STABILITY OF VOICE FREQUENCY MEASURES IN SPEECH

W. J. Barry (1), M. Goldsmith (2), A. J. Fourcin (1), H. Fuller (2)

(1) University College London, (2) National Physical Laboratory

ABSTRACT

The stability of personal average laryngeal frequency during speech was examined in two-minute and fifteen-minute recordings made at different times of the day and for four differing speech production tasks. One and a half hours of speech from each of four subjects were analysed with respect to mean and modal frequency and frequency range.

The work was funded under Alvey Project MMI/132, Speech Technology Assessment.

1. INTRODUCTION

Voice source characteristics are an important part of the expressive component in the speech communication chain. The most commonly used of these is the frequency of vocal-fold vibration, which is considered both subjectively [1] and objectively ([9], p.82) important for the identification of a speaker. However, although there have been a large number of studies devoted to the measure of average voice fundamental frequency in groups of 50 or more speakers ([9, 2] for survey information), and there is general recognition that a speaker's voice pitch varies with the situation, little has been done to clarify the general question of individual stability in that or other voice frequency measures.

This paper presents results from an in-depth study [4] of the laryngeal patterns in speech of four speakers, and investigates the question of stability in a number of ways. Firstly, on the assumption that two minutes of continuous speech are sufficient to characterise a speaker's voice frequency, [12, 11, 10, 8], the "stability" of two-minute stretches over a 15-minute period was examined. Secondly, in view of evidence that that system-atic longer term fluctuations occur [7], a series of 15-minute recordings were made at different times of one day. Thirdly, a number of different types of speech tasks were set in order to examine the stability of average voice frequency over tasks.

2. SPEECH MATERIAL

Larynx and speech-signal recordings were made with four speakers (2F, 2M) on two days under anechoic conditions using a two-channel PCM-VCR recorder. Two speakers took part in 6 recording sessions of approximately 15 minutes each during the course of each of the two days. A standard extended reading task (P+B) was given at 9.40am, 12.10pm, and 3.30pm (times ±20 minutes), namely the reading of the "Environmental Passage" [3] followed by 13 minutes of continuous reading from a book of short stories. These were interspersed (at 10.20, 11.20 and 12.30 ±20 minutes) with three further tasks for comparison purposes: 1) Eight consecutive readings (RP) of the "Environmental Passage" to provide a constant textual structure and avoid fluctuations in interest and excitement. 2) Free monologue (Mon.) for 15 minutes on a subject of each speaker's choice to give a longer term comparison between read and freely spoken monologue. 3) A dialogue with pairs of speakers (Dial.) to compare both with reading and with free monologue.

3. ANALYSIS PROCEDURES

Analysis of the larynx signal (Lx) was carried out in the standard way [4, 3, 5] to give laryngeal frequency distributions over 128 logarithmically equal bins on a scale from 30 to 1000Hz. "Cleaned" distributions were used for all quantitative statements and statistical calculations. They were derived by including Lx cycles only if their durations were within 10% of the preceding cycle. This eliminates laryngeal irregularities from the distribution while retaining a maximum number of data points. In addition, the distributions are time-weighted to correspond to our auditory impression of pitch movement as frequency change in time, and to conform to other voice-frequency studies, which have obtained their data from fixed *frame*-based rather than from fundamental- or laryngeal-*period*-based analysis.

Processing of the 15-minute sessions was done on sections of approximately 2 minutes, and overall distributions for the whole session was obtained by cumulation of the shorter sections.

4. RESULTS

Two aspects of larynx frequency stability are addressed: Firstly, the extent to which a "minimal stable period" (i.e. 2 min.) still fluctuates within a longer stretch of speech. Secondly, whether there are genuine systematic baseline changes in the longer term.

4.1. Two-minute sections

There are considerable differences between the mean values for the 2-minute sections both within and across the different tasks. Table 1 gives the range in semitones between the lowest and highest mean and modal frequency for a two minute stretch during each of the tasks.

The largest shifts in mean frequency appear to be randomly distributed over tasks and speakers. The comparison of the repeated-passage reading (Task 2) with the other conditions does not support the assumption that shifts in mean frequency are a function of text-type alone. The difference in Task

Table 1. Difference in semi-tones between the highest and lowest mean & (*modal*) frequency for a two-minute stretch during each 15-minute task.

TASK	FC	CP	AH	TJ
1.	1.26	0.45	0.53	1.15
	1.01	*1.04*	*3.41*	*1.68*
2.	1.36	0.19	0.72	0.91
	1.01	*0.00*	*0.86*	*0.70*
3.	0.51	1.05	1.93	0.60
	0.00	*1.04*	*0.86*	*0.80*
4.	0.75	0.81	1.35	2.12
	2.10	*1.10*	*2.60*	*1.60*
5.	1.56	0.95	0.66	0.19
	1.01	*1.04*	*0.85*	*0.00*
6.	0.91	0.82	0.87	0.87
	2.10	*1.04*	*0.88*	*4.23*

2 is very low for speaker CP, but not for the other speakers, and in fact has the highest value of all tasks for FC. In general, the pattern in time for 15 minutes is more regular than for the book readings, but in no uniform way across speakers. CP has one small shift, FC has a steadily rising frequency, and the two male speakers AH and TJ have intermittent fluctuations. Modal values, however, do not reflect the expected task-dependence any more clearly.

The differences between the speakers in the repeated-passage task, and the random distribution of large and small pitch differences for 2-minute stretches across the different tasks, indicate that there is, strictly speaking, no such thing as a generally valid personal voice-frequency value, either mean

or mode (compare [1]). Each situation appears to have its specific and individual effect on a speaker's voice frequency.

4.2 Longer-term stability

To test for variability in the longer term, the values found for the standard passage in tasks 1, 4 and 6 in this study were compared with those for the same speakers reading the same passage (P) some weeks earlier [3]. Table 3 lists the mean and modal values for all the *single* readings.

Table 3. Mean and (*modal*) Fx values (Hz) for the standard passage in Tasks 1, 4 and 6 and in P.

TASK	FC	CP	AH	TJ
1	213	229	114	104
	197	*222*	*108*	*103*
4	223	239	122	109
	209	*222*	*114*	*103*
6	228	234	123	110
	209	*222*	*108*	*114*
P	208	225	130	102
	209	*236*	*126*	*98*

The overall picture of variation from situation to situation is confirmed by this overview. For each speaker, values for the same text vary from one recording to the next as much as the different 2-minute stretches varied within the 15-minute readings from a book. However, table 3 shows a certain regularity in the values from task 1 to task 6. For all four speakers there is an increase in mean Fx from task 1 to task 4. This is followed by a fall-off in task 6 for CP while TJ and AH maintain approximately the same mean excitation frequency and FC shows another increase. Given the timing of the tasks (morning, mid-day, afternoon) this progression is similar to that reported in [7], namely an increase in fundamental frequency for all speakers during the first part of the day (9am-noon). In partial contrast to our results, however, an increase was found there for the male speakers during the afternoon with a fall-off restricted to female speakers.

These shifts in the 2-minute standard passages are related to shifts in mean frequency over the 15-minute stretches of which the standard readings are a part. Three of the four speakers show a marked upward shift from morning to mid-day and fall back from mid-day to afternoon. The one speaker (TJ) who has no marked shift in overall mean pitch from morning to mid-day *does* have a significant shift from mid-day to afternoon (Mann-Whitney, U' = 6, p = 0.05).

Modal values both for the 2-minute passage and the whole 15-minute reading tasks confuse this fairly clear picture. In the passage (table 3) CP deviates from the pattern, in that her modal value remains constant throughout the day. In the 15-minute values, TJ is the exception, with a steady decrease of mode from 108Hz in task 1, 98Hz in task 4, and 94Hz in task 6.

4.3 Task dependence

Table 4 gives the overall mean values for monologue, dialogue, repeated reading, and tasks 1,4,6 together (compare the latter with table 3 values for the separate free reading sessions).

Table 4. Mean/*modal* values (Hz) for tasks 3, 5, 2 and 1+4+6.

TASK	FC	CP	AH	TJ
3	191	237	114	89
	186	*222*	*108*	*89*
5	209	222	104	100
	197	*222*	*98*	*98*
2	228	232	117	101
	197	*222*	*98*	*98*
1,4,6	223	239	121	104
	209	*232*	*106*	*98*

There is a strong tendency for mean voice frequency to be lower for spontaneous than for read speech. All four dialogue values

and three of the four monologue values (exception CP) are lower than for any of the reading conditions. This finding parallels the results from the STA Normative Study [3] and those from the literature discussed in it. The modal values, however, show a different pattern. Neither the monologue nor the dialogue are consistently lower than the reading tasks, and AH rather than CP is the exception.

5. SUMMARY AND CONCLUSIONS

The results from this investigation of voice-frequency stability indicate very clearly that while mean frequency has been shown to stabilise over a speech sample of 2 minutes or more in duration, fluctuations from one sample to another forbid the use of such a measure as an absolute personal characteristic. It can only serve as a characterisation of the sample in question. Across speech tasks, and even between different 2-minute samples of the same extended task, mean frequency was shown to vary by as much as 15%. Within speakers, some regularity was found in mean-frequency change during the course of the day, and between spontaneous and read speech. This supports pre-vious findings in the literature, but individual variation in the patterns found indicate that these trends also need to be treated with caution.

In the light of these results, the conclusion is unavoidable that reliance on mean voice frequency as an indicator of personal identity is inadvisable. The possible alternative measure, modal frequency, appears less sensitive to task variation, but it still does not offer a reliable voice-frequency characterisation of speakers.

6. REFERENCES

[1] E. ABBERTON & A.J. FOURCIN, Intonation and Speaker Identification. *Lang&Sp. 21*, 305-318 (1978)

[2] R. J. BAKEN, *Clinical Measurement of Speech and Voice*. Boston: College Hill Press, (1987)

[3] W.J. BARRY, M. GOLDSMITH, A.J. FOURCIN, & H. FULLER, H. (1990): Larynx Analyses of Normative Reference Data. *Project Report, Alvey Project MMI 132*, London: University College.

[4] W.J. BARRY, M. GOLDSMITH, A.J. FOURCIN, & H. FULLER (1990): Stability of Laryngeal Measures in Speech. *Project Report, Alvey Project MMI 132*, London: University College.

[5] A.J. FOURCIN, Laryngographic assessment of phonatory function. In: C.L. Ludlow (ed.) *Conference on the Assessment of Vocal Pathology, Maryland*: ASHA Reports 11, (1981)

[6] H.C. FULLER, A.J. FOURCIN, M.J. GOLDSMITH & M. KEENE (1990): A Database of Normative Speech Recordings. *Proceedings Institute of Acoustics* 12, part 10, 1-6

[7] K.L. GARRETT & E.C. HEALEY, An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of the day. *J. Acoust. Soc. Amer.* 82 (1), 58-62, (1987)

[8] S. HILLER, J. LAVER & J. MacKENZIE, Durational aspects of long-term measurements of fundamental frequency perturbations in connected speech. *Work in Progress 17, 59-76, Dept. of Linguistics*, Univ. of Edinburgh, (1984)

[9] H.J. KÜNZEL, *Sprechererkennung. Grundzüge forensicher Sprachverarbeitung*. Heidelberg, (1987)

[10] J.D. MARKEL & S.B. DAVIS, Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base. *IEEE Transactions, ASSP-25*, 330-337 (1979)

[11] K. O. MEAD, Identification of speakers from fundamental frequency contours in conversational speech. *JSRU Report 1002*, Ruislip, Middlesex, (1974)

[12] M. STEFFAN-BATOG, W. JASSEM & GRUSZKA-KOSCIELAK, Statistical distribution of short-term F0 values as a personal voice characteristic. In: W. Jassem. (ed.) *Speech Analysis and Synthesis* 2, 195-206, Pol. Acad. Science, (1970)