THE SYNTHESIS OF COMPONENTS IN VOICE QUALITY

JOHN LAVER*

It has been said that the distinction between phonetic quality and personal quality, or voice quality in my usage of the term, is one of the basic assumptions of phonetics.[1]

Although we take phonetic quality as the basic datum of our subject, we know less than we might about voice quality, and I have been trying to construct a simple descriptive model of this area.

The essence of the approach is that voice quality is analysable into components, in much the same way that a consonant is described by reference to the voicing, place and manner components in its production.

Briefly, voice quality is conceived as deriving, acoustically, from two main sources: firstly, the permanent and ephemeral organic foundation of the speaker's anatomy and physiology; and secondly, the long-term muscular adjustments, or "settings",[2] once voluntarily acquired perhaps, now unconscious, of the larynx and supralaryngeal vocal tract.

This second section, the long-term muscular settings of the larynx and supralaryngeal vocal tract, is an area open to traditional phonetic techniques of imitation and kinesthetic introspection, and one can thereby form hypotheses about the physiological actions necessary to produce particular qualities heard from a speaker. These hypotheses can be tested by acoustic synthesis, and I have carried out some preliminary simulation of some voice qualities, using PAT, the Edinburgh University resonance analogue speech synthesiser designed by Walter Lawrence.[3]

Different voice qualities were synthesised as backgrounds to a standard linguistic message, a longish sentence of synthetic speech from "The North Wind and the

---

* Department of Phonetics, University of Edinburgh.

[1] Ladefoged, P.: "The Perception of Vowel Quality" Ph. D. Thesis, University of Edinburgh, 1960.

[2] Honikman, B.: "Articulatory Settings" in "In Honour of Daniel Jones", D. Abercrombie et al. (eds.), Longmans, London, 1964.

[3] Anthony, J. and Lawrence, W.: "A Resonance Analogue Speech Synthesiser", *Proceedings of the 4th International Congress on Acoustics*, Copenhagen, 1962.

Sun",[4]—"Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him, and at last, the North Wind gave up the attempt."

Three different fundamental frequency ranges were used, which I shall call Voice, (50—250 c/s), Falsetto, (120—480 c/s), and Creak, (32—128 c/s).

Besides the normal unmodified larynx and vocal tract settings for PAT, some extra optional laryngeal and supralaryngeal modifications were superimposed on the three fundamental ranges. Laryngeal modifications were: harshness and whisperiness. Supralaryngeal modifications included nasalisation; a latitudinal distortion of the vocal tract,—velarisation; and a longitudinal distortion of the vocal tract,—raised larynx.

Thus voices with up to five different components can be simulated: optional harshness, optional whisperyness, and optional nasalisation; one of the three configurations of the vocal tract,—unmodified (i.e. normal), velarised, or raised larynx; and one of the three fundamental frequency ranges, voice, falsetto or creak.

All possible combinations of these components give a total of seventy two different voice qualities, such as: Nasalised Falsetto, Harsh Whispery Creak, Harsh Nasalised Velarised Voice, Whispery Nasalised Raised Larynx Falsetto, and so on. I'd like to play you some of these, reserving acoustic details of the syntheses for later, should anyone be interested. I would especially appreciate any comments on the naturalness or otherwise of the qualities,—judging naturalness by a readiness, for example, to ascribe characteristics of personality to the voice in question.

List of items on Demonstration Tape

1. Normal Voice
2. Normal Falsetto
3. Normal Creak
4. Normal Whispery Voice
5. Normal Harsh Voice
6. Normal Harsh Whispery Voice
7. Normal Nasalised Voice
8. Velarised Voice
9. Raised Larynx Voice
10. Whispery Velarised Creak
11. Harsh Whispery Nasalised Velarised Falsetto

In conclusion, I'd like to say that a componential description seems to be a useful approach to structuring this area. In the synthetic work so far, only categorical

differences between voices have been attempted; the synthesis of scalar, quantitative differences within a component is the eventual aim of the project.

DISCUSSION

Fant:

1. Have you made an acoustic analysis of various pathological voices?
2. Your nasal quality might have benefitted from a simultaneous shift up in F_1.

Mazzarella:

The artificial voices used in the demonstration often seemed to suggest a certain age of the speaker.

Laver:

ad Fant: I thank Dr. Fant for his suggestion about raising F1 to give a better nasal quality.

I have been less concerned with pathological voices than with the voice quality variations which are susceptible of imitation by normally equipped speakers — i.e. voluntary long-term muscular settings of the larynx and vocal tract. I have personally made no acoustic analyses yet, relying more on the somewhat restricted literature, and on physiologically-based hypotheses.

Clearly the next step is to proceed to the acoustic analysis of individual voices.

---

[4] Uldall, E.: "The Synthesis of a Long Piece of Connected Speech", Stockholm Speech Communication Seminar, 1962.