SOME SPEECH AND ACOUSTIC MEASURES OF THE AGING VOICE

L. Penny, A. Russell and C. Pemberton
Department of Speech Pathology
Flinders University of South Australia

ABSTRACT Age is one of the personal attributes which may be judged with reasonable accuracy from samples of a person's speech. The features that the listener uses in making these judgements are of interest not only to the speech scientist, but also to the speech pathologist and gerontologist interested in physiological aging. Longitudinal studies are the method of choice when seeking data on age-related processes, but in the human case such studies are often impracticable for reasons of the time scale, and so cross sectional studies, interpreted cautiously, may be resorted to. Information is offered on some aspects of voice and speech, obtained from a longitudinal study that has so far spanned 45 years, with additional data from subjects at relevant ages, but so far studied at only one time interval.

INTRODUCTION

The source-filter theory of speech production posits a causal relationship between the structure and functioning of the speech mechanisms and the quality and characteristics of the radiated speech signal. The speech signal is a product of both the source energy and the resonating filters through which it passes. Any changes therefore, temporary or permanent, in the biomechanics of these speech organs will be reflected in the spectral and temporal properties of the articulated speech sounds. Age-related changes to the structure and functioning of body systems involved in speech production are well attested. These changes include a reduction in respiratory efficiency (due to atrophy and weakening of the muscles of respiration, and possibly to the calcification of rib cartilage), changes in the larynx due to laryngeal muscle atrophy and changes in the elasticity of laryngeal tissues, and to the pharyngeal and lingual muscles, again due to degenerative changes). Velar efficiency may be reduced. Age-related changes in the tissues of the cheeks and lips, and changes to dentition, may alter the shape and size of the oral cavity and its radiating properties. The production of saliva decreases, further altering the properties of the oral tissues. Additionally, central nervous system changes, involving decreasing neural cells and diminishing release of neural transmitters may result in a slowing of the motor processes of speech production. The result of all these changes are to alter both voice quality and the efficiency and precision of articulatory gestures. Many of these changes will not alter the speaker's ability to communicate a message, but they will contribute to the characteristic sound of the elderly voice, and speech may slow and appear more effortful. Other consequences of aging include changes to hearing acuity, and this may impact on the quality of speech since hearing impairment reduces feedback necessary for the long term control of the highly skilled behaviour which comprises speech. It is possible that kinaesthetic feedback may be impaired too.

The timing of normal age-dependant changes to body systems varies with the particular system in question; some systems show effects much earlier than others. It might be expected that structures and functions which serve vital purposes will 'hold up' much better than others. In this respect, as speech is dependant on structures which serve other, and vital, functions, then speech may resist changes relative to other fields of performance, or may show differing rates of change, depending on the particular vocal aspects involved. For example, changes in laryngeal tissues bringing about changes in the pitch of the voice may precede changes brought about by deficits in the motor control of speech. It is also possible for the speaker to adjust to, and compensate for, certain deficits in performance. The well known trade-off among the elderly of speed for accuracy in skilled performance (Welford, 1977), is part of this process of adjustment, and in the case of speech might be expected to lead to over-all slower rates of utterance before alterations to the actual temporal relationships between speech segments occurs. The only way to plot such changes is through the study of the same people over time, for individuals age at different rates (hence the concept of 'physiological age', analogous to the developmental age of children) although the sequence of changes may be presumed to be constant. Cross-sectional studies of different people at the same age may smooth out important aspects of the progression. It is quite possible that longitudinal plots of a highly skilled behaviour such as speech may contribute to our understanding of physiological aging as well as speech performance across the life span.

Very long-term longitudinal studies of a human behaviour such as speech are by their very nature difficult to maintain. Losses of subjects, and changes in research personnel, equipment, theoretical orientation and funding all make the process rather hazardous. A happy coincidence of events has allowed as to make some efforts in this direction though, and we report on a study which, in effect, began in 1945. In that year, Olive de Pinto arranged with the Australian Broadcasting Commission in Adelaide for recordings to be made of the 27 members of her first year class from the teachers' college. She employed a set of standard sentences and a free speech task. In late 1992 and into 1993 we traced and re-recorded 15 of the 27, using the same tasks as well as some others. We also recorded our own students at the same age as the de Pinto subjects were in 1945, and plan to follow these people through at intervals, for the same time. Additionally, The Australian Longitudinal Study of Aging, which is based at Flinders University in Adelaide, has allowed us to have access to its subjects and to record them in the same tasks. Very detailed histories of these people are available, as well as their current clinical status. The ALSA study commenced collecting data in 1992 and its follow up period is scheduled only for two years, but it has a very large randomly-drawn sample of people starting at age 65 and continuing to over 100 years. We are thus able to augment our longitudinal study with cross sectional data on subjects about whom a great deal is known. This present study reports on some aspects of the longitudinal data, notably the prediction that formant bandwidths and formant relative amplitudes will change as a result of changes to the vocal tract, and that rate of utterance will slow as part of the trade between speed and accuracy that characterises performance in the aging. Data concerning speaking fundamental frequency and its range have been reported elsewhere, (Russell, Pemberton and Penny, 1993).

METHOD

Subjects, materials and equipment.

The subjects were all South Australian-born women speakers of general Australian English. They were all in good general health. both in 1945 and 1992. A 55% recovery rate after an interval of more than 45 years must be considered good. We are not aware of any factors that would lead to there being important differences in speech between the people we traced and the people we could not find, though four people are known to have died. We are unable to determine whether this has resulted in an unrepresentative surviving sample, but with our good numbers, we hope not.

The recordings made by de Pinto have been dubbed on to digital tape by a method described elsewhere (Russell, Pemberton and Penny, 1993). We worked from these tapes rather than from the original steel and acetate discs. Recordings made by us employed a Sony TCD-D10 digital audio-tape recorder and Sony ECM 959 DT microphone. Most recordings were made in a sound-treated room, but some, where the subjects were found living interstate, were made in the best circumstances we could manage. Formant frequencies, formant bandwidths and relative amplitudes were extracted using the Soundscope programme on the Centris 650. Rate of utterance was obtained using a stopwatch.

The standard sentences are those in Ward, 1929.

RESULTS

Formant measures

Formants are plotted in figures 1 and 2. The perhaps striking result is how very similar they are, even though the vowels of later years are made by a voice with a much lower fundamental frequency. The position of /u/ (vowel 10) is perhaps of great interest to students of Australian English.

fig. 1 FORMANT VALUES (connected speech): YOUNGER FEMALE SPEAKERS, 1945

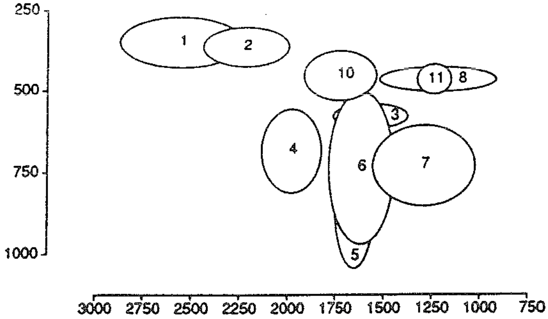
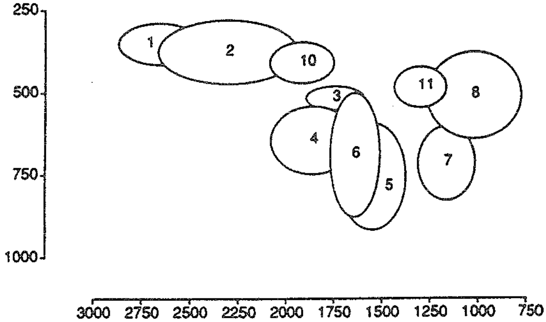


fig. 2 FORMANT VALUES (connected speech): OLDER FEMALE SPEAKERS, 1992



Formant bandwidths

Subject	Bandwidth 1, 1945	Bandwidth 1, 1992	Bandwidth 2, 1945	Bandwidth 2, 1992
1	223	191	370	244
2	185	215	191	455
3	170	239	354	213
4	256	288	528	256
5	171	223	411	178
6	231	174	508	253
7	260	187	488	223
8	203	179	370	260
9	284	215	361	301
10	219	244	415	329
11	183	329	374	374
12	187	207	423	382
13	243	203	439	268
14	329	309	452	402
15	268	321	423	341
	X=227:SD=46	X=234:SD=52	X=407:SD=80	X=298:SD=979

A paired samples t-test on bandwidth 2, yielded t=3.11, 14df, and p=0.0039 (one-tail).

Relative amplitudes

The 15 subjects in 1945 produced 23 instances of 'disordered' relative amplitudes, wherein the relative amplitude of an adjacent higher formant exceeded that of a lower. In 1992 the number was 46. A paired-samples t-test yielded t=3.192, 14df, p=0.0014 (one tail).

Rate of utterance (in syllables per second, and on the standard sentences)

Subject	1945	1992
1	4.68	3.51
2	4.45	3.79
3	4.69	4.32
4	3.65	3.74
5	4.62	3.62
6	4.08	3.20
7	4.43	4.14
8	4.36	4.22
9	4.86	3.95
10	4.37	4.07
11	4.87	3.93
12	4.45	4.32
13	4.90	4.44
14	4.19	3.88
15	4.44	4.27

A paired samples t-test yielded t=4.78, with 14df, p=0.002 (one tail).

. DISCUSSION

The plot of vowel formants shows very little change indeed between 1945 and 1992. It should be remembered that the formants were extracted from connected speech, and that the actual positions of some of them are therefore pulled away from their assumed target positions as deduced from /h_d/ frame data. It had been predicted that bandwidths would increase over the period because of an assumption that the younger, more elastic vocal tract would have higher damping properties. This is born out for B2 but not B1. It will be noted that all the bandwidths reported here are much bigger than the earlier literature suggests (Fant, 1967, and Flanagan, 1972). More recently, Kent and Read (1992) have drawn attention to the wider bandwidths of women. This may be related to the power difference between the voices of men and women. Klatt and Klatt (1990) suggest 'tracheal coupling and source/tract interactions' may be responsible. Whatever the reasons for the wider bandwidths of women, it appears that bandwidth 2 narrows with age. Changes to control over vocal effort may be responsible for the somewhat irregular pattern of formant relative amplitudes which the older voice displays.

This, combined with bandwidth changes, may contribute to the differences in timbre between the older and younger voice that the ear is sensitive to. Rate of utterance changes are as expected. It is of some interest to note that young women speakers in Adelaide today have a rate of utterance for similar material of 4.17 syllables per second. This is a little slower than the equivalent age group was in 1945, and if the longitudinal data werre not available, the slowing of the older group would not seem so marked. The slowing is almost certainly the result of trading speed for precision in a highly skilled act making considerable, if well practised, demands on the performer.

REFERENCES

Fant, C. Gunner M. (1967). On the predictability of formant levels and spectrum envelopes from formant frequencies. In Lehiste, I. (ed) Readings in acoustic phonetics. The M.I.T. Press, Cambridge, Mass.

Flanagan, J.L. (1972). Speech Analysis Synthesis and Perception. 2nd ed. Springer-Verlag, Berlin.

Kent, Ray D. & Read, Charles, (1992) The acoustic analysis of speech. Singular Publishing Group, San Diego.

Klatt, D.H.& Klatt, L.C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 67,979-995.

Russell, A., Pemberton, C & Penny, L, (in press). Journal of Speech and Hearing Research.

Ward, I.C. (1929). Phonetics of English. Cambridge: W Heffer & Sons.

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.