# The role of F0 variation in the intelligibility of Mandarin sentences

Aniruddh D. Patel[1], Yi Xu[2], and Bei Wang[3]

1. The Neurosciences Institute, 2. Univ. College London, 3. Minzu Univ. of China
(apatel@nsi.edu, yi.xu@ucl.ac.uk, bjwangbei@googlemail.com)

**Abstract**
This study tested the importance of F0 variation for tone language comprehension. The intelligibility of Mandarin sentences with natural F0 contours was compared to the intelligibility of monotone (flat-F0) sentences created via speech resynthesis. In a quiet background, flat-F0 speech was just as intelligible as natural speech (about 94% intelligible), highlighting the robustness of the language comprehension system. However, when babble noise was added (0 db SNR) flat-F0 speech was substantially less intelligible than natural speech (60% vs. 80% intelligible), indicating that F0 variation is very important for Mandarin sentence intelligibility in noise.

## 1. Introduction

A number of recent studies have examined the role of fundamental frequency (F0) contours in the intelligibility of English sentences (e.g., Binns & Culling, 2007; Watson & Schlauch, 2008). These studies have found that intact F0 contours improve the intelligibility of speech in noise, compared to monotone (flat) or inverted F0 contours. However, to date no such study has examined tone languages. In tone languages the lexical meaning of most content words depends on their F0 pattern, and thus F0 variation seems likely to be quite important to sentence intelligibility. Indeed, theoretical studies suggest that the functional load of tones in distinguishing Mandarin words is as high as that of vowels (Surendran & Levow, 2004).

The current study examined the intelligibility of Mandarin sentences with natural vs. flat-F0 contours. In addition to studying intelligibility in noise, we also examined intelligibility in a silent background, to address whether F0 variation is crucial for intelligibility in Mandarin. Based on studies of whispered sentences, it is known that Mandarin without F0 information can be highly intelligible in a quiet background. However, whispered speech contains cues to tone other than F0, such as duration and amplitude (Holbrook & Lu, 1969; Liu & Samuel, 2004). Indeed, these cues may be promoted by speakers when whispering. Unlike whispered speech, flat-F0 speech has voicing, and studies indicate that that when F0 is present it provides the dominant cue for tone perception (e.g., Whalen and Xu, 1992). Thus the intelligibility of flat-F0 Mandarin cannot be predicted from studies of whispered speech.

A neuroimaging study of the perception of flat-F0 vs. normal German sentences indicates that flat-F0 speech elicits greater activation in a number of brain regions in both hemispheres (Meyer et al., 2004). This suggests that processing flat-F0 speech is cognitively demanding, even in an intonation language. Hence one might expect that flat-F0 speech in tone languages such as Mandarin would be even more cognitively demanding, and thus may be less intelligible than natural speech even in a quiet background.

## 2. Methods

40 news-like sentences from the corpus of Nazzi et al. (1998) were translated into Mandarin and read by a male native speaker (author YX). Sentences were 18 syllables long on average (mean duration 3.2 s, mean rate 5.6 syll/s). Mean F0 was 131 Hz. The "pitch width" of each sentence [= $12*\log_2(\text{maxF0}/\text{minF0})$] was 15.4 semitones on average. Table 1 shows 3 example sentences from the study.

**Table 1.** Example sentences from this study

| | |
|---|---|
| A hurricane was announced this afternoon on the TV. | 今天下午的电视里说明天要有台风。 |
| The last concert given at the opera was a tremendous success. | 这家歌剧院里的上一场音乐会十分成功。 |

| | |
|---|---|
| The latest events have caused an outcry in the international community. | 最近的几个事件已经引起了国际社会的强烈反响。 |

Sentences were digitized at 44.1 kHz and F0 analyses and manipulations were done using Praat. A flat-F0 contour was created for each sentence at the sentence's mean F0 using Praat's pitch manipulation tools. The resulting flat-F0 sentence was resynthesized using the PSOLA (pitch-synchronous overlap and add) method. To control for possible changes in voice quality introduced by resynthesis, the original sentences were also resynthesized using their extracted F0 contours, and these "natural-F0" resynthesized sentences were used in the perception experiment. (Informal comparison of the natural-F0 and original sentences suggested no noticeable changes in voice quality.)

Each sentence was combined with babble noise at 4 SNR levels (no noise, +5, 0, -5 dB SNR), and amplitude-normalized to .5 volts RMS, using SIGNAL (Engineering Design). 6-talker Mandarin babble noise was used, consisting of 3 females and 3 males reading semantically anomalous sentences (Van Engen & Bradlow, 2007). This resulted in 320 stimuli (40 sentences x 2 F0 types x 4 noise levels). The stimuli were organized into 8 lists such that each of the 40 original sentences occurred in each list, distributed equally across the SNR and F0 conditions (i.e., 10 sentences at each SNR level, 5 with natural-F0 and 5 with flat-F0). Since each participant heard just one list, to achieve proper counterbalancing (so that every sentence appeared in every condition, across participants), each list was used equally often in the study. Within each list, sentences were ordered so that background noise gradually increased in intensity (i.e., sentences 1-10: no noise; 11-20: +5 dB SNR; 21-30: 0 dB SNR; 31-40: -5 dB SNR). Furthermore, natural-F0 and flat-F0 sentences were presented in an alternating fashion.

Participants (n=24, 12 female, mean age 21.6 y) were native Mandarin speakers enrolled as students at Beijing Normal University. Listeners were tested individually in a quiet room while facing a computer monitor, and heard sentences over Edifier R18 loudspeakers at a comfortable listening level. The instructions were to "listen to each sentence and write down the words you hear". Listeners pressed a key to advance to the next trial, and wrote their responses on paper. Each sentence could only be heard once. Example sentences were provided before the experiment, sampling all conditions.

Author BW transcribed the written responses into typed format, then author YX scored the responses (both authors were blind to the condition that each response came from). A perfect score (1) was given to sentences with no errors or only wrong homophone characters (equivalent to spelling errors in English). For sentences containing errors, the intelligibility score "s" was assigned as shown in Table 2.

**Table 2** Intelligibility scoring. The score for each sentence was assigned as:
s = (N – S – M – I – O) / N, where N is the number of syllables (equivalent to morphemes in Chinese), S is the total number of syllables with either wrong consonant, wrong vowel, or wrong tone, M is the total number of missing syllables, I is the total number of wrongly inserted syllables, and O is the total number of wrong syllable orders. In the example below, the first three syllables lǐ shì huì were omitted, yào was moved leftward, and zài was inserted.

| Error code | MMM O I |
|---|---|
| Answer | 。。。要在今天下午开会举行专题辩论 |
| Pinyin | 。。。yào zài jīn tiān xià wǔ kāi huì jǔ xíng zhuān tí biàn lùn |
| Original | 理事会今天下午要开会举行专题辩论 |
| Pinyin | lǐ shì huì jīn tiān xià wǔ yào kāi huì jǔ xíng zhuān tí biàn lùn |
| English | The committee will meet this afternoon for a special debate |
| Score | s = (16 – 0 – 3 – 1 – 1) / 16 = 0.6875 |

**3. Results**

Intelligibility scores are presented in Figure 1 as a function of F0 type and SNR level.

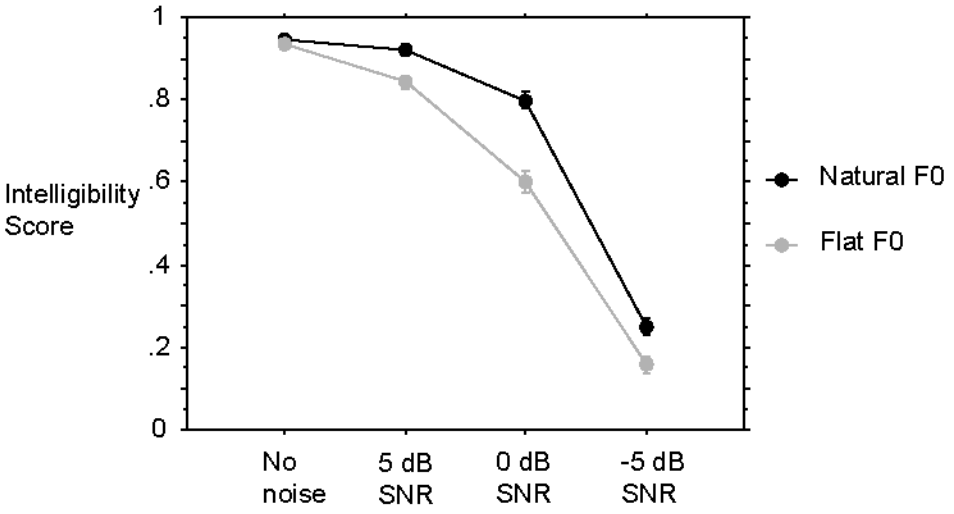


**Figure 1.** Intelligibility of natural-F0 vs. monotone (flat-F0) Mandarin sentences, as a function of background babble noise. SNR = Signal-to-noise ratio. Error bars represent standard errors.

A two-factor analysis of variance revealed main effects of both F0 and SNR, i.e., intelligibility was reduced by flat F0 ($F(1,23) = 63.22$, $p < 0.001$) as well as by noise ($F(3,23) = 659.02$, $p < 0.001$). A significant interaction indicated that intelligibility degraded faster with increasing noise for flat-F0 vs. for natural-F0 sentences ($F(3,23) = 12.24$, $p < 0.001$). More surprising was the high intelligibility of flat-F0 speech without noise. Indeed, flat-F0 speech was just as intelligible as natural-F0 speech in this condition (both ~ 94% intelligible). A 2-tailed t-test showed no difference between the two (flat-F0 = 93.6%, natural-F0 = 94.5%, $p = 0.54$). In conditions with noise, flat-F0 speech was significantly less intelligible than natural-F0 speech, as shown by Bonferroni-corrected 2-tailed t-tests (SNR = +5 dB: 84.3% vs. 92.1%, $p < 0.002$; SNR=0 dB: 60.1% vs. 79.8%, $p < 0.001$, SNR=-5 dB: 15.8% vs. 25.0%, $p < 0.005$).

**4. Discussion**

The current study removed all pitch variation from sentences of Mandarin Chinese, a tone language in which pitch makes lexical distinctions between words. Despite this severe manipulation, the resulting monotone sentences were as intelligible as speech with natural F0 contours, when heard in a quiet background. This finding highlights the robustness and flexibility of spoken language comprehension. The comprehension system presumably uses the remaining segmental information, plus secondary cues to tone (such as amplitude and duration), and lexical constraints provided by sentence context, to aid speech perception in the face of a degraded speech signal. The current finding appears to support models which include an important role for top-down information in guiding speech perception in a predictive manner (Poeppel et al., 2008). One way to test this idea in future work would be to repeat the current study using semantically anomalous sentences (e.g., “Colorless green ideas sleep furiously”). If lexical prediction plays an important role in helping the speech perception system compensate for lack of pitch variation, the intelligibility of semantically anomalous flat-F0 speech should be substantially lower than that of semantically anomalous sentences with natural F0 contours, even in the absence of noise.

Consistent with prior work on English, the current study found that when babble noise was added to tone language sentences, the intelligibility of flat-F0 speech was lower than that of natural-F0 speech. Of particular interest in this regard is the large intelligibility difference between natural-F0 and flat-F0 speech at 0 dB SNR (80% vs. 60%), when signal and noise were equal in amplitude. This difference appears to be more than twice the size of the intelligibility difference between natural-F0 and flat-F0 English when heard in white noise at 0 dB SNR (Watson & Schlauch, 2008). While direct comparison of the current study and that of Watson & Schlauch (2008) is difficult due to different types of noise, different stimulus materials (they used sentences from the SPIN test), and different methods for intelligibility scoring,

it seems plausible that flat F0 contours would be more detrimental to speech perception in noise for tone vs. intonation languages. Further experiments are needed to test this possibility directly.

Finally, based on the current findings we can make a prediction about speech perception in individuals with congenital amusia. Amusia is a deficit of music processing which can also impair linguistic pitch contour perception (Patel et al., 2008; Liu et al., submitted). In a quiet setting, where pitch variation may not be essential for normal sentence comprehension, amusics may exhibit no loss of speech intelligibility. However, we predict that in noisy environments amusics will suffer a greater loss of speech intelligibility than non-amusic listeners, due to their deficits in linguistic F0 pattern perception. Perceptual research with amusics on speech intelligibility in quiet and in noise can be used to test this idea empirically.

**Acknowledgments**

We thank Kristin Van Engen and Ann Bradlow for providing Mandarin babble noise, and John Iversen for comments. Supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior Fellow, and in part by NIH Grant No. DC006243 to YX. Thanks to Prof. Hua Shu at Beijing Normal University for providing facilities for the perception experiment.

**References**

Binns, C., & Culling, J.F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *J. Acous. Soc. America,* 122:1765-1776.

Holbrook, A. & Lu H-T. (1969). A study of intelligibility in whispered Chinese. *Speech Monographs,* 36: 464-466.

Liu, F., Patel, A.D., Fourcin, A., & Stewart, L. (submitted). Speech processing in congenital amusia: Production, perception, & imitation.

Liu, S. & Samuel, A.G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Lang. & Speech,* 47:109-138.

Meyer, M., Steinhauer, K., Alter, K.., Friederici, A.D., & von Cramon, D.Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain & Language,* 89:277-289.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination in newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance,* 24(3):756-777.

Patel, A.D, Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception,* 25:357-368.

Poeppel, D., Idsardi, W.J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Phil. Trans. of The Royal Soc. B.,* 363:1071-1086.

Surendran, D. & Levow, G-A. (2004). The functional load of tone in Mandarin is as high as that of vowels. *Proc. Intl. Conf. Speech Prosody*, 1:99–102.

Van Engen, K., & Bradlow, A.R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acous. Soc. America,* 121:519-526.

Watson, P.J. & Schlauch, R.S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *Am. J. of Speech-Language Pathology,*17:348-355.

Whalen, D.H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica,* 49:25-47.