English Learners’ Phonation in Chinese Narrow Focus Syllable “-/a/ / C_”

YANG Jin¹, KONG Jiangping

1 Dept. of Foreign Languages
USST, SHISU, PKU
Shanghai, China

Abstract
This paper conducts acoustic experiments from a signal processing perspective, to examine the “-/a/ /C_” syllable narrow focus processing strategy of English L2 learners. The research discusses F0, OQ and SQ as key phonatory acoustic parameters. L2 strategies are examined by measuring deviations from L1 group to L2 group. The matrix distance reflects L1’s traces in the interlanguage system.

Keywords: F0, OQ, SQ, narrow focus, L2

1. Introduction
As in the past, phonetic researches on narrow focus are mainly focused on the acoustic parameters such as F0, duration and amplitude. Ladeforged (1982) and Fant (2004) claimed that power is an important parameter for judging voice quality and is closely related with F0 change. Chen (1974) found that the average pitch range of the four Chinese speakers was at least 1.5 times wider than that of the four English-speaking subjects when they spoke their native languages. Zhang et al. (2008) reported Mandarin speakers’ production of lexical stress contrasts in English is influenced partly by native-language experience with Mandarin lexical tones, and partly by similarities and differences between Mandarin and English vowel inventories. There are also a few researches on the acoustic parameters of Chinese focus types. However, there seems to be little research on the narrow focus perception studies, not to say phonation studies. To the author’s knowledge, only Yin (2011) used a nine-syllable Chinese sentence to test sentence final narrow focus processing strategy. Within the scope of Chinese sentences of fixed length, his result showed that male speakers mainly process narrow focus by raising high frequency energy while female speakers mainly realize narrow focus by raising low frequency energy. Yin’s research is a pioneering work in Chinese L1 speakers’ phonation pattern on narrow focus. Based on this, our study intends to make one step further. We will add two variables into the current study, i. e., L2 speakers and sentence length. We intend to follow Kong’s (2001) definition of OQ and SQ in his research on Chinese tones and regards F0, OQ and SQ as three key parameters in locating English and Chinese “-/a/ / C_” Structure narrow focus as an important prosodic characteristics and therefore conducts experiments on English L2 learners of medium-high level. And three-dimensional Matlab plots of cross language F0, OQ and SQ are presented for the first time on narrow focus processing. We hope this crisscross study will shed light on the phonation nature of speech production and L2 phonetic acquisition.

2. Method
2.1 EGG

EGG equipment is composed of several components. High Frequency Oscillator is a current-controlled oscillator to yield micro high-frequency current. The current yielded passes through the Electrode Circuit that in our case is glottis and vocal folds. Then there is Automatic Gain Control to ensure a stable and appropriate sized electric signal feedback. At the output end is AM Detector to detect output signal. Output signal at this place is Gx signal which usually contains low-frequency jitter. Due to the fact that most EGG has set High-Pass Filters, Gx signal passed through this filter and changes into Lx signal. Researches show that EGG signal is positively related with vocal folds attack area. When vocal folds attack area (hereafter referred to as V AA) grow, EGG signal becomes stronger. It weakens while V AA decreases. This has been proved by high-speed glottis photography (Fourcin 1974, Baer, T. et al., 1983 & Gilbert, H. R. et al., 1984).

This paper is sponsored by USST Social Science Project 11XSY22, Shanghai Municipality Key Course “Translation” & “Interpretation for Science & Technology” 1K10305001 Project, and National Social Science Junior Project 09CYY017.
2.2 Parameters

F0 or pitch, is the inverse of a vocal folds vibration period. OQ refers to Open Quotient, is the opening phase of the glottis divided by the period as indicated in this formula – OQ = ac / ad. SQ refers to the Speed Quotient, is the opening phase divided by the closed phase as indicated in this formula – SQ = ab / bc. In the following prototypical glottis airflow signal illustrated in Fig. II.B.1 (courtesy of Kong, 2001), which could be analogical to glottis area signal, the abscissa axis indicates a timeline while the vertical axis indicates vocal folds attacking area. The whole figure reflects a change of glottal airflow or glottis area along the timeline. Point ‘a’ and ‘d’ stand for the initial point of glottis opening phase where glottis airflow value augments from zero. ‘b’ is the point when the glottal width reaches a peak value, thus corresponding to a peak area of glottis. ‘c’ represents the closure of vocal folds when glottal airflow resets to zero. Period ‘ac’ is the opening phase of glottis while ‘cd’ is the complete closure phase and the glottal airflow or glottis area is zero. Contacting event and de-contacting event are highlighted as the beginning of a complete closure of the glottis and the opening of the glottis. The area in-between is the closing phase of the glottis.

\[ OQ = \frac{ac}{ad}, \quad SQ = \frac{ab}{bc} \]

![Fig. II.B.1 EGG signal](image)

Another term is CQ – Contact Quotient in EGG signal processing. CQ refers to the ratio of glottis closure phase to the period. Fig. II.B.2 (courtesy of Kong, 2001) is a typical EGG signal. Its abscissa is time and the vertical axis is vocal fold contact area.

\[ CQ = \frac{B}{A} \times 100\% \]

![Fig. II.B.2 Defining EGG Signal](image)

In this figure, A stands for a vocal fold vibration cycle, starting from the closing point of the glottis to the next closing point. It covers a period between two contacting events. B signifies a closed phase. It covers the period between the initial of glottis closure phase and the opening point. It’s between contacting event and de-contacting event. C represents an open phase. It’s between the glottis opening point to the next closure point, i.e., the period between de-contacting event to the next contacting event. D is a closing phase starting from the closure of glottis to vocal folds contact area augments to a peak value. E is the opening phase starting from the maximum contact area to the open point of the glottis. Therefore in this figure, OQ (Open Quotient) = C/A*100%. CQ (Contact Quotient) = B/A*100%. SQ (Speed Quotient) = E/D*100%. Since B + C = A, OQ + CQ = 1. In this sense, the formulas also state the complementary property of OQ and CQ (Henrich 2004).

From the above two illustrations, we can easily achieve this result – sound wave is in a reverse phase to EGG signal. This is because EGG signal is obtained via the resistance between the pair of vocal folds, i.e., it’s determined by the opening degree of the glottis. Therefore, this will lag a semi-period to the sound wave which is yielded from vocal folds attack.

2.3 Participants

Four English and Chinese bilinguals participated in this experiment. They are two male and two female native American English speakers and two male and two female native Beijing dialect speakers. The four American English speakers were born and grow up in California in local American families. The four Chinese speakers were born and grow up in local Beijing families. They speak Beijing dialect whose phonetic system is the basis of Standard Mandarin Chinese, i.e., Chinese Putonghua, the official language of China. All of them are students recruited from Peking University. Their age ranges from 20 to 24. All American students have been studying Chinese for two to four years. Their Chinese proficiency is skilled. Since oral proficiency is not surely positively correlated with written language proficiency, we invited Oral English teachers to evaluate their spoken proficiency. The American students are of medium-high Chinese level.

None of subjects reported having any speech disorders. All subjects received payment.
2.4 Materials

Forty-four sets of question/answer pairs were constructed for this experiment. Questions were designed to set a disambiguating context to point out clearly the sentence final narrow focus marking location. The subject will click on the mouse to continue PPT questions and answer orally. Answers are all statements. All narrow focused syllables are composed of a truly existing consonant followed by /a/, thus 22 syllables in Chinese. Sentences are arranged randomly to appear before subjects, each appear totally for twice to ensure a reliable outcome.

Tab. II. D. 1 contains all the token forms. Ch_l means Chinese default last syllable at the long sentence. Ch_lf means Chinese narrow focus at the last syllable of the long sentence. Ch_s and Ch_sf refer to their counterparts in Chinese short sentences.

<table>
<thead>
<tr>
<th>Ch_l</th>
<th>Zhang Xiaosan zai wei qiang bian de kongdi shang xie le “-/a/ /C_.”. – (English equivalence: Zhang Xiaosan wrote “-/a/ /C_” on the ground by the wall.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch_lf</td>
<td>bu, Zhang Xiaosan zai wei qiang bian de kongdi shang xie le “-/a/ /C_.”. – (English equivalence: No, Zhang Xiaosan wrote “-/a/ /C_” on the ground by the wall.)</td>
</tr>
<tr>
<td>Ch_s</td>
<td>wo xie de shi “-/a/ /C_.”. – (English equivalence: I wrote “-/a/ /C_”).</td>
</tr>
<tr>
<td>Ch_sf</td>
<td>bu, wo xie de shi “-/a/ /C_.”. – (English equivalence: No, I wrote “-/a/ /C_”).</td>
</tr>
</tbody>
</table>

Tab. II. D. 1 Read Speech Material

2.5 Procedure

Trained proctors recorded all speech data. Questions appeared randomly in a laptop computer. Subjects wore Sony clip microphone, EGG neck belt, breast belt, abdomen belt, finger voltage collector and heart pulse collector. The Sony clip microphone is all-directional and located about 15 cm away from their mouths, and they were instructed to speak naturally at a normal rate and volume. Recording are carried out in a sound-attenuated room. The main recording equipment is a 16 bit Myoelectrigraph & Electroencephalograph Information Gatherer produced by Australia Powerlab Company. The recording software is Chart 5 which is carried by the equipment itself. Recording collects two channels of signals: 1) sound file collected by Sony microphone; 2) graph collected by EGG produced by Kay Corporation.

2.6 Results

Altogether 22*8*2*2 = 704 tokens are obtained for analysis. All tokens are normalized to 30 points before averaged according to F0, OQ and SQ separately. Results are shown in the following figures. In each graph, a red line denotes the “-/a/ /C_” syllable cumulative average value at the sentence final narrow focus place. The blue line is for their counterpart in the non-narrow focus sentence final position. Legends are indicated in this model such that “mA_ch_lf_F0ave” indicates male American English speakers’ average F0 across all “-/a/ /C_” syllables at the narrow focus position. “mC” group are male Chinese speakers. However, considering limit space of this paper, we’ll only extract English and Chinese male speakers’ reaction to the stimuli of Chinese long sentence. Other data are shown in Tab. II. F. 1 by function coefficient.

For other data, different levels of variables are fitted into a curve function y = cx^2 + tx + v (c stands for curvature, t for tilt rate and v for intercept). In Tab. II. F. 1, Ch_l stands for Chinese long sentences which contain fourteen syllables. Ch_s stands for Chinese short sentences which contain five syllables. mA is
male American speakers’ narrow focus token. mA is male American speakers’ default token (or, token at non-narrow focus position). Collected data are as follows.

<table>
<thead>
<tr>
<th>F0</th>
<th>OQ</th>
<th>SQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>mA</td>
<td>-0.027</td>
<td>0.55</td>
</tr>
<tr>
<td>mC</td>
<td>-0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>mA</td>
<td>-0.027</td>
<td>2.401</td>
</tr>
<tr>
<td>mC</td>
<td>0.034</td>
<td>-0.975</td>
</tr>
<tr>
<td>mE</td>
<td>-0.057</td>
<td>1.8</td>
</tr>
<tr>
<td>tA</td>
<td>0.017</td>
<td>-1.716</td>
</tr>
<tr>
<td>tC</td>
<td>0.042</td>
<td>-0.874</td>
</tr>
<tr>
<td>tC</td>
<td>0.149</td>
<td>-0.147</td>
</tr>
</tbody>
</table>

Tab. II. F. 1 F0, OQ and SQ parameters

F0, OQ and SQ are also put into a three dimensional plot edited by Matlab software for convenience. In Plot II. F. 1, the red circles represent male American speakers Chinese speech signal, briefed in the legend as “mAC”. The green plus represent male Chinese speakers’ data, the yellow diamonds for female American speakers’ and the blue squares for female Chinese speakers’ signal.

Plot II. F. 1 Matlab 3-D distribution of F0, OQ and SQ

3. Analysis and discussion

First, F0 has a robust effect in narrow focus representation, despite L1 or L2, sentence length or sex difference. For details, the male American speakers’ F0 variation seems smaller than the other groups. In other words, the difference between the highest and the lowest value is comparatively smaller than the other groups. This is true for either at narrow focus position or not.

The common feature of male and female Chinese speakers F0 variation proves that Chinese speakers distinguish clearly between narrow focus and non-narrow focus syllables. The common feature of female Chinese and English speakers F0 variation proves that female speakers share a good distinction on narrow focus. The male American speakers are neither L1 speakers nor female, therefore they are weak in F0 variation for a good narrow focus.

Sentence length is found contributive to narrow focus identification. We found in short Chinese sentences containing five syllables, F0 variation is not obvious as in long sentences. This may due to the fact that each syllable weighs more in the short sentence unit than the long sentence unit (the weight 1/5 > 1/14). This weight is correlated with contribution. So the rest four syllables in the short sentences are not so identifiably different from the final syllable “/a/ / C_”, thus causing less F0 variation.

Second, OQ distinguishes sex. Male speakers, American or Chinese, perform similar OQ curves at either narrow focus or not. Male OQ has an obvious rising tilt rate. Female speakers, on the other hand, perform typical OQ difference on narrow focus difference. Their OQ is lower at narrow focus position and higher at non-narrow focus position.

Chinese speakers, no matter sex, have a little bigger OQ variation range than Americans.

All male speakers do not vary much at OQ values. All females vary, though. Females’ OQ are lower at narrow focus position. But the variation is not much, only slightly bigger than male speakers. Meanwhile, although females’ OQ differ at narrow focus position, they do not take a rapid growth as males do. We observed only one exception – the female American speakers, when reading Chinese short sentences, suddenly rise from the one fourth point along the time axis. This might be the result of lab effect. We suppose the subject may psychologically hint herself of reaching a supposed expectation of experiment purposes.

The male Chinese and female American speakers rise more rapid in Chinese short sentences than long sentences. Since the short sentence only contains five syllables. Subglottal air pressure is still very strong at the sentence final position. There are still plenty of energy causing a rapid growth of OQ.

Third, SQ is stable on most occasions, differing only slightly for gender difference. Male American speakers’ SQ are mostly stable but rise rapidly at the rear end. The rest three
groups are fairly stable. Male Chinese speakers and female American speakers have a slight decline at the rear end of SQ. They fluctuate a little and don’t have rapid changes. However, this could not prove that male Chinese speakers are more close to female American speakers. In Chinese long sentences, all females decline at the rear end. But the male Chinese speakers SQ fluctuate and have a rising tilt rate at the rear end. In this sense, we believe all male speakers’ SQ rise fairly obviously at the rear end but Chinese’s rise is moderately lower in degree.

Above we have mentioned changes. Now let’s move to have a closer observation of the exact values. Males’ SQ is between 350-400 while females’ is about 250. This difference reveals that males have more low frequency energy than females.

Two male groups have higher SQ value in long Chinese sentences than short ones. This proves that at long sentence final position when subglottal air pressure declines, energy decline rapidly, therefore male compensate with a higher vocal folds vibration frequency. This is a male speaker strategy different from female speakers. However, we will not exclude another possible cause in this process. We’d like to report a byproduct in our experiment. We have observed a double peak phenomenon at the rear end of “-/a/ / C_” syllable. Since vocal folds vibrate quasi-periodically, frequency should also be fairly stable. Therefore double peaks within one period might be due to the ventricular folds vibration. Ventricular folds are above the vocal folds. When speech proceeds to the ending of speech, the subglottal air pressure declines to a minimum and the glottis begins to fall. The vocal tract withdraws downward. This will cause vocal folds to close first and pulls the ventricular to vibrate for a couple of times. Sometimes on other occasions ventricular vibration is intentionally used as a specific artistic phonation pattern as in Noh, a traditional Japanese opera.

Another surprising finding is a compensatory effect of OQ, SQ to F0. In other words, if F0 at narrow focus is higher than at non-narrow focus position, then a lower OQ, or a lower SQ, or a combination of both OQ and SQ, is witnessed across almost all conditions in our experiment. The only one exception is male American speakers. Even for them, the proposition is only moderately violated. They perform similarly at OQ and SQ. Hence according to this proposition we infer a similar F0 performance. And we do witness a smaller F0 variation range than all other groups.

For the rest three groups, compensatory effect are obvious. Male Chinese speakers have a stable OQ, so their F0 and OQ is negatively correlated, i.e., compensatory distributed. All female groups have lower OQ and higher F0 at narrow focus position. Since OQ varies not in a symmetrical degree as F0, their SQ adds weights on the scale for the smaller OQ variation. SQ and OQ together compensate F0 variation.

Female Chinese and American speakers’ F0 difference (F0 average peak minus low) is very close in the two positions:

\[
\begin{align*}
\text{f1_ch_if}_F0\text{ave} – \text{f1_ch_i}_F0\text{ave} &= 73.2526 \\
\text{f2_ch_if}_F0\text{ave} – \text{f2_ch_i}_F0\text{ave} &= 66.9312 \\
\text{f1_ch_sf}_F0\text{ave} – \text{f1_ch_s}_F0\text{ave} &= 47.9098 \\
\text{f2_ch_sf}_F0\text{ave} – \text{f2_ch_s}_F0\text{ave} &= 51.7211
\end{align*}
\]

Male and female Chinese speakers F0 difference in Chinese syllables are stable.

American speakers judge by pitch height, not only pitch height will decrease, but also in uttering Chinese. First they will emphasize on the first high-level tone and pinpoint the Chinese syllable pitch height at a stable level without much tilt rate. Then they come to the second step – to decide F0 height. But they are not sure. So the usual strategy is to narrow F0 difference between narrow and non-narrow focus. Other possible strategies are that they’ll first fix the narrow focus F0 height, then try to extend to a lower F0 height for non-narrow focus syllable. But this time again they are not sure. Or to fix a low F0 height, then extend to a higher F0 for narrow focus correspondent. They are not so good at distinguishing L2 narrow focus because they are not sure of pitch height, despite the fact that all examples in our case is level tone and excluded the tone effect. That is, even for level tones, L2 learners are not sure of pitch height in our case. Their strategy is to resort to a safer smaller range to erase the Chinese specific F0 range. They cannot judge F0 height.

Our research result partly conforms to Yin (2011). He studied Chinese L1 speakers’ performance for a nine syllable length material, with the same CV structure. He reported female F0 range of 80Hz, OQ rising about 7% and SQ decreasing tremendously. Our result conforms to his F0 part and contradicts to OQ and SQ data. Yin also reported males’ performance not that obvious as females. F0 raise but range is smaller than females’. It’s only 25Hz. Males’ OQ raise slightly and less than females. We haven’t observed such. This might owe to the different sentence length. We observed male speakers’ SQ raising slightly which conforms to his study.
4. Acknowledgment

Sincere thanks go to Prof. Kong Jiangping, Phonetics Lab, PKU, who have provided experimental surroundings, and Mr. Ye Zehua, who have developed Matlab programming for this study platform. Without these, this study is impossible.

5. References


