THE TRANSFORMATION OF BIRD SOUNDS INTO 'SPEECH'

J.P. Scanlan

ABSTRACT - Budgerigars mould their imitations of speech on species-specific vocalizations. This process is demonstrated, and some of its mechanisms examined, by analysing intermediate forms in which the transformation is incomplete.

INTRODUCTION

The ability of some birds, when kept as pets, to imitate the sounds of human speech raises intriguing questions concerning both avian perception of speech sounds and human perception of avian "speech". In attempting to address these questions, several authors have drawn attention to certain speech-like features in the species-specific vocalizations of some "talking" birds - e.g. Bertram (1970), in a study of the Indian hill mynah (*Gracula religiosa*), and Nottebohm (1976) in discussing the orange-winged Amazon parrot (*Amazona amazonica*).

This paper approaches the question from a new perspective: it examines a form of avian vocalization which is intermediate between bird sound and "speech", and thus documents the transformation of the former into the latter. In doing so, it demonstrates that even minimal modifications to some bird sounds can result in a distinctly speech-like aural impression.

Brockway (1964) distinguished 10 types of vocal behaviour in the budgerigar (*Melopsittacus undulatus*), and Wyndham (1980) eight. One of these vocal patterns - the one she called "loud warble" - Brockway described at "the 'chatter' so commonly associated with this species". The use of the word "chatter" suggests the speech-like quality of budgerigar warble, and Wyndham's (1980) spectrograms confirm this similarity.

The two *Melopsittacus* subjects of this study had been carefully trained by their owners to repeat a fixed repertoire of phrases and verses. Both these birds, before entering upon their repertoires, would go through a preparatory performance which began as species-specific warble and included an ever-increasing number of speech-like sounds until the repertoire itself was begun. This sequence, which I have called "pre-performance warble", contained speech phrases and syllables which were not part of the performance repertoire. This conforms with the eclectic nature of warble as described by Brockway and Wyndham.

SUBJECTS AND INSTRUMENTS

What follows is an analysis of several sequences from the "pre-performance warble" of two pet budgerigars (Budgerigar A and Budgerigar B - each trained by a different adult female owner) which graphically illustrate the transformation of bird sounds into "speech". The birds were recorded in their homes, together with their respective trainers. Budgerigar A was recorded with a Uher Report tape recorder, and Budgerigar B with a Uher CR 240 cassette recorder. The microphone used for both recordings was a Sony Electret condenser microphone, ECM-30.

The spectrograms were produced on a Kay Sona-Graph 6061B, using the 300 Hz bandwidth filter.

ACOUSTIC ANALYSIS

Whistles and the second formant

I have demonstrated (Scanlan, 1988), by relating acoustic and anatomical data, that a budgerigar is capable of generating in its syrinx pulse trains acoustically similar to those involved in human vowel production, using a mechanism analogous to the periodic adduction of the human vocal folds. I have also shown, however (Scanlan, *ibid.*), that this is not the only mechanism available to the budgerigar for the imitation of human vowel sounds. An alternative mechanism involves the amplitude modulation of a sinusoidal carrier, the carrier frequency being within the second-formant range of human vocalic spectra, and the modulation frequency within the range of human fundamental frequencies.

Sequences of quasi-sinusoidal whistles are common components of budgerigar warble. It is by amplitude modulation of whistles such as these that a bird can create the auditory impression of low-frequency "voicing". At the pre-performance warble stage, however, patterns of amplitude modulation can still be more bird-like than voice-like, and the main adaptation towards "speech" is a matching of whistle frequency contours with the contours of spectral peaks - principally those representing F2 - in the speech signal.

Figure 1, taken from the pre-performance warble of Budgerigar A, shows that only minimal alterations are required to transform species-specific whistles into crude representations of speech sounds. The spectrogram represents the bird's imitation of a phrase - "pretty pretty boy" [prɪtɪ prɪtɪ bɔɪ] - followed by a sequence of three whistles to which it bears a strong resemblance. The plosive burst and upward "second-formant transition" of the first [prɪ] have their counterparts in the rapid release and initial upward glide of the first whistle (at 1.2 sec on the spectrogram). The plosive burst, aspiration noise, and falling second-formant transition of [tɪ] are precisely mirrored - in both frequency range and transition slope - in the rapid release and downward glide in the first whistle. The second [prɪtɪ] is a simplified version of the first. In the third word, the rising second-formant transition of [ɔɪ] has its counterpart in an excursion of similar slope over an identical frequency range (1,200 Hz - 3,300 Hz) at the end of the final whistle. These segments bear a strong resemblance to the second-formant transition in the human-speech version of [bɔɪ] (Fig. 4).

Figure 2 was made from the recording of Budgerigar B, which also produced this phrase in pre-performance warble and also followed it by three whistles resembling it in both temporal and spectral dimensions. This and the previous example suggest that, for both birds, there is a strong association between the species-specific and speech-imitation sounds.

The "chrp" and high front vowels

Perhaps the most convincing evidence for the transformation of a natural sound pattern into an imitation of speech is to be found in those sequences in which speech-like sounds have replaced only part of the total species-specific pattern or, conversely, in which a species-specific element has been reintroduced into an imitated word or phrase. In the pre-performance warble of budgerigars, this is most common in the case of the high front vowels [ɪ] and [i], which share some important acoustic characteristics with the most common call of the budgerigar, the "chrp" (named by Wyndham, 1980). These vowels, like the budgerigar's "chrp", have a strong concentration of energy (high F2 + F3) at about 2,000-3,500 Hz. The "chrp" is also characterised by rapid glides which, in speech imitation, can be adapted to serve as transitions to a high second formant. Indeed, the second-formant transitions of normal human speech are so chirp-like, if isolated, that they have been labelled "chirps" in the literature of phonetics (Mattingley *et al.*, 1971).

Figure 7 shows the insertion of a "chrp" in place of the [ɪ] sound in the phrase "clever boy" [klevə bɔɪ]. Comparison of this phrase, having a mixture of speech-like and avian elements, with a completely "spoken" version (Fig. 6) produced by the same budgerigar (Budgerigar A) demonstrates the similarity between the rapid upward glide at the beginning of the "chrp" and the second-formant transition from [ɔ] to [ɪ] in the word [bɔɪ]. The energy in the [ɪ] sound in the spoken phrase is centred on about 2,800 Hz, as is that in the "chrp". Compare Figure 6 with the human-voice version of [bɔɪ] in Figure 4, and note the similarity of F2 transition pattern.

Further evidence that the "chrp" in Figure 7 represents the final section of the word [bɔɪ] is the presence on the original spectrogram (although not clearly visible on its reproduction here) of a click, like the plosive burst of [b], about 0.13 sec before the onset of the "chrp". Relative timing is also an important clue: the onset of the "chrp" in Figure 7 occurs 0.83 sec after the initial plosive burst in [klevə], and exactly the same interval separates the beginning of the [ɪ] sound in Figure 6 from the beginning of the phrase. The durations of the [ɪ] and "chrp" sounds are similar.

Figure 5 shows the word [bɔɪ] followed immediately by a "chrp" which retains many of the acoustic features of the [ɔɪ] sound. The main area of intensity is within the same frequency range (2,000-3,500 Hz, the F2-F3 region of high front vowels) and the contour of the second-formant transition in [ɔɪ] resembles, albeit in a lower frequency range, the steep upward glide of the "chrp". As in the whistle sequences, this juxtaposition of species-specific and speech-like sounds emphasises not only their physical similarity, but also their close association, for the bird, as components of its vocal repertoire.

Harmonic spectra

Another common component of budgerigar warble is a rather strident vocalization consisting of an array of harmonically related components. This is probably the "brr" sound named by Brockway (1964); I have referred to it as "harmonic warble" (Scanlan, 1988). Figure 3 demonstrates a transition from harmonic warble to speech imitation in which the main factors in the transformation are the spectral and temporal qualities of the initial burst of noise, and the pitch contour of the succeeding vocalic segment. The harmonic warble (1) is followed by an intermediate stage (2) and then the speech imitation proper (3). It is obvious from the spectrogram that, in terms of acoustic patterning, the bird has modified (1) only minimally to produce the very different perceptual effect of (3). Note particularly the fricative noise at the beginning of (3) which is of similar frequency range (mainly between 2,000 Hz and 4,500 Hz), intensity, and duration (about 46 msec) to that in the same word - the bird's name, "Joey" - pronounced by his trainer (Fig. 4). The contour of the harmonic components has been modified in (3) to resemble the falling then rising contour of F2 in Figure 4.

CONCLUSIONS

There are significant acoustic similarities between bird sounds and speech - particularly in the second-formant frequency range. Only minimal modifications are required to transform species-specific sounds into speech-like sounds. Acoustic cues associated with F2 are of great importance in human perception and discrimination of speech sounds. (See, for example, Liberman *et al.*, 1954.) The avian sound-production system allows direct, precise control over "F2" generation (Scanlan, 1988). The avian auditory system in general (Dooling, 1982) and the budgerigar auditory system in particular (Dooling and Searcy, 1985) are selectively responsive to spectral information in the F2 range. "Talking" birds are thus well equipped, at the levels of both sound perception and generation, for the production of "speech" sounds which are recognisable as such to human ears.

The analyses presented here have revealed correspondences between some F2-related speech-perception cues (e.g. second-formant transitions) and basic vocal patterns of the avian imitator. I have suggested (Scanlan, 1988) that these cues may be of communicative significance to both birds and man. This hypothesis is supported by experimental evidence that budgerigars can use spectral cues to discriminate among vowels (Dooling and Brown, 1990), and that they can categorise consonant-vowel syllables according to boundaries along the voice-onset-time continuum similar to those employed in human speech perception (Dooling *et al.*, 1989).

I have argued (Scanlan, 1988) that, once a "talking" bird has formed the necessary emotional bond with a human owner, it actually *perceives* the communication sounds (i.e. speech) of the owner according to special neural processes normally reserved for the perception of species-specific sounds. In other words, for the "imprinted" bird the distinction between species-specific and speech sounds has at least been modified, if not entirely removed. This process, which is facilitated by the factors outlined in the preceding two paragraphs, can be observed in the budgerigar vocalizations analysed above.

REFERENCES

Bertram, B. (1970) "The vocal behaviour of the Indian hill mynah, *Gracula religiosa*", *Anim. Behav. Monogr.* 3, 79-192.

Brockway, B.F. (1964) "Ethological studies of the budgerigar: reproductive behaviour", *Behaviour* 23, 294-324.

Dooling, R.J. (1982) "Auditory perception in birds", in *Acoustic Communication in Birds*, D.E. Kroodsma and E. H. Miller, eds (Academic Press, New York and London).

Dooling, R.J. and Brown, S.D. (1990) "Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels", *Percept. & Psychophys.* 47, 568-574.

Dooling, R.J., Okanoya, K. and Brown, S.D. (1989) "Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction", *Percept. & Psychophys.* 46, 65-71.

Dooling, R.J. and Searcy, M.H. (1985) "Nonsimultaneous auditory masking in the budgerigar (*Melopsittacus undulatus*)", *J. Comp. Psychol.* 99, 226-230.

Liberman, A.M., Delattre, P.C., Cooper, F.S. and Gerstman, L.J. (1954) "The role of consonant-vowel transitions in the perception of the stop and nasal consonants", *Psychol. Monogr.* 68(8), Whole No. 379.

Mattingley, I, Liberman, A., Syrdal, A. and Halwes, T. (1971) "Discrimination in speech and nonspeech modes", *Cog. Psychol.* 2, 131-157.

Nottebohm, F. (1976) "Phonation in the orange-winged Amazon parrot (*Amazona amazonica*)", *J. Comp. Physiol.* 108, 157-170.

Scanlan, J.P. (1988) *Analysis of Avian "Speech": Patterns and Production*, Ph.D. thesis, University of London.

Wyndham, E. (1980) "Diurnal cycle, behaviour and social organization of the budgerigar (*Melopsittacus undulatus*)", *Emu* 80, 25-33.

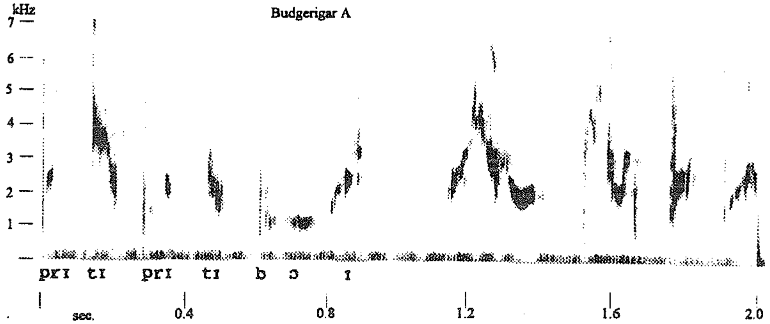

Figure 1. Pre-performance warble

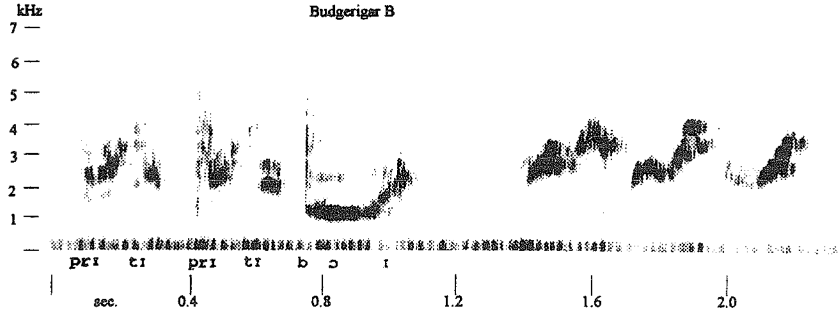

Figure 2. Pre-performance warble

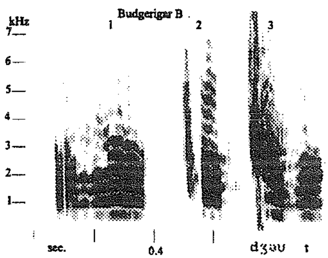

Figure 3. Pre-performance warble

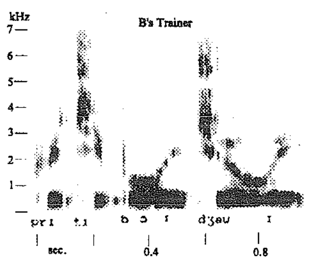

Figure 4. Training exemplar

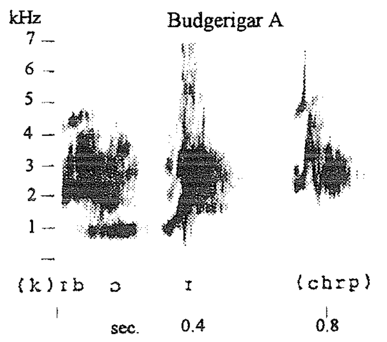

Fig. 5. Pre-performance warble

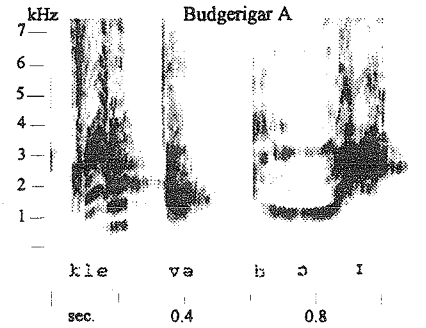

Fig. 6. Performance

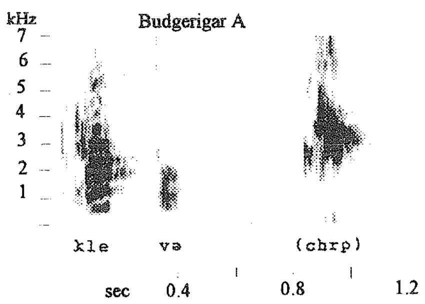

Fig.7. Pre-performance warble