The function of Declination in Spontaneous Speech

K.M. Hird

Deaprtment of Speech and Hearing Science
Curtin University of Technology

ABSTRACT - The assumptions about the universality of declination were tested in the context of spontaneous speech. Acoustic analyses provided support for a theory of declination that involves both passive physiological processes and processes that are speaker controlled and context specific. The results also suggested that damage to the right cerebral hemisphere impairs the use of processes that are required for the control of declination in spontaneous speech.

INTRODUCTION

The overall aim of the experiments reported in this paper was to test models of declination in the context of spontaneous discourse samples. Two main theoretical orientations dominate the literature. Explanations based on physiological and linguistic theories of phonatory control have been proposed to explain declination and resetting of fundamental frequency. The 'breath-group theory' of declination reported by Lieberman (1967) has motivated extensive research investigating the neurophysiological processes involved in the control of phonation. The literature also includes several linguistic theories of prosodic control. The tone sequence, contour interaction and the peak feature theories described by Ladd (1983) are examples of linguistic perspectives on the control of declination in speech. Each paradigm specifies research methodologies that reflect underlying theoretical assumptions. This paper is concerned however with evaluation of the breath-group theory of declination.

Declination was originally identified on the basis of perceptual judgements (Cohen et al., 1982). Advances in technology then permitted the use of acoustic analysis. Frequency was selected as the acoustic correlate of pitch. Frequency has therefore been selected as the most relevant acoustic parameter in the study of declination. The perception of pitch is not entirely independent of changes in intensity however. Pierrehumbert (1979) demonstrated that listeners relied on changes in intensity to identify phrasal boundaries. In addition to the analysis of the frequency contour in the experiments reported in this paper, evaluation of the corresponding intensity contours was undertaken. Analysis of intensity was included in order to increase the probability of measuring the parameters relevant to the perception of declination in spontaneous discourse.

The aim of the experiments reported here was to test the effect of inspiration on declination and resetting of frequency. According to the breath-group theory (Lieberman, 1967), the gradual declination of frequency across an utterance is the consequence of gradually diminishing subglottal pressure. Subglottal pressure declines naturally as air is expired during phonation. Fundamental frequency is either maintained until the end of the utterance, when there is a rapid drop prior to inspiration, or it declines gradually over the utterance. Then, when speakers inspire, they restore subglottal pressure. An increase in subglottal pressure causes the vocal cords to vibrate more rapidly. Fundamental frequency following inspiration is therefore at its highest level. In addition, it is reported that most speakers replenish air supply at major linguistic boundaries (Vaissiere, 1983). However, if resetting of fundamental frequency is the result of a rapid increase in subglottal pressure it should occur every time breath is replenished. When inhalation also coincides with linguistic boundaries it is difficult to determine the differential effects of inspiration and linguistic boundaries on resetting of fundamental frequency. The purpose of the experiments reported in this paper was to test the basic assumptions of the breath-group theory of declination whilst controlling for the potential confounding effect of major linguistic boundaries.

EXPERIMENT 1.

Subjects.

Eleven monolingual Australian males participated in this experiment. Mean age was 49 years (SD = 7.46). Each subject was an experienced public speaker and had been interviewed on several previous occasions. The subjects were participants in the Australian Broadcast Commission "Sunday" programme or "Lateline" programme. Each subject was interviewed separately by one of two experienced political journalists. Eleven consecutive interviews were selected over a 10 week period.

Procedure.

Eleven interviews were recorded over a five week period separately from the 'Sunday' and 'Lateline' television programmes. Interviews were between 7 and 20 minutes in length. Each audio tape recording of the interviews was transcribed verbatim by the researcher. Moments of inspiration were identified from the audio tape and verified by physiological correlates observed from the video tapes. Reliability of places of inspiration was determined by an independent rater. Few discrepancies occurred but each was resolved by reviewing both the audio and video tapes of the interviews until there was complete agreement between the raters.

Each breath-group was given a number in sequence for each subject. Each breath-group was then coded according to the place of inspiration. Specifically, each breath-group was classified into the BS (Breath coinciding with a sentence boundary), BC (Breath coinciding with a clause boundary) or B (A Breath occuring at neither a sentence or a clause boundary) categories. Sentence and clause boundary judgements were made according to the LARSP guidelines (Crystal, Fletcher & Garman, 1976). Consecutive segments were categorised until there were approximately 10 samples in the smallest data set. Consequently the number of observations in each category was not equal.

Apparatus.

A Panasonic Video Cassette Recorder VHS NV - L25 series was attached to a Sony Trinitron Colour Television KV 2184AS. A Sony TC-D5 Pro II cassette recorder was then attached at the video audio output (Sony Corporation, 1986). A 90 Figi minute chromium dioxide audio cassette and 90 min Sony VHS video cassette were used to record the interview. The Sony TC-D5 Pro II cassette recorder was attached to a Macintosh Quadra. Analogue recordings were digitised using the Audiomedia (1990) software using 16 bit and 28KHZ sampling rates. The beginning and end of each segment was determined by acoustic verification via the Audiomedia software's play back facility. Digitised samples were then saved on a SyQuest 44mb removable cartridge.

Digitised speech samples were then opened using Signalyse 2.04 (Keller, 1992) acoustic analysis software. Fundamental frequency was extracted using Fast Fourier Transform routine at 100 msec sampling windows. Intensity was extracted from the amplitude envelope at 100 msec sampling windows. The frequency contours were saved as numerical text in Hertz onto floppy discs for further analysis. Descriptive data from analysis of the intensity contour was also saved. The minimum, maximum, average, standard deviation, range and slope of each frequency and intensity contour were calculated automatically by selecting the statistics option from the Signalyse analysis menu.

Statistical Analysis.

Numerical data saved from the frequency and intensity contours were analysed using Statview Student (Feldman & Gagnon, 1991), SuperAnova (Gagnon, Roth,Finzer,Hofman,

Haycock,Simpson & Feldman, 1990) and Systat (Wilkinson, Mullins, Michael, Wille, Coward,et al.,1992) statistical software for the Macintosh.

RESULTS - EXPERIMENT 1

Resetting of fundamental frequency was more frequently observed in speech segments where inspiration coincided with the beginning of sentences. Resetting was also observed following inspiration at non- linguistic boundaries. This effect was not consistent for all subjects, however. Analysis of the trough data provided evidence to suggest that fundamental frequency declined across segments where inspiration had occurred at either a sentence or clause boundary. Analysis of the intensity data provided further support for the idea that frequency and intensity are more highly correlated in the speech segments where inspiration coincides with a sentence boundary. Together, the results from all frequency measurement techniques support the proposition that resetting and declination of fundamental frequency may be controlled by independent processes. Specifically, the results demonstrated context specific frequency effects. This result supports Umeda's (1981) conclusion that resetting may only occur in specific contexts.

The results of the analysis of intensity values provided further support for the proposition that resetting and declination may be controlled by independent processes. Intensity values declined across all speech segments regardless of where inspiration had taken place. Frequency was also shown to decline across segments where inspiration coincided with a linguistic boundary, but not in segments where inspiration did not coincide with a linguistic boundary. Thus, declination of frequency and intensity across speech segments are correlated to some extent. Resetting of fundamental frequency was of a greater magnitude however at the beginning of sentences that coincided with inspiration relative to the other speech segment categories. In addition, the tally of negative slopes associated with regression lines calculated to fit the intensity contours was higher in segments where inspiration coincided with the beginning of a sentence.

The results of this experiment provides support for Gelfer et al.'s (1985) proposition that the processes involved in resetting of fundamental frequency may be different from those controlling declination. Further support for this idea is found in both the linguistic and physiological research. Vaissiere (1983) suggested that the height of resetting of fundamental frequency is under speaker control and carries communicative significance. Lieberman (1967) proposes that declination is a passive response to the limitation of the physiological system. These two propositions may not be mutually exclusive. Hauser and Fowler (1992) suggest that declination of fundamental frequency secondary to declining subglottal pressure could be the default state within the phonatory apparatus. Changes in the relationship between these two acoustic parameters would therefore require some increased effort on the part of the speaker.

It is possible that the special relationship between intensity and frequency observed at the beginning of sentences reflects speakers' attempts to change the usual relationship between the acoustic parameters for some purpose. Umeda (1981) also found evidence to support context specific use of resetting in spontaneous speech samples. Umeda (1981) suggested that the function of resetting in discourse was to highlight new ideas within the discourse structure. It is proposed that the motivation for changing the 'default' relationship between fundamental frequency and intensity is speaker controlled and directed by higher order processes underlying communicative intent. The nature of the 'default' state proposed by Hauser and Fowler (1992) does not fully account for these results.

Intensity was shown to decline across all speech segments regardless of category membership. There was no evidence to support declination of frequency across all segments however. Analysis of the frequency data demonstrated high intra-subject and inter-subject variability.

In addition, there were many missing values in the fundamental frequency contours. Acoustic analysis of eleven male subjects suggests that 'poor' voice quality was not an aberration specific to a particular subject. It appears that voice quality in spontaneous speech is not of consistent quality. Periods of breathiness were frequently observed in the acoustic analysis of all speakers. In addition, frequency data were rarely present at the end of the segments prior to inspiration. The breathy voice quality was not obvious, however, in a perceptual analysis of the tape recordings. On the basis of these observations it is proposed that the perception of declination in spontaneous speech may actually depend on declination of intensity. The perception of declining values of frequency may be the consequence of pitch periodicity where fundamental frequency is perceived even when there is no energy present that corresponds to the fundamental. That is, listeners perceive declining pitch values even though evidence of declining fundamental frequency is inconsistent in the acoustic signal. It is also possible that the harmonics of the fundamental provide sufficent acoustic infromation when energy at the fundamental is missing. The declining values of intensity are however, consistently present in the acoustic signal and the listener may add frequency information using auditory closure techniques.

EXPERIMENT 2.

The second experiment was designed to test the proposition that higher order processes were involved in the control of resetting of fundamental frequency. Spontaneous speech samples were collected from subjects with damage to the right cerebral hemisphere. This patient group was selected as the right cerebral hemisphere has been implicated in the control of prosodic processes (Ryalls & Behrens, 1988). Pragmatic language processes are also thought to be specialised to the right cerebral hemisphere (Tompkins & Mateer, 1985). It was hypothesised that if the function of resetting was to communicate pragmatic information (Umeda, 1981), and if the underlying skills required for this communicative intent were specialised to the right cerebral hemisphere, then damage to the right cerebral hemisphere should impair the use of resetting in spontaneous speech.

Subjects

Nine subjects with lesions to the right cerebral hemisphere following a single stroke participated in the study. Mean age was 68.6 years (SD = 11.4 years). Site of lesion was identified by CT scan in seven of the nine cases. A Neurologist determined site of lesion using clinical examination in the remaining cases. Only subjects who were neurologically stable, had experienced a single unilateral lesion, had no other neurological disturbances, and spoke English as their only language were included in the study. Five subjects spoke with Australian accents, three subjects spoke with English accents and another subject spoke with an Anglo-Burmese accent. The literature provides evidence to discount the effect of 'accent' on the more global prosodic attributes typical of a specific language (Vaissiere, 1983; Lieberman, 1967). In addition all subjects included scored 4 or more on the Delayed Word Recall test (Knopman & Ryberg, 1989). A score of 4 or more ensures that subjects are not suffering from mild dementia. Seven of the eight patients were in-patients at a rehabilitation hospital. The remaining two patients were members of a community 'stroke group' and lived at home with support from their spouse.

Procedure

Each person was interviewed in a quiet place free from distractions. The researcher spent time 'getting to know' the patient and explaining the general aims of the research. Each subject was informed that the research was designed to investigate whether strokes in the right side of the brain had any effect on their voice. They were encouraged to participate even if they did not feel they had acquired a speech or voice problem as a consequence of their

stroke. Each subject gave permission for audio-recording. Video recording was not possible in the hospital setting.

Each subject completed the Delayed Word Recall Test (Knopman & Ryberg, 1989) according to the author's instructions. All subjects appeared relaxed and glad to participate in an informal social conversation. Topics ranged from women's role in society to family events and travel experiences. The researcher terminated the conversation at a natural break in the discourse, hence some samples are longer than others. Conversations were typically about 30 minutes in duration. Each audio tape was transcribed verbatim by the researcher. Moments of inspiration were identified from the audio tape. Speech segments were categorised into the breath groups according to the same criteria used in the previous experiment.

RESULTS - EXPERIMENT 2

The results of the third experiment provided evidence to support the hypothesis that damage to the right cerebral hemisphere impairs the use of resetting in spontaneous speech. Resetting of fundamental frequency was not evident in speech segments where inspiration coincided with the beginning of a sentence. Two explanations for this result were discussed. The first explanation is based on the literature supporting the right hemisphere's involvement in pragmatic language processing. Absence of resetting was interpreted as a consequence of impairment in processes involved in planning pragmatic messages. That is, in the case of resetting of fundamental frequency, impairment was thought to affect higher order processes that are involved in making given/new distinctions. The results provided some support for this explantion. A basic impairment in voice quality was postulated as the alternative explanation for lack of resetting observed in brain damaged subjects (Ryalls, 1988). That is, the restricted variability in the frequency contours produced by right hemisphere damaged subjects was thought to be a consequence of poor motor control. There was less support for this second explanation for the results of the third experiment. Intensity contours produced by the right cerebral hemisphere damaged subjects did however decline over speech segments in the same way as normal subjects. The presence of declining intensity and normal voice volume suggests that the right cerebral hemisphere subjects had no severe impairment in motor control processes.

In summary, the presence of declining values of intensity and the absence of evidence for resetting of fundamental frequency supports the hypothesis that right cerebral hemisphere damage interferes with higher order processes involved in the control of resetting in spontaneous speech. Declination of intensity values was not however, affected by right cerebral hemisphere cortical damage. This finding supports the proposition that the perception of declination may be influenced by the declining values of intensity and that this function is secondary to physiological limitations.

CONCLUSION

The experimental results support a theory of resetting declination that involves both voluntary and involuntary processes. Resetting of fundamental frequency is thought to be under voluntary control. It is proposed that speakers use resetting of fundamental frequency to introduce new information (Farnetani et al., 1988; Terken, 1984; Umeda, 1981) or unpredictable relationships between units (Redecker, 1990) in the discourse structure. Specifically, resetting of fundamental frequency may be important in providing information about the relationship between discourse structures that would on a surface analysis appear unrelated. Relationships that appear unrelated but refer to a shared underlying representation are described by Redecker (1990) as rhetorical or sequential discourse units. It is hypothesised therefore that resetting of fundamental frequency is controlled by higher order processes involved in the formulation and maintenance of the speaker's communicative intent throughout the discourse. The results of the experiments in this paper also provide evidence that these processes are specialised to the right cerebral hemisphere. That is, the right cerebral

hemisphere appears to be specialised for processing global prosodic modulations that facilitate the interpretation and production of deep messages within the discourse structure (Brownell, Potter, Bihrle & Gardner, 1986; Emmorey, 1987).

REFERENCES

Audiomedia 2.02. (1990) Digidesign. 1360 Willow Rd. Suite 101. Menlo Park. CA. 94205.

Brownell, H., Potter, H., Bihrle, A., & Gardner, H. (1986). Inference deficits in right brain damaged patients. Brain and Language, 27, 310-321.

Cohen, A., Collier, R., & 't Hart, J. (1982). Declination: construct or intrinsic feature of speech pitch? Phonetica, 39, 254-273.

Crystal, D., Fletcher, P., & Garman, M. (1976). The grammatical analysis of language disability. Great Britain: Billing & Sons Ltd.

Emmorey, K. (1987). The neurological substrates for prosodic aspects of speech. Brain and Language, 30, 305-320.

Farnetani E., Torsello C.T., & Cosi P. (1988). English Compound versus non-compound noun phrases in discourse: An acoustic perceptual study. Language and Speech, 31(2), 157-180.

Gelfer, C., Harris, K., Collier, R., & Baer, T. (1985). Is declination activelycontrolled? In I. R. Titze and R. Scherer (Eds.). (1985). Vocalfold physiology: biomechanical acoustics and phonatory controlDenver: The Denver Centre for Performing Arts.

Hauser, M. D., & Fowler, C. (1992). Fundamental frequency declination is not unique to human speech. Evidence from nonhuman primates. J. Acoust. Soc. Am., 91(1), 363 - 369.

Keller, E. (1992). Signalyse 2.04. Switzerland: Infosignal Inc.

Knopman, D.S., & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. Archives of Neurology, 46, 141 - 145.

Ladd, D. (1983). Peak features and overall slope. In A. Cutler & D. Ladd (Eds.). Prosody: Models and Measurements. New York: Springer-Verlag.

Lieberman, P. (1967). Intonation, Perception and Language. Research Monograph 38. Massachusetts: M.I.T. Press.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. Journal of the Acoustical Society of America, 66(2), 363-369.

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. J. of Pragmatics, 14, 367-381.

Ryalls, J., & Behrens, S. (1988). An overview of changes in fundamental frequency associated with cortical insult. Aphasiology, 2, 107-115.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure. Language and Speech, 27 (3), 269-290.

Tompkins, C., & Mateer, C. (1985). Right hemisphere appreciation of prosodic and linguistic indications of implicit attitude. Brain and Language, 24, 185-203.

Umeda, N. (1981). The influence of segmental factors on fundamental frequency in fluent speech. J. of the Acoustic Society of America, 70, 350-355.

Vaissiere, J. (1983). Language - Independent prosodic features. In A. Cutler & D. Ladd (Eds.). Prosody: Models and Measurements. New York: Springer-Verlag.