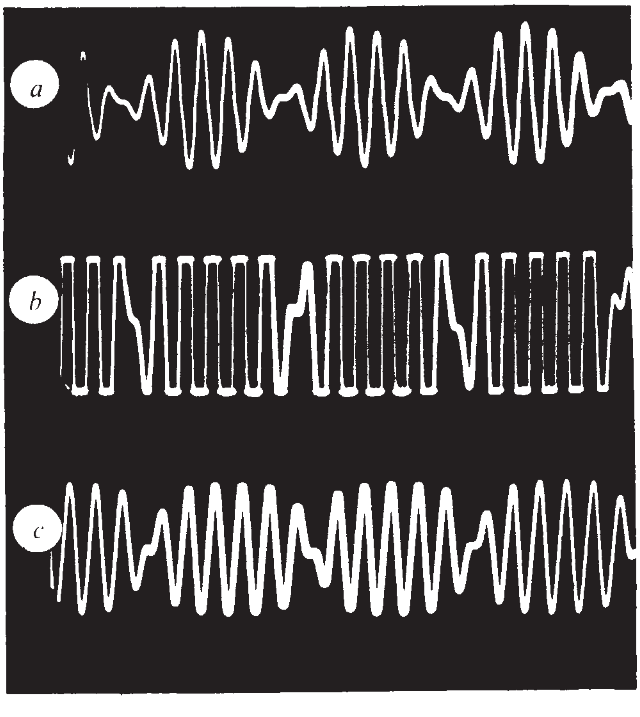

Fig. 1 Unprocessed and processed wave forms, from "beating" a pair of AF sine waves. *a*, The unprocessed wave. *b*, Processed by audio frequency peak clipping. *c*, Processed by the carrier clipping system.