EFFECT OF F0 FLUCTUATION AND AMPLITUDE MODULATION OF NATURAL VOWELS ON VOWEL IDENTIFICATION IN NOISY ENVIRONMENTS

Kentaro Ishizuka and Kiyoshi Aikawa

NTT Communication Science Laboratories, NTT Corporation
3-1, Morinosato-Wakamiya, Atsugi, Kanagawa, 243-0198 Japan
ishizuka@atom.brl.ntt.co.jp, aik@idea.brl.ntt.co.jp

ABSTRACT

This paper describes findings showing that the fundamental frequency (F0) fluctuation and amplitude modulation (AM) included in natural vowels contribute to improving the vowel identification rate in the presence of interferer sounds. A vowel identification experiment revealed that even very small F0 fluctuations and AM of vowels significantly improved the vowel identification rate in the presence of interferer sound. In addition, it was found that the effect is not the result of reducing the masked threshold. A vowel detection experiment revealed that the effect of reducing the threshold depends on the kinds of interferers, and that the vowel identification rates do not correlate with the masked thresholds.

1. INTRODUCTION

The human auditory system can perceive speech in noisy environments. This noise-robustness of the human auditory system has been studied in relation to the ability to segregate target sound from mixtures of sounds using both spectral and temporal acoustic cues.

McAdams [1] investigated the effect of the frequency modulation (FM) coherence among three vowels to increase the prominence of the target vowel in the presence of the two other vowels. The result showed that the FM of the target vowel increased its prominence. This study suggests that the FM of the target vowel increases vowel perception in the presence of other vowels.

In contrast, Hall et al. [2] found that the presence of noise components at frequencies different from the signal frequency improved the signal detection threshold in the noise when the amplitudes of the noise components were modulated coherently by low-frequency modulator. This resulting masking release has been called ‘co-modulation masking release’ (CMR). This result suggests that the effect of reducing thresholds is due to auditory grouping caused by the coherent amplitude modulation (AM) of the frequency components of the noise. As evidence that CMR is related to speech perception, Carrell and Opie [3] showed that the intelligibility of time-varying sinusoidal sentences [4] was improved by amplitude co-modulation of the sinusoidal component frequencies. This study suggests that the AM of speech contributes to improving speech perception.

In fact, natural vowels actually have coherent FM and incoherent AM. The F0 modulation of natural vowels includes not only a gradual, wide-band modulation, such as the intonation, but also a rapid, narrow-band, and 1/f characteristic F0 fluctuation, such as the ‘jitter’ observed in the sustained part of vowels. The AM of natural vowel frequency components are not the same across the frequencies; namely, they are not co-modulated.

However, there have been few findings related to the effect of the F0 fluctuation and the AM of natural vowels in noisy environments. In the case of McAdams’s study [1], because he used both an artificial vibrato and an artificial jitter, it was not clear whether the prominence improvement was due solely to the jitter. In addition, he focused on target vowel identification in the presence of two other vowels, not background noise. Therefore, it is not obvious whether or not the F0 modulation of vowels increases the prominence of the target vowel in other noisy environments. Moreover, although the effect of coherent AM has been investigated in the field of auditory grouping research, the AM of natural vowels that is not coherent across the frequencies has not been considered.

The purpose of this study is to reveal the effect of the F0 fluctuation and AM of natural vowels in the presence of interferer sounds through vowel identification experiments.

2. METHOD

To analyze the effect of the F0 fluctuation and AM of natural vowels on vowel perception in the presence of interferer sounds, a vowel identification experiment was undertaken. In this experiment, vowel identification rates were measured as a function of male vowels synthesized conditions in the presence of interferer sounds. Four vowel synthesized conditions were used: with no modulation, with only the AM of natural vowels, with only the F0 fluctuation of natural vowels, and with both F0 fluctuation and AM. Hereafter, the term “vowel condition” is used to refer to these synthesized conditions in this paper. The vowels were the five Japanese vowels /a/, /i/, /u/, /e/, and /o/. The interferer sounds were a band noise and a harmonic complex tone with a flat spectral pattern. These two interferer sounds were employed because they are considered to be typical and contrastive models of environmental sounds, and because they allowed us to eliminate the effect of the spectral envelope shape of the interferer. Both sounds have contrastive temporal/spectral properties and similar spectral properties. It is well known that vowel identification is related to the peak position of the spectral envelope of the vowel in the spectral region. Therefore, the flat spectral envelopes of these sounds allow us to eliminate the effect of the spectral envelope shape of the interferer sounds on vowel identification in the spectral region. The effect of the F0 fluctuation and AM of natural
vowels on vowel identification is examined by comparing the vowel identification rates for vowel conditions. Our previous study [5] used the 1/f characteristic artificial F0 fluctuation, whereas this study used F0 fluctuation extracted from natural vowels. This study also used AM extracted from natural vowels for synthesizing the vowels.

3. EXPERIMENTS

3.1. Stimuli

As the first step of stimulus generation, natural male vowels were recorded and digitized before synthesizing the vowels. The speaker was a 26-year-old Japanese male. He was asked to speak the vowels with a stationary pitch when recording. The sampling rate of the natural vowels was 48 kHz. Then, linear predictive coding (LPC) filter coefficients for each vowel were generated from each recorded natural vowel. After that, glottal pulse period sequences were extracted from the digitized vowel samples using a peak picking method [6]. The extracted F0s ranged from 124.4 to 131.1 Hz, and the mean F0 was 127.7 Hz. The bandwidth of the F0 fluctuation was 6.7 Hz (5.2% of the mean F0). Lastly, the amplitude envelopes for each frequency band were extracted as follows. The natural vowels were filtered through a filter bank consisting of 6th-order Butterworth bandpass filters each with a center frequency that was an integer multiple of the mean F0 of the vowels, i.e. 127.7 Hz, and whose bandwidth was 127.7 Hz, and the waveforms for each frequency band were extracted. The extracted waveforms were half-wave rectified and passed through a 2nd-order Butterworth lowpass filter with a cutoff frequency of 50 Hz. In this way the amplitude envelope for each frequency band of the natural vowels was extracted, and the four types of male vowels were synthesized as detailed below. All the vowels were correctly identified in a silent environment by an untrained female subject in a pre-experiment.

Vowel condition A: synthesized vowels with a constant F0 and constant amplitude for each frequency band, i.e. with no modulation. The vowels were synthesized by LPC filtering the glottal pulse sequences with an identical periodicity. The synthesized vowels were filtered using the filter bank, and then the vowels for use were generated by adding together all the filter outputs. The F0 was constant at 127.7 Hz.

Vowel condition B: synthesized male vowels with a constant F0 and with the AM for each frequency band of natural vowels. The vowels were synthesized by LPC filtering the glottal pulse sequences with an identical periodicity. The synthesized vowels were filtered using the filter bank, and then the vowels for use were generated by adding together all the filter outputs. The F0 was constant at 127.7 Hz.

Vowel condition C: synthesized male vowels with constant amplitude for each frequency band and with the F0 fluctuation of natural vowels. The vowels were synthesized by LPC filtering the glottal pulse period sequences of the natural vowels. The synthesized vowels were filtered using the filter bank, and then the vowels for use were generated by adding together all the filter outputs. The F0s fluctuated from 124.4 to 131.1 Hz.

Vowel condition D: synthesized male vowels with both the AM for each frequency band and the F0 fluctuation of natural vowels. The vowels were synthesized as follows. First, the fluctuated vowels were synthesized by LPC filtering the glottal period sequences of the natural vowels. Then, they were filtered using the filter bank, and the amplitude envelope for each frequency band of the natural vowels was multiplied by each filter output of the fluctuated vowels. Finally, the vowels for use were generated by adding together all the filter outputs with AMs. Therefore, the vowels included the F0 fluctuation and AM for each frequency band of the natural vowels. The F0s fluctuated from 124.4 to 131.1 Hz.

The interferer sounds were a band noise with a cutoff frequency of 5361.7 Hz, and a harmonic complex tone composed of a pure tone with a frequency at the mean F0 of the synthesized vowels, i.e. 127.7 Hz, and its 42 harmonics each with the same amplitude in the Schroeder-negative phase [7].

Stimuli were produced by adding one of the interferer sounds to each synthesized vowel at various SNRs. The powers of the vowels were constant for all the vowels in all the vowel conditions. When the harmonic complex tone was used as the interferer sound, the SNRs were –6, –3, and 0 dB. When the band noise was used as the interferer sound, the SNRs were adjusted so that the peak of the band noise power spectrum was the same as the harmonic complex tone. As a result, the SNRs were –12.5, –9.5, and –6.5 dB. The total number of stimuli was 120. The stimuli included a 10 ms taper and the duration was 300 ms. The sampling frequency of the stimuli was 48 kHz.

3.2. Procedure and subjects

The stimuli were presented diotically through headphones (STAX SR-Λ Signature) in a soundproof room at intervals of six seconds. The vowels included in the stimuli were presented at 60 dB SPL. The subjects were forced to identify the vowel they heard for each stimulus and indicate each vowel in turn on a mark sheet. The same stimulus was presented three times to each subject, making a total of 360 tests. For each subject, the experiment consisted of four experimental sessions according to the vowel conditions lasting a total of about two hours. One experimental session consisted of 90 trial tests. The order of the stimuli in one experimental session and the order of the sessions were randomized. A computer was employed to produce the stimuli and to control the experiment.

Forty subjects participated in the experiment, 20 male and 20 female Japanese between 18 and 23 years old. All subjects were paid. None had a history of hearing difficulties.

3.3. Results

The mean vowel identification rate for each vowel condition, interferer sound, and SNR is presented in Fig. 1. There was a 20% chance of guessing the answer correctly. An analysis of variance (ANOVA) confirmed that there were significant differences among the vowel identification rates for each vowel condition in the presence of the band noise at SNRs of –6.5 \( \left[ F(3,156) = 0.36, p < .0001 \right] \) and –9.5 dB \( F(3,156) = 4.33, p < .01 \). At an SNR of –6.5 dB, a Tukey-Kramer multiple comparison test confirmed that the identification rates for the vowel conditions with the F0 fluctuation of natural vowels, namely conditions C and D, were significantly higher than the rates for the conditions without it, namely conditions A and B \( \left[ \alpha = .05 \right] \). At an SNR of –9.5 dB, the identification rate for vowel condition D was significantly higher than the rates for
vowel conditions A and C \[\alpha = .05\]. At an SNR of \(-12.5\) dB, there was no significant difference between the vowel identification rates in the presence of band noise \[F(3,156) = 1.39, p > .10\]. In contrast, in the presence of a harmonic complex tone, there were significant differences among the identification rates at every SNR \[F(3,156) = 27.12, p < .0001\], for 0 dB SNR; \(F(3,156) = 21.81, p < .0001\), for \(-3\) dB SNR; \(F(3,156) = 18.39, p < .0001\), for \(-6\) dB SNR. At every SNR, a Tukey-Kramer multiple comparison test confirmed that the identification rate for vowel condition D was significantly higher than that for the other vowel conditions, and that the rate for vowel condition C is significantly higher than the rate for vowel conditions A and B \[\alpha = .05\].

3.4. Discussion

In spite of the small F0 fluctuation (5.4% of the mean F0), in the presence of the harmonic complex tone, the result suggests that the F0 fluctuation of the natural vowels significantly increased the vowel identification rates. In addition, the vowel identification rates for vowel condition D were always (at every SNR) significantly higher than the rates for vowel condition C. Since the differences between the two vowel conditions was related to whether or not the vowel includes the AM for each frequency band, this result suggests that the AM for each frequency band also increased the vowel identification rate. However, there were no significant differences between the identification rates for vowel conditions A and B. Therefore, the AM for each frequency band of the natural vowels increased the vowel identification rates only when the vowels had F0 fluctuation in the presence of the harmonic complex tone as the interferer sound. In contrast, when there was band noise present, the result suggests that the F0 fluctuation of the natural vowels also increased the vowel identification rate. However, the AM for each frequency band of the natural vowels did not increase the vowel identification rates significantly in any case.

The probable reason why the vowel identification rates were increased by the F0 fluctuation and AM of the natural vowels in the presence of the interferer sounds is as follows. A reduction in the audibility threshold of the vowels because of the F0 fluctuations and AM of the natural vowels may account for the effect. In other words, there is a possibility that the F0 fluctuations and AM made it easier for the human auditory system to detect the vowels from the interferer sounds, and that this increase of detectability led to the effect. The phenomenon whereby the coherent AM of noise reduces the signal detection threshold in noise is well known as CMR \[2\]. As described in the introduction, CMR is considered because of the auditory grouping caused by detecting the correlation among modulations for each frequency band from the co-modulation of noise. To confirm whether the increase in vowel identification rates was related to such masking release, the masked thresholds of vowel detection with the interferer sounds were measured as a function of synthesized male vowel conditions in the following experiment. It should be noted that the AMs for each frequency band of the natural vowels were not modulated coherently, although their harmonics fluctuated coherently.

3.5. Verification experiment: Masked thresholds

3.5.1. Stimuli

The four vowel conditions and two types of interferer sounds in the experiment described earlier were also used in this experiment. The synthesized male vowel /a/ was used as a signal vowel. The band noise was presented at 66.5 dB SPL and the harmonic complex tone at 60 dB SPL. These sound pressure levels correspond to the levels in the preceding experiment at SNRs of \(-6.5\) dB for the band noise and 0 dB for the harmonic complex tone. The power of each interferer sound was constant for all vowel conditions. Stimuli were produced by adding one of the interferer sounds to the signal vowel. The stimuli included a 10 ms taper and the duration was 300 ms. The sampling rate of the stimuli was 48 kHz.

3.5.2. Procedure and subjects

The stimuli were presented diotically through headphones (STAX SR-A Signature) in a soundproof room. Masked thresholds for the vowel were determined in a two-interval, two-alternative forced-choice (2AFC), three-down-one-up procedure, estimating 79.4% detection \[8\]. The initial step size of 2 dB was reduced to 1 dB after three reversals. A run was terminated after 12 reversals, and the average of the last eight reversals was taken as the threshold. Each trial was preceded by a 300-ms pure tone of 1,000 Hz at 40 dB SPL. Intervals were one-second silences. Each subject had two sessions for each vowel condition. A computer was employed to produce the stimuli and to control the experiment.

Twenty subjects participated in this experiment, 10 male and 10 female Japanese between 19 and 22 years old. All subjects were paid. None had a history of hearing difficulties. Half the subjects had sessions for vowel conditions without AM, namely vowel conditions A and C, and the other half had
among the vowel identification rates for each vowel condition, band noise was present, in spite of the significant differences thresholds for vowel conditions C and D. In addition, when the condition C, there were no differences between the masked condition D was significantly higher than that for vowel release. Although the vowel identification rate for vowel in the vowel identification rates was not due to such a masking however, at the same time, the result suggests that the increase is, a masking release because of the F0 fluctuation of the vowels. presence of the harmonic complex tone, the effect whereby the F0 fluctuation of natural vowels increased the vowel identification rates. The result strongly suggests that the F0 fluctuation of natural vowels significantly increases vowel identification rates in the presence of interferer sounds. In addition, the AM of natural vowels also increased vowel identification rates. An experiment was undertaken to verify whether a reduction in the masked threshold because of the F0 fluctuation and AM of natural vowels accounts for this increase in vowel identification. The masked thresholds were measured as a function of the vowel conditions. The results showed that the thresholds were different even when the identification rates were the almost same. In contrast, the thresholds were the almost same even when the identification rates were different. Thus the results indicate that vowel identification rates do not correlate with directly with the masked thresholds in the presence of the interferer sounds. Therefore, the effect of the F0 fluctuation and AM of natural vowels in increasing vowel identification rates must be due to something other than masking release.

3.5.3. Results

The thresholds for each vowel condition, interferer sound, are presented in Fig. 2. In the presence of the band noise, an ANOVA confirmed that there was no difference among the masked thresholds for each vowel condition \[ F(3,76) = 2.01, p > .01 \]. In contrast, when the harmonic complex tone was present, there were significant differences between the masked thresholds for each vowel condition \[ F(3,76) = 9.43, p < .0001 \]. A Tukey-Kramer multiple comparison test confirmed that the masked thresholds for vowel conditions C and D were significantly lower than those for vowel conditions A and B \[ \alpha = .05 \].

3.5.4. Discussion

When the harmonic complex tone was present, the F0 fluctuation of natural vowels caused a significant decrease in the masked thresholds. Therefore, the result suggests that, in the presence of the harmonic complex tone, the effect whereby the F0 fluctuation of the natural vowels increased the vowel identification rates was due to a reduced masked threshold; that is, a masking release because of the F0 fluctuation of the vowels. However, at the same time, the result suggests that the increase in the vowel identification rates was not due to such a masking release. Although the vowel identification rate for vowel condition D was significantly higher than that for vowel condition C, there were no differences between the masked thresholds for vowel conditions C and D. In addition, when the band noise was present, in spite of the significant differences among the vowel identification rates for each vowel condition, there were no significant differences among the masked thresholds. Therefore, the results of these experiments suggest that the effect of the F0 fluctuation and AM of natural vowels in increasing vowel identification rates was due to reasons other than the masking release such as CMR.

4. CONCLUSION

This paper described findings showing that F0 fluctuation and AM of natural vowels increase vowel identification rates in the presence of interferer sounds. In a vowel identification experiment, vowel identification rates were measured as a function of the vowel conditions in the presence of interferer sounds. The result strongly suggests that the F0 fluctuation of natural vowels significantly increases vowel identification rates in the presence of interferer sounds. In addition, the AM of natural vowels also increased vowel identification rates. An experiment was undertaken to verify whether a reduction in the masked threshold because of the F0 fluctuation and AM of natural vowels accounts for this increase in vowel identification. The masked thresholds were measured as a function of the vowel conditions. The result showed that the thresholds were different even when the identification rates were the almost same. In contrast, the thresholds were the almost same even when the identification rates were different. Thus the results indicate that vowel identification rates do not correlate with directly with the masked thresholds in the presence of the interferer sounds. Therefore, the effect of the F0 fluctuation and AM of natural vowels in increasing vowel identification rates must be due to something other than masking release.

Acknowledgements: The authors thank Dr. Makio Kashino for helpful advices and discussions on the design of the experiments and the direction of this research. They also thank Dr. Norihiro Hagita, Dr. Hiroshi Murase, and Dr. Ken’ichiro Ishii for supporting our research.

REFERENCES