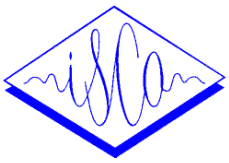
ISCA Archive
http://www.isca-speech.org/archive



2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

# ACOUSTIC AND PRODUCTION PILOT STUDIES OF SPEECH VOWELS PRODUCED IN NOISE

*Jean-claude JUNQUA**

Speech Technology Laboratory, Division of Panasonic Technologies, Inc.
3888 State Street, Santa Barbara, California 93105

*currently working at Central Research Laboratories, Matsushita, Japan.

## ABSTRACT

We conducted two pilot studies: 1) to evaluate the influence of speech loudness on acoustic parameters, and 2) to analyze inter-articulatory relationships in vowel production in noisy and non-noisy conditions. The acoustic study revealed that the tense and lax vowel quadrangles (for the male and female speakers) tend to shift upward with increased vocal loudness (especially for the lax vowels). The production study brought to light that: 1) the type of noise influences speech production, and 2) the modification of speech production due to background noise is speaker-dependent and context-dependent.

10.21437/ICSLP.1992-263

## 1 INTRODUCTION

In the presence of noise, speech is masked and its production is modified by what is called the Lombard effect. Several studies have found that acoustic differences between speech produced in quiet and speech produced in noise modify the intelligibility of speech [1, 2]. In this paper, we present two pilot studies which extend the work reported in [2]: 1) an acoustic study in English vowels produced in noise at different loudness levels, and 2) a study of articulatory movements in English vowels produced in several noisy or non-noisy conditions. The second study extends the work of Lindau and Ladefoged [3] to speech produced in noise. Both acoustic and production studies aim for a better understanding of speech produced in noise.

## 2 THE ACOUSTIC STUDY

### 2.1 Method

This acoustic study focuses on the *changes in acoustic parameters as a response to different levels of increased vocal loudness*. To that end, nine vowels (see table 1) in the h-d context were produced (in response to a stimulus tape) by a child (eight years old), a man, and a woman at five loudness levels (quiet to loud speech). A VU meter was used during the recording of the stimuli to ensure that successive repetitions of each word increased in intensity. However, no attempt was made to ensure equal increments in either loudness or intensity from one word to another. In only three cases, consecutive productions of the stimuli do not increase in loudness. The stimulus tape was played to the subjects through sound field. Two sets of stimuli were recorded onto the stimulus tape. In the first set, each word was presented at the rate of one per second, with a pause between the last repetition of one stimulus word and the first repetition of the next word. This set was played as an example to the subject. In the second set, a five second pause occurred between consecutive repetitions of the same stimulus word. The subjects were instructed to repeat the words during the five second pause, increasing the loudness of their voices (as in the stimulus tape). After all the words had been repeated once, the stimulus tape was rewound and the process was repeated. The following parameters were extracted: the loudness (dBA), the fundamental frequency, the first four formants, and the duration of the vowel. Speech data was digitized at 10 KHz. The formant and the pitch values were extracted by computer and compared to those obtained with a Kay digital spectrograph. The difference between the formant values obtained with the two methods was within the bounds of normal variability, as were the pitch values. Measurements of the loudness were obtained with a Bruel and Kjaer (B&K) digital sound level meter in the dBA scale.

| IY | EH | AE | AO | UW | AH | AA | IH | UH |
|---|---|---|---|---|---|---|---|---|
| heed | head | had | hawed | who'd | hudd | hod | hid | hood |

Table 1 The nine vowels of our stimulus tape. The vowels are written using the Arpabet notation.

### 2.2 Results

The ability of the subjects to match the loudness levels of the taped stimuli was studied. Histograms were prepared in which the frequency of the occurrence of the difference in decibel value (in the A-scale) of each stimulus word and its corresponding production by each subject was plotted. The majority of the male and female subjects' responses fell within five decibels of the stimuli. The child's responses were generally between five and fifteen decibels louder than the stimuli. These results suggest that there is a developmental trend associated with the ability to control vocal loudness.

For the fundamental frequency, as it has already been reported [1, 2], we observed a direct relationship between increased loudness

and increased fundamental frequency. The relationship holds for the three subjects in this study, although it is strongest for the male and weakest for the child. The vowel duration followed a similar tendency as the fundamental frequency. However the duration of tense vowels was generally greater than that of lax vowels at the same loudness level. As expected, the variability of the child's responses was greater than that for either of the adults.

For the formants, a signal processing package [4] was used to display a wideband spectrogram and to compute an LPC analysis. Data were extracted from three frames selected at random from within the stable portion of the formants observed in the wide band FFT spectrogram. The average of the three values was considered to be the frequency of the formant. Our analysis did not reveal any significant relation in the differences between formant values as a function of speech loudness. Nor does there appear a correlation between changes in any formant frequencies, except for F1 and increased vocal loudness in the male subject. In this case F1 values tend to increase with vocal loudness. A plot of the first formant versus the difference between the first and second formant (F1 versus F2–F1) revealed that the male and female speakers use approximately the same amount of vowel space (with a slight tendency to a decrease when the loudness increases), although the female's vowel quadrangles reveal the use of higher formant values. It seems that females, when speaking louder, tend to produce more open articulated vowels. For both the male and female subject most of the changes in the tense vowel quadrangle appeared to be attributable to the decrease in the difference between the formant values for /AE/ and /AA/. This is revealed graphically as a pivoting from the high front vowel /IY/ as loudness increases (see figure 1). Additionally, the vowel quadrangles for the male and female speakers tend to shift upward with increased vocal loudness, especially for the lax vowels. Data from the child is generally more variable than that of the adults. There is considerable change in the child's formant values but these changes do not appear to be consistent with increases in loudness. Additionally, the child uses much more acoustic space than either adult (see figure 2). It is interesting to note that the child's tense vowel triangle intersected that of the male subject but was a superset of the female's. More data must be collected and analyzed to substantiate these findings.

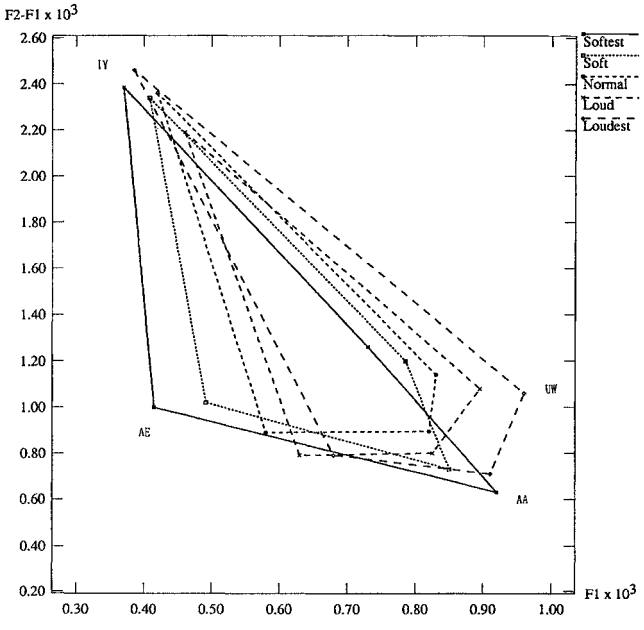



Figure 1 Female tense vowel quadrangle as a function of the speech loudness level.

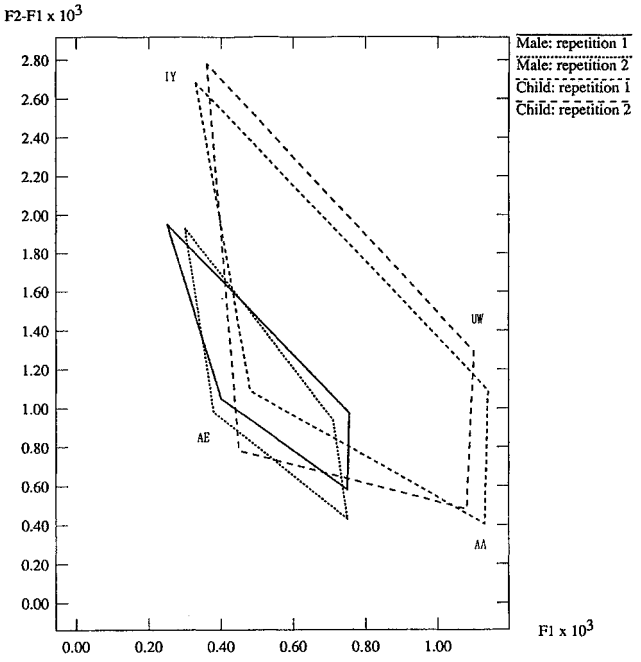



Figure 2 Tense vowel quadrangle for two repetitions of the male and the child speakers in the case of normal speech production.

## 3 THE PRODUCTION STUDY

### 3.1 Method

To study inter-articulatory relationships in vowel production in noisy and non-noisy conditions, we recorded, using the University of Wisconsin X-ray microbeam system, 10 different vowels (the vowels in Table 1 plus /eI/ as in the word "day") in the d-d and s-s contexts for four conditions: 1) normally articulated in a quiet environment, 2) clearly articulated in a quiet environment, 3) when the speaker was exposed to 85 dB SPL white-Gaussian noise, and 4) when the speaker was exposed to 85 dB SPL multitalker babble noise. Two male and two female speakers produced these vowels in a sentence frame (e.g. "Say *d*id to me"). Each speaker produced several repetitions of the vowels in a 5 sec window. We studied the movements of seven pellets: three on the tongue, one on the lower teeth as an indicator of the jaw position, one on the mandibule molar, and two, respectively, on the upper and lower lips. We extracted the mid vowel position

(point of maximum excursion of one of the tongue points) and its amplitude. We are particularly interested to understand: *1) what are the changes in speech production between normal speech and speech produced in noise? 2) Is there any difference between speech produced in noise and articulated speech? 3) Is the influence of noise on speech production dependent on the type of noise?* To this end, we conducted three different analyses where we examined: 1) the pellet positions between various repetitions of the same vowel produced in the same condition; 2) the correlation between the changes in the movement of the articulators; 3) the correlation between the pellet positions across the recording conditions. To this end, multiple regression analyses and *t* tests were applied.

### 3.2 Results

To assess the variation between several repetitions of the same vowel, for each speaker we computed the $r^2$ correlation values using a bivariate paired point analysis. The results, presented in table 2, represent the global variation of the different pellets in the two consonantal contexts.

| conditions ↓ | S1 | S2 | S3 | S4 |
|---|---|---|---|---|
| normal | 0.98 | 0.98 | 0.99 | 0.98 |
| clear | 0.97 | 0.96 | 0.97 | 0.99 |
| white noise | 0.98 | 0.97 | 0.91 | 0.98 |
| babble noise | 0.97 | 0.98 | 0.99 | 0.99 |

Table 2 Correlations of the pellet positions in two tokens, one in the first repetition and one in the second repetition, for each speaker (S1: male, S2: female, S3:male, S4: female) and the four conditions. For this analysis, the two consonantal contexts have been considered together.

These results show that there is little variation between repetitions of the same word for the different conditions studied. This consistency makes the differences that we will report in the following paragraphs of this paper more valid.

The second analysis examines the correlation between the movement of the articulators for, successively, the four conditions and the two consonantal contexts. Specifically, the jaw height was compared to the height of all three points of the tongue. A similar study has already been reported by Lindau and Ladefoged in [3] in the case of normal speech. Our study extends their work to articulated speech and speech produced in noise. From multiple regression analysis, we extracted correlation values ($r^2$). Table 3 and table 4 show the results obtained.

Some speakers exhibit a stronger correlation than others (e.g. speaker 2 versus speaker 3). This result is in agreement with Lindau and Ladefoged [3] who suggested that some speakers behave in accordance with a jaw-based model of the tongue, while some others moved the tongue and the jaw independently of each other. For a given speaker, this result is valid across the different recording conditions. It is interesting to note that the recording conditions influence the correlation values. However, this phenomenon seems to be speaker- and condition-dependent. The consonantal context seems also very important. Table 3 and table 4 show that better correlation values are obtained in the case of the s-s context.

Finally, we examined, for each context, the correlation between the pellet positions across the recording conditions. The results obtained are plotted in figure 3. It can be seen that, generally, clearly articulated speech has a higher correlation with speech produced in presence of babble noise than with the two other kinds of speech (normal speech and speech produced in presence of white noise). Depending on the type of speech the speakers seem to adopt different strategies to move their articulators. When the correlation is computed between normal speech and two other kinds of speech, it is interesting to notice that normal speech seems to be generally closer to clear speech than to the two other kinds of speech in the d-d context. However, normal speech seems to be generally closer to speech produced in presence of babble noise than to the two other kinds of speech in the s-s context. As it is shown, the consonantal context plays an important role in the results obtained.

| d-d context | | | | |
|---|---|---|---|---|
| conditions ↓ | S1 | S2 | S3 | S4 |
| normal | 0.88 | 0.97 | 0.59 | 0.61 |
| clear | 0.81 | 0.97 | 0.63 | 0.90 |
| white noise | 0.78 | 0.87 | 0.71 | 0.76 |
| babble noise | 0.91 | 0.72 | 0.59 | 0.74 |

Table 3 $r^2$ values from multiple regression of the jaw height versus the height of the tongue for consonantal context d-d, the four speakers, and the four conditions.

| s-s context | | | | |
|---|---|---|---|---|
| conditions ↓ | S1 | S2 | S3 | S4 |
| normal | 0.88 | 0.91 | 0.87 | 0.89 |
| clear | 0.91 | 0.96 | 0.81 | 0.93 |
| white noise | 0.82 | 0.89 | 0.63 | 0.94 |
| babble noise | 0.84 | 0.70 | 0.70 | 0.95 |

Table 4 $r^2$ values from multiple regression of the jaw height versus the height of the tongue for consonantal context s-s, the four speakers, and the four conditions.

## 4 FINAL REMARKS

In the acoustic study, we focused on the influence of the speech loudness on acoustic parameters. It was interesting to notice that the amount of vowel space used by the speakers stayed almost constant when the loudness increased. There was a slight tendency to a decrease but it was less obvious than the decrease noted by Bond and Moore [5]. However, both studies involve only a few speakers and more data are necessary to confirm these results.

The production study showed that the conditions in which speech is produced (e.g. type of background noise) has a lot of influence on speech production itself. Each speaker seems to adopt its own compensation method to produce an intelligible output. As the compensation method seems also context-dependent, it is difficult to extract some reliable rules which could describe speech production variation across the various conditions.

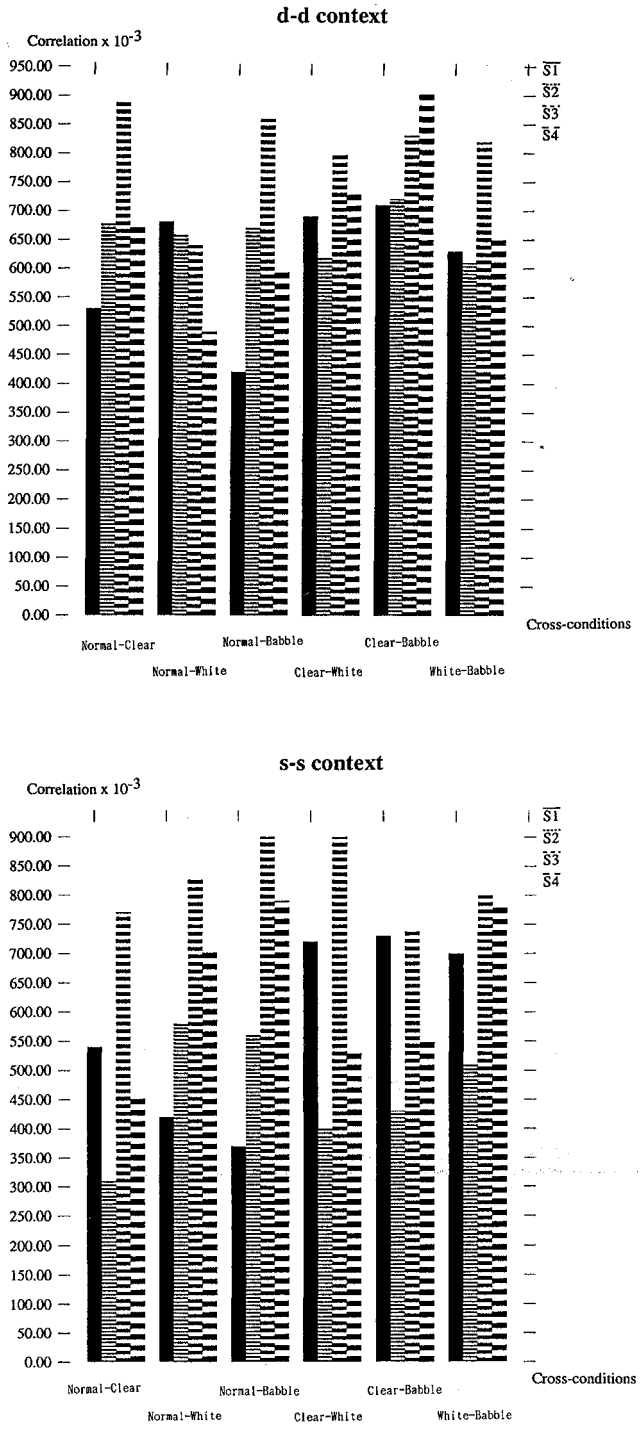



Figure 3 $r^2$ values representing the average correlation between the pellet positions across the recording conditions for the s-s and d-d contexts. For the computation of the average, the pellet located on the upper lip was not included because of the low correlation values obtained for this pellet as compared with the other pellets.

## 5 CONCLUSIONS

The main conclusions of our work are the following: 1) the acoustic study revealed that the child uses much more acoustic space than either adult, and that the tense and lax vowel quadrangles (for the male and female speakers) tend to shift upward with increased vocal loudness (especially for the lax vowels); 2) for each different recording condition, the production study suggested that there is a strong correlation in the movement of the articulators between repetitions of the same word; 3) the type of noise influences speech production; 4) the modification of speech production due to background noise is speaker-dependent and context-dependent; 5) in our study, clearly articulated speech seems to be generally closer (according to the correlation values) to speech produced in presence of multitalker noise than to the two other kinds of speech. However, this speech style has also its own characteristics.

Our future studies aim at the correlation between our acoustic and production results.

### Acknowledgments

The help provided by Lillian Stuman during the parameter extraction is sincerely appreciated. The technical staff of the X-ray Microbeam facilities at Wisconsin University was also of great help during the data collection of the second part of this study. Special thanks to Hisashi Wakita for his useful comments during the first stage of this research.

## Bibliography

[1] W. Summers, D. Pisoni, R. Bernacki, R. Pedlow, and M. Stokes. Effects of Noise on Speech Production: Acoustic and Perceptual Analyses. *J. Acoust. Soc. Am.*, 84(3):917–928, 1988.

[2] J. Junqua and Y. Anglade. Acoustic and Perceptual Studies of Lombard Speech: Application to Isolated-Word Automatic Speech Recognition. In *ICASSP-90*, pages 841–844, 1990.

[3] M. Lindau and P. Ladefoged. Interarticulatory relationships in vowel production. Technical report, UCLA Working Papers in Phonetics 74, 1990.

[4] Y. Laprie. Notice d'Utilisation de Snorri. Technical report, CRIN, 1988.

[5] Z. Bond and T. Moore. A note on loud and lombard speech. In *ICSLP-90*, pages 969–972, 1990.