SPEECH PRODUCED UNDER ADVERSE CIRCUMSTANCES

Z. S. Bond
Department of Linguistics
Ohio University
Athens, OH USA

Thomas J. Moore
Armstrong Aerospace Medical
Research Laboratory
Wright-Patterson AFB, OH USA

Speakers sometimes are required to function under adverse or demanding speaking circumstances. We have been examining the effects of two physically adverse conditions, acceleration and high noise levels, on the acoustic-phonetic structure of speech.

INTRODUCTION

Over the past few decades, the acoustic-phonetic structure of speech has been investigated in considerable detail. Almost all of this work has described speech produced carefully, with minimal distraction or disturbance of the speaker; in short speech in benign circumstances. Yet speakers sometimes have to function under circumstances which either impose or require changes in speech. First, speech may be produced in different styles, such as very clear, slow, loud, and so forth, as seems appropriate for a specific audience. Speakers may also be influenced by psychological states such as excitement, fatigue, discomfort or distraction. Finally, speakers may be influenced by the physical circumstances under which they are required to function, such as high ambient noise levels or various forms of physical motion.

The acoustic-phonetic consequences of style differences have received some attention (for example, Schulman, 1985; Picheny, et al., 1986); states of psychological arousal have been investigated primarily from the point of view of assessing the condition of a speaker, the question of interest being whether it is possible to detect stress by examining characteristics of speech. The effects of physically adverse circumstances on the acoustic-phonetic characteristics of speech have received the least investigation.

PHYSICALLY ADVERSE CIRCUMSTANCES

Over the past two years, we have been studying the speech of speakers in physically adverse circumstances: while hearing high noise levels, experiencing high sustained acceleration or whole-body vibration. We have found that speech produced in adverse circumstances differs systematically from speech produced in benign environments. In this report, we wish to characterize briefly the physical environments which we have investigated and summarize our findings to date.

Acceleration.

Acceleration vectors are classified according to the direction in which they act on the human body. Headwards acceleration, which tends to displace body tissue footward, is termed positive G or +Gz. High sustained acceleration is defined as exposure to acceleration forces of 6G or greater for periods in excess of 15 seconds (Burton, et al., 1974). High sustained accelerations would be encountered in some aircraft.

Method. We have examined the acoustic-phonetic structure of isolated words as produced by two male speakers in two conditions: 1) while sitting in the gondola of the centrifuge at the Armstrong Aerospace Medical Research Laboratory without acceleration (at 1G) and at an acceleration level of +6Gz. The speech of the subjects was recorded through the M-101 noise-cancelling military microphone which was located within a standard Air Force oxygen mask. Speech analysis was performed using the program SPIRE (Zue and Cyphers, 1985) on the Symbolics 3670 computer. Measurements from SPIRE displays were made of the first three formants of vowels, word durations, vowel durations, intervocalic obstruent durations, and fundamental frequency in stressed and unstressed syllables.

Results. Speech produced under high acceleration sounds quite normal, even natural. Whether a word has been produced under acceleration is not obvious to a listener. However, some acoustic-phonetic characteristics appear to differ between speech produced under acceleration and speech produced in a benign environment. Differences were detectable both in the timing and spectral composition of segments.

The formant structure of vowels shifted under acceleration. Only the third formant did not exhibit systematic changes. The first formant of most vowels was somewhat higher; the second formant was lower for front vowels and higher for back vowels. The F1/F2 vowel space of one speaker is given in Fig. 1. The vowel space shrinks, suggesting lessened mobility of the articulators.

Mean fundamental frequency in stressed syllables increased for both speakers, by 10 Hz for Speaker 1, somewhat less for Speaker 2.

All but one of the words measured increased slightly in mean duration for Speaker 1; for Speaker 2, however, the mean duration of some word increased, of others decreased. Word duration shifts resulted almost entirely from shifts in the duration of vowels, so that under acceleration Speaker 1 produced longer vowels while Speaker 2 used variable vowel durations. The duration of intervocalic obstruents decreased slightly for both speakers under acceleration. Because of variability in response, it is difficult to determine whether changes in segment and word durations are a function of speaker characteristics or of acceleration levels. Further details of this study are available in Bond, Moore, and Anderson, 1986.

High Ambient Noise.

When in the presence of high ambient noise, speakers tend to increase the level of their speech, presumably to maintain what they judge to be an appropriate level of sidetone. This increase in loudness is typically accompanied by an increase in pitch, reflected in fundamental frequency. While these relationships have been noted repeatedly and described in the extensive literature on the Lombard effect, only recently have other changes of speech produced under noise received attention. Pisoni, et al., (1985) have reported that vowels as defined by the first two formants become less distinct and that the distribution of energy within the speech spectrum shows an increase in high frequency components.

Method. We have examined the speech of one male speaker, a 20 year old student at a Midwestern university, in a number speaking conditions in conjunction with noise exposure. The speaker was recorded on four separate days in five speaking conditions and two recording environments.

The recording environments can be characterized as 'operational', the speaker wore a standard Air Force helmet equiped with an oxygen mask and an M-101 microphone, and 'laboratory', the speaker was wearing a boom microphone.

The speaker recorded two repetitions of ten spondee words in three noise exposure conditions: white noise at 85dB, 95dB and 100dB SPL; he also recorded the same materials in quiet and when instructed to be 'loud'. We will limit our report to a description of speech produced under the two highest noise levels, in comparison with speech produced in quiet and speech intended to be 'loud'. In all cases, the noise was presented over earphones.

The durations of words and segments, the fundamental frequency and energy at the mid-point of both syllables, and the formant structure of vowels and diphthongs were measured from SPIRE displays.

Results. The speaker reported some discomfort while speaking in the operational condition, particularly when he was also exposed to noise. In all conditions, however, his speech was intelligible and produced in a relatively casual conversational style. The first syllable of each word, receiving stress, was produced with a higher fundamental frequency than the second. Since the speaker was producing words in isolation, the second syllable was longer than the first, a result of pre-pausal lengthening.

Speech in noise. Average word duration varied by approximately 100 msec. from speaking in quiet to speaking in noise. The majority of the variability was a function of vowel durations. In quiet, the average duration of the first syllable was 156 msec. and of the second syllable, 234 msec. The first syllable was longest when speaking in 100 dB noise, increasing to 178 msec. The second syllable increased to 265 msec. There was considerable variability, however; vowel segments did not invariably lengthen in noise.

The second syllable was produced with less energy than the stressed first syllable in all noise conditions. In quiet, the second syllable was produced 9dB lower than the first. In the two noise conditions, the differences between the two syllables decreased to 2 dB and 4 dB.

As would be expected, the fundamental frequency of both syllables increased when speaking in noise, though there was some variability and the increases were not directly proportional to noise levels. In quiet, the first syllable was produced at an average F0 of 138 Hz, the second syllable at 109 Hz. At 100 dB noise, the two syllables were produced at an average F0 of 147 and 119 Hz. The absolute levels differ, but the F0 difference between the two syllables is roughly proportional.

Speaking in noise had a detectable effect on vowel formants. Noise was associated with a higher F1, and a lower F2 for front vowels. The effects were most marked for high vowels. The formant shifts associated with noise are given in Fig. 2.

Oxygen mask. Wearing an oxygen mask had a detectable effect on speech in and of itself and the mask also tended to modify some of the changes associated with noise.

Word and segment durations tended to be longer when the speaker was wearing the mask. With the mask but in quiet, mean word duration was 775 msec; at 100 dB noise, it was 768 msec., effectively the same. The stressed first syllable was produced at a mean duration of 172 msec. in quiet, 173 msec. at 100 dB noise; the lengthened second syllable varied from 248 msec. in quiet to 259 msec. in noise. On the average, therefore, wearing the mask tended to cause the speaker to lengthen segments but noise exposure had no additional effect.

The same can be said of energy differences between the two syllables. The second syllable was produced 5 dB lower than the first in quiet, 4dB lower at 100 dB noise.

The average fundamental frequency for the stressed syllable in quiet was 129 Hz, almost 10 Hz lower than without the mask. Noise at 100 dB increased average F0 to 150 Hz, a value comparable to the increase without the mask. The unstressed second syllable was produced at a mean 109 Hz in quiet, 121 Hz at 100 dB noise, values comperable to speaking without the mask.

The vowel space associated with the oxygen mask is given in Fig. 3. The oxygen mask appears to have an effect similar to noise in that F2 tends to lower, particularly for front vowels relative to speech produced without the mask. Noise, however, does not seem to have any additional effects on vowel formants over those associated with the oxygen mask.

Loud speech. When asked to be deliberately loud, the speaker produced words with average vowel durations and fundamental frequency and amplitude values comparable to those characteristic of speech in noise, speech which might be characterized as unconsciously loud. Without the oxygen mask, 'loud' vowels in the two syllables were 181 msec. and 273 msec. in duration. The corresponding fundamental frequency values were 158 Hz. and 123 Hz. The second syllable was 6 dB lower than the first.

When wearing the oxygen mask while attempting to be loud, the speaker produced similar values: mean vowel durations were 170 and 240 msec.; mean F0 was 150 Hz and 119 Hz; the difference in engergy between the syllables, however, was only 3 dB.

The vowel space plot for loud speech is given in Fig. 4. The vowels of loud speech were very similar whether the speaker was wearing an oxygen mask or not. The vowels were shifted, however, from the values of quiet speech: F2 for front vowels lowered and F1 raised, particularly for high vowels.

DISCUSSION

Our primary observation is that the acoustic-phonetic structure of speech can be systematically affected by the physical environment under which it is produced. The observed changes can be correlated with the specific circumstances of speech production.

In order to maintain vision and consciousness at higher accelerations, so called anti-G maneuvers are necessary. These involve pulling the head down, tensing the skeletal and abdominal muscles as much as possible, and increasing intrathoracic pressure by forcibly exhaling against a partially or completely closed glottis. These straining maneuvers undoubtedly affect laryngeal tension and vocal tract configuration, and may be responsible for the changes observed in speech under acceleration. Increased laryngeal tension would be responsible for the observed increase in fundamental frequency. Tension in the pharyngeal region would tend to reduce tongue mobility, resulting in a decreased vowel space.

When speaking under high levels of noise, the speaker increased loudness (energy) and pitch (fundamental frequency). These same changes were associated with deliberately loud speech. The inference is that loud speech is the same, whether due to external physical circumstances or to speaker intent.

According to our subject, the oxygen mask restricts the mandible so that there is some resistance to jaw lowering. However, in previous work (Shulman, 1985), an increase in loudness was associated with a larger mouth opening and a raised F1. We would hypothesize that a speaker who is increasing the loudness of his speech and using a larger mouth opening would tend to shift the point of maximum constriction towards the back, raising F1 and lowering F2 for front vowels. When jaw movement is restricted by the mask, tongue mobility would decrease with approximately the same acoustic effects.

REFERENCES

Bond, Z. S., Moore, T. J., and Anerson, T. R., The effects of high sustained acceleration on the acoustic-phonetic structure of speech: a preliminary investigation. Armstron Aerospace Medical Research Laboratory Technical Report, AAMRL-TR-86-011, 1986.

Burton, R. R., Leverett, S. D., and Michaelson, E. D., Man at high sustained +Gz acceleration: a review, Aerospace Medicine 45 (1974) 1115-1136.

Picheny, M.A., Durlach, N.I., and Braida, L. D., Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech, Journal of Speech and Hearing Research 29 (1986) 434-445.

Pisoni, D.B., Bernack, R. H., Nusbaum, H. C., and Yuchtman, M., Some acoustic-phonetic correlates of speech produced in noise, IEEE International Conference on Acoustics, Speech, and Signal Processing, 1985.

Schulman, R., Dynamics and perceptual constraints of loud speech, paper presented at 110th meeting of the Acoustical Society of America, Nashville, Tennessee, 1985.

Zue, V. W., and Cyphers, D. S., The MIT SPIRE system, Proceedings of Speech Tech, '85, New York: Media Dimensions, 277-279.

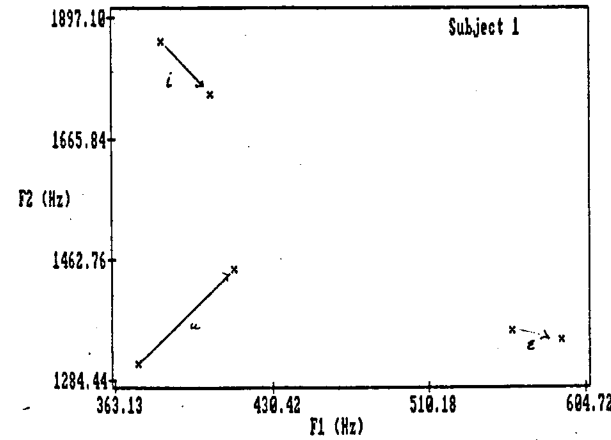

Fig. 1. F1 and F2 at 1G and 6Gz; the arrow points toward the 6Gz values.

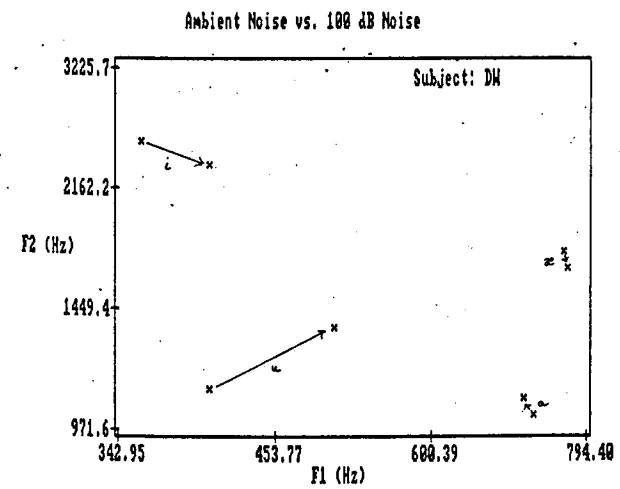

Fig. 2. Formant shifts associated with speaking in noise.

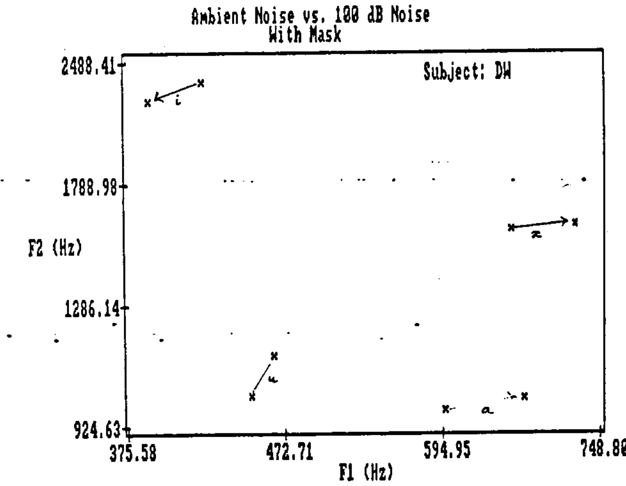

Fig. 3. Formant shifts associated with noise. The speaker is wearing an O_2 mask.

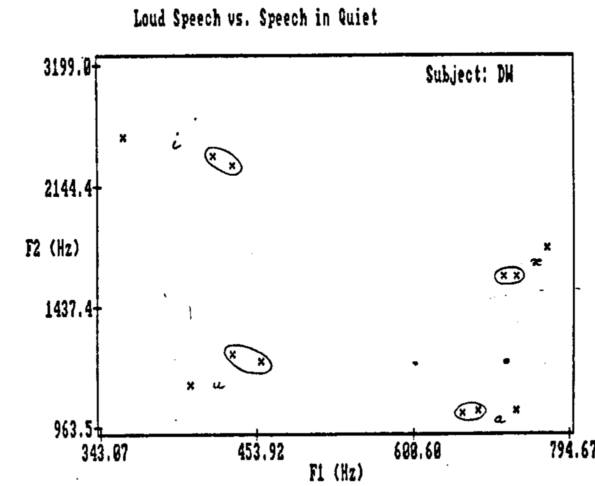

Fig. 4. Speech intended to be loud is enclosed in ellipses.