Improving air traffic control speech intelligibility by reducing speaking rate effectively

Hou, Nana; Tian, Xiaohai; Chng, Eng Siong; Ma, Bin; Li, Haizhou

2017

Hou, N., Tian, X., Chng, E. S., Ma, B., & Li, H. (2017). Improving air traffic control speech intelligibility by reducing speaking rate effectively. Proceedings of the 2017 International Conference on Asian Language Processing (IALP). doi:10.1109/IALP.2017.8300578

<https://hdl.handle.net/10356/144767>

<https://doi.org/10.1109/IALP.2017.8300578>

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/IALP.2017.8300578>

Downloaded on 26 Apr 2023 22:45:40 SGT

Improving Air Traffic Control Speech Intelligibility by Reducing speaking Rate Effectively

Nana Hou¹, Xiaohai Tian^{1,2}, Eng Siong Chng^{1,2}, Bin Ma³, Haizhou Li^{2,4}

¹*School of Computer Science and Engineering, Nanyang Technological University, Singapore*

²*Temasek Laboratories, Nanyang Technological University, Singapore*

³*Human Language Technology Department Institute for Infocomm Research (I2R), A*STAR, Singapore*

⁴*Department of Electrical and Computer Engineering, National University of Singapore, Singapore*

Email: {nana001, XHTIAN}@e.ntu.edu.sg

Abstract—Low intelligibility of Air Traffic Control (ATC) speech is one major cause of aircraft accidents every year. Many factors can affect speech intelligibility, among which the most prominent aspects is the high speaking rate commonly present in ATC speech. Hence, a possible solution would be to improve intelligibility by artificially lengthening the spoken utterance to lower the speaking rate.

In this work, we explore the lengthening of clean recorded ATC utterances by first identifying phoneme sequences in a given utterance. Such identified phoneme segments can then be lengthened. We will examine effects of lengthening vowels-only, consonants-only, or homogeneous lengthening. To verify our approach, we will conduct human listening test to evaluate the intelligibility. The results show 74.67% was obtained in AB preference test.

Keywords—speech intelligibility, speaking rate, vowels and consonants, air traffic control

I. INTRODUCTION

Speech intelligibility (SI) measure the comprehension of a listener to the spoken utterances and conveys a measure to indicate how much a listener understands the received speech [1]. Intelligibility in air traffic control (ATC) speech is becoming extremely important as investigations have found that low intelligibility in communications between pilots and ground controllers have resulted in considerable number of aircraft accidents [2]. Many factors contribute to low intelligibility of ATC speech, e.g the quality of spoken utterance, background or channel noise, speech accent, fast speaking rate and so on [3]. Among these, fast speaking rate is an important factor affecting ATC speech intelligibility.

speaking rate is a term measuring how fast we talk. Calculating the number of syllables per seconds is commonly used to measure speaking rate [4]. As speaking rate increases, fast speakers blur many phonetic characteristics and syllables. For stressed vowels, their duration will be reduced sharply and for distressed vowels, although their duration is as short as that of sped up stressed vowels, their amplitude and fundamental frequency still are reduced to some extent [5]. For consonants, their energies are also decreased, and their duration are similarly shortened like vowels but are reduced less [6]. speaking rate also affects longer segments of speech (more than one word), e.g, stressed groups in utterances - the stressed group's syllable duration are both decreased and realized incompletely. In addition, phrase groups are also affected - too much

co-articulation and fundamental frequency deviation make their boundaries decline and hard to reach clear pronunciations. All these phenomena affect speech intelligibility as speed rate increases [7, 8].

Previously, researchers have proposed to use text-to-speech (TTS) system in place of human generated speech, i.e, the pilots and ATC types text which are then generated by TTS system to overcome the issues of accent and speaking rate [9]. To date, we have not found studies which reduce ATC speaking rate to improve intelligibility. This lack of prior work may be attributed to the fact that ATC speech is usually very noisy, and it is not straightforward to analyze and stretch such utterances.

In this work, we aim to improve the intelligibility of fast ATC speech for human listening by artificially stretching a given utterance. We examine the process to lengthen a given utterance by lengthening identified vowels and consonants separately, as well as homogeneously to determine their impact on speech intelligibility. We will only focus on modifying vowel and consonant phonemes duration without considering supra-segmental aspects of articulation.

Our approach to lengthen a given utterance is described as follows:

- 1) The STRAIGHT [10] features are used to analyse the original given speech. This process converts the temporal signal into frames of STRAIGHT features.
- 2) The phonetic posteriorgram (PPG) is used to identify vowel and consonant class on a frame level and then applied on the STRAIGHT features for lengthening and scratching.
- 3) The identified frames belonging to vowels and/or consonants class can now be modified independently. E.g, if we wish to lengthen consonant segments, frames belonging to the consonant class will be duplicated and inserted into the array of frames to realize the process to lengthen the particular sound. We will explore vowel-only, consonant-only, and all phoneme lengthening.
- 4) The modified STRAIGHT features are then used to synthesis the lengthen utterance.

To verify our approach, the AB test is performed on the original fast utterances vs the various lengthened synthesized utterances to a) reduction of speaking rate can

improve intelligibility, b) if vowel-only, consonant-only, or homogeneous lengthening will yield better intelligibility.

The organization of the paper is as follows: Section two discusses the reasons for fast ATC speech and its phonemena. Section three explains our experimental setup to reduce speaking rate as well as listening test and the found results. Section four concludes the paper.

II. ATC SPEAKING RATE

One common measure of speaking rate is to count the number of syllables spoken per second (sps). The normal speaking rate has been reported to be in the range of approximately 4 sps [11]. However, for ATC speech, it has been shown that speaking rate is significantly faster, e.g, at rates above 6 sps is not uncommon and contributes significantly to low intelligibility.

The reason for fast ATC speech has been investigated by many works, e.g, [11] proposed that pilots begin to speak faster from the period when they are trainees to achieve proficiency qualification requirement of International Civil Aviation Organization (ICAO). ICAO quantifies the proficiency and fluency of language to six levels and trainees must attain level four and above to pass. As such, non-native English speakers gradually speak faster to obtain higher scores in order to achieve required fluency. For native English speakers, their articulation rate became even faster.

Past works have investigated how to reduce fast speaking rate effectively [12]. Based on [5], there are four acoustic factors that directly influence speech intelligibility - duration, overall amplitude, fundamental frequency and vowel color. Their experiments show that duration of each phoneme is affected most, with vowel duration more affected than consonant. This agrees with [6, 13] who also reported the similar findings. These works inspire us to examine approaches to reduce speaking rate by considering different rates to slow down speech based on their phoneme class.

III. REDUCING THE SPEAKING RATE

In this work, we specifically focus on the analysis of speech features to vowel and consonant class for improving the speech intelligibility.

The system diagram (Figure 1) explains our phoneme lengthening method. Our method is implemented in spectrum level and consists of three stages: speech analysis, speech scratch and speech synthesis. In speech analysis stage, frame-based Phonetic PosteriorGrams (PPGs) feature and the corresponding Vocoder features were first extracted. In speech scratch stage, the PPGs were used to identify the vowel and consonant class for each frame. Then the scratch process was applied on the corresponding Vocoder features. At synthesis stage, the resultant speech was reconstructed by the scratched Vocoder features.

In our implementation the STRAIGHT method [10] is based on pitch-adaptive spectral analysis and fundamental frequency (F0) calculating by instantaneous frequency. It not only can protect information in time-domain but also

Figure 1: System diagram of speaking rate reduction.

Table I: Text of every utterance and corresponding reading speed.

S/No.	Text	Fast speed (sps)	Normal speed (sps)
1	five four zero three vacate left and turn left, whiskey papa, contact Singapore ground one two four point three	6.0	3.9
2	Thresh star ratio five niner six, surface wind two zero zero degrees eight knots, runway two zero right, clear to land	6.1	4.1
3	Redcap seven zero six line up, runway two zero right via whiskey two	6.0	4.2
4	Center niner seven niner one, vacate whiskey seven, turn left whiskey papa	6.3	4.0
5	Singapore two three eight, Singapore tower, good afternoon, continue your approach, runway two zero right	6.2	4.3

produce high-quality features. Thus, features extracted by STRAIGHT method is suitable for exploring relations between physical promoters and human perception.

Our setup of STRAIGHT and PPG features are as follows: 513-dimensional spectrum were extracted using STRAIGHT method, then concatenated with 5 band aperiodicity measures and $\log[F_0]$. The corresponding acoustic features (e.g. MFCC) were feed into a pretrained phoneme recognizer (PPG) to obtain a 42-dimension PPG frame to identify vowel and consonant class for each frame.

IV. EXPERIMENTAL RESULTS

A. Database

For the experiments, we develop a small corpus of a single male L1 speaker reading five ATC type of exchange in two different speeds, i.e fast speaking rate and normal speaking rate. The details of the utterance and reading speed is in Table I. The speaking rates of 6 syllables per second for fast speech and 4 syllables a second for normal rate agree with previous published works [11]. The fast speaking rate utterances are used as sourced utterances in Experiment 1 (Table II) while the normal speech utterances are the target X examples in Experiment 2 (Table III).

B. Reducing Speaking Rate by 3 criteria

To measure if speaking rate reduction is effective, we apply the proposed method of analysis and re-synthesis to lengthen our original fast speech utterances to have the same duration as the corresponding normal speed

utterances. We examine three different ways to lengthen the duration, specifically, we duplicate frames belong to either consonant-only, vowel-only, or both-consonant and vowel jointly. We call these three methods:

- 1) lengthening vowels-only,
- 2) lengthening consonant-only, and
- 3) lengthening both vowel+consonant.

In other words, for the case of vowel only lengthening, only frames belonging to vowels are replicated. These frames belonging to vowels are identified through the PPG features. After lengthening, the desired stretched utterance is reconstructed by synthesis.

In total, five fast speed rate utterances were converted to 3 corresponding sets with normal speed rate by the above criteria. In the next section, we discussed the listening tests setup.

C. Listening Test

To evaluate the effectiveness if speaking rate reduction improve intelligibility, listening tests by 5 listeners were conducted. Two types of test - AB and XAB were conducted.

The AB test is the intelligibility test that compares 5 listeners' preference between the original fast speech to the three lengthened speech by the 3 criteria - vowel-only, consonants-only, and vowel+consonant. For each pair of sentence, the listeners are asked to choose their preference only based on the intelligibility criterion. The objective of this test is to reveal if artificially reducing the speaking rate is effective, and if so, which criteria for stretching is preferred. During the listening tests, each A,B utterances are played only once, and their order are presented randomly to the listener.

For the XAB tests, also known as similar test, the listener are presented with the normal speaking rate of the utterance known as X, as well as lengthened utterance by 2 different criteria known as A and B condition. The objective of this test is to determine which is the preferred criteria (A or B) for lengthening such that the resultant generated signal sounds closer to the target utterance. For the XAB trials, we conducted 3 groups of XAB trials for the 3 criteria of lengthening:

- 1) In first XAB test, A is consonant-only lengthening constructed utterance, and B is vowel+consonant lengthening.
- 2) In the second XAB test, A is consonant-only lengthening, and B is vowel-only lengthening.
- 3) In the third test, A is vowel-only lengthened utterance and B is is vowel+consonant lengthening. .

During the listening tests, X utterances were played first once, followed by A or B utterances randomly only once. The listeners are then asked to decide which sample is closer to the normal utterance. The results of the listening test are discussed in the following sub-section.

D. Result

Table II summarizes the results for the intelligibility tests, which aim to reveal if lengthened synthesized ut-

Table II: Results of intelligibility preference tests with 95% confidence intervals for different sample groups

Intelligibility test (AB test)	Preference score(%) (95% confidence interval)
A) original fast utterance B) lengthening Vowel and Consonant	16 84
A) original fast utterance B) lengthening Vowel	28 72
A) original fast utterance B) lengthening Consonant	32 68

Table III: Results of intelligibility preference tests with 95% confidence intervals for different sample groups

Similar test (XAB test)	Preference score(%) (95% confidence interval)
A) lengthening Consonant B) lengthening Vowel and Consonant	36 64
A) lengthening Consonant B) lengthening Vowel	64 36
A) lengthening Vowel B) lengthening Vowel and consonant	40 60

terances yield better intelligibility than original fast utterances. Each row of the table compares the three criteria of lengthening, and the corresponding right column shows the preference score between the AB trials. The results clearly show that all the B-utterances are preferred - that is, artificially reducing the speed improves intelligibility. Finally, the average preference score of 74.67% was obtained in AB preference test. Additionally, the criterion of jointly lengthening vowel and consonant was preferred over the other two criteria.

Table III summarizes the similarity test. The objective here is to determine which lengthening criteria will generate utterances closer to the actual normal speed utterance. Each row of the table indicate the two AB criteria used to generate the lengthened utterance from fast utterance and the corresponding right column shows the preference score between the AB trials. The results show that listeners prefer joint vowel and consonant lengthening criterion followed by consonant lengthening criterion, and there is no significant difference between selection of vowel or consonant lengthening criterion.

V. CONCLUSION

This paper proposed a method to reduce the speaking rate in ATC domain by identifying vowels and consonants and lengthening them to test whether it can improve speech intelligibility. The subjective evaluation results show that the proposed speaking rate reduction approach will improve intelligibility, and that jointly lengthening vowel and consonant, or consonant only criteria are preferred.

Our paper has only carried out a preliminary investigation of speaking rate reduction of ATC speech. The followings list possible future works:

- 1) Different approaches for speaking rate reduction.

- 2) Currently, this study has been conducted on clean speech. Further studies on actual noisy ATC utterances should be carried out.

ACKNOWLEDGMENT

The author would like to thank Air Traffic Management Research Institute of Nanyang Technological University for supporting this project.

REFERENCES

- [1] Jae Lim. "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.5 (1978), pp. 471–472.
- [2] DAVID O'HARE et al. "Cognitive failure analysis for aircraft accident investigation". In: *Ergonomics* 37.11 (1994), pp. 1855–1869.
- [3] María TP Zaballo et al. "Air traffic controllers' long-term speech-in-noise training effects: A control group study". In: *Noise & health* 18.85 (2016), p. 376.
- [4] Frieda Goldman-Eisler. "The significance of changes in the rate of articulation". In: *Language and Speech* 4.3 (1961), pp. 171–174.
- [5] Thomas Gay. "Effect of speaking rate on vowel formant movements". In: *The journal of the Acoustical society of America* 63.1 (1978), pp. 223–230.
- [6] Rob JJH Van Son and Louis CW Pols. "An acoustic profile of consonant reduction". In: *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*. Vol. 3. IEEE. 1996, pp. 1529–1532.
- [7] Donata Moers et al. "Integrating a fast speech corpus in unit selection speech synthesis: experiments on perception, segmentation, and duration prediction". In: *Speech Prosody 2010-Fifth International Conference*. 2010.
- [8] Valerie Hazan and Duncan Markham. "Acoustic-phonetic correlates of talker intelligibility for adults and children". In: *The Journal of the Acoustical Society of America* 116.5 (2004), pp. 3108–3118.
- [9] Thierry Dutoit. *An introduction to text-to-speech synthesis*. Vol. 3. Springer Science & Business Media, 1997.
- [10] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain De Cheveigne. "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds". In: *Speech communication* 27.3 (1999), pp. 187–207.
- [11] H Said. "Phraseology Report. Pilots/Air Traffic Controllers Phraseology Study. IATA. IFALPA". In: *International Air Transport Association* (2011).
- [12] Akiko Amano-Kusumoto and John-Paul Hosom. "A review of research on speech intelligibility and correlations with acoustic features". In: *Center for Spoken Language Understanding, CSLU-011-001* (2011).
- [13] James M Hillenbrand, Michael J Clark, and Robert A Houde. "Some effects of duration on vowel recognition". In: *The Journal of the Acoustical Society of America* 108.6 (2000), pp. 3013–3022.