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

International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

GLOTTALIZATION IN INVENTORY CONSTRUCTION: A CROSS-LANGUAGE STUDY

H. Ding, O. Jokisch and R. Hoffmann

Laboratory of Acoustics and Speech Communication
Dresden University of Technology
{hongwei.ding; oliver.jokisch; ruediger.hoffmann}@ias.et.tu-dresden.de

ABSTRACT

In this paper we present the study of glottalization in three languages: Mandarin Chinese, English and German. The motivation of this study originally comes from the selection of inventory speaker for Mandarin Chinese, and the design of phoneset for English and German synthesis. Because glottalization is characterized with a discontinuity in fundamental frequency, which can degrade the quality of synthesis during pitch manipulation of a concatenative synthesis, we thus investigated the common phenomenon of glottalization in these languages, with the focus on Mandarin Chinese. We illustrated the contexts where it often occurs and propose the implication to deal with glottalization in inventory building for concatenative speech synthesis.

1. INSTRUCTION

1.1. What is Glottalization

Glottalization is often referred to as "creaky voice", "glottal stops", or "laryngealization", which is reported in many languages. We can find the phonetic description in Ladefoged [1] "Creaky voice is the term we will use for a mode of vibration of vocal folds in which the arytenoids cartilages are much closer together than in modal voice." The glottalization is characterized by a significant drop in the amplitude in the waveform, and more importantly, there is an abrupt change in the periodicity of the signal. We have observed this phenomenon both in tonal languages such as Mandarin Chinese and non-tonal languages such as English and German.

1.2. The Implication in Inventory Building

If the inventory units for synthesis contain irregular pitch period, any prosodic manipulation in concatenative synthesis can lead to a degradation of the synthesized speech in general [2]. Because of its discontinuity of fundamental frequency, glottalization can bring some trouble for a concatenative speech synthesis at the stage of pitch manipulation. To investigate the significance of glottalization, and how to deal with it in the construction of inventory is vital to the quality of synthesis.

2. THE STUDY OF GLOTTALIZATION

We will illustrate the phenomenon in these three languages separately. We conducted experiments to find out, whether there is glottalization in the particular language, in which condition it occurs, and propose our approach to deal with it in the design of phoneset and selection of speakers.

2.1. Mandarin Chinese

2.1.1. Investigation of Glottalization

The effect of glottalization in Mandarin Chinese has already been described early in 1956 by the great linguistician Chao, Yuen-Ren [3]: "A peculiar effect that often results from the effect to lower the pitch is the loss of voice at the lowest point, resulting in a sort of a grunt or sometimes a glottal stop." In our study in Chinese [4], we come to the conclusion that glottalization occurs more often in the middle of an isolated tone 3 and sometimes at the end of a tone 4.

2.1.2. Material

Isolated syllables and natural texts were designed in order to find the answers to the questions and to see whether the rate of speaking and the coarticulation effects have any influence on glottalization.

We took efforts to select such syllables that are distinguished by the prosodic features of the tone as well as by distinctions of vowel quality. 400 isolated syllables with an equal distribution of each tone for 100 syllables were chosen. The isolated syllables were written in Chinese characters, prepared in a randomized version of the list. The aim was to find whether the special articulation of the particular sound has different degree or different patterns of glottalization. The texts were simple articles, where third tone popped up frequently under various circumstances (about 140 syllables with third tone).

2.1.3. Speakers

Eight speakers, among which four male speakers and four female speakers, from different origins were selected. They were all Chinese students or researchers in Germany. They were asked to read these prompts with Mandarin Chinese. The recording was carried out in a quiet room with a DAT-recorder.

2.1.4. Results

Material of the analysis included all the isolated syllables from the eight speakers, and third tone syllables in texts of two speakers, one with least glottalization, the other with the most glottalization, both of them come from Beijing. The statistics were made on 3480 syllables (3200 isolated syllables and 280 from texts).

It is clear from the results: the irregular period of pitch, or glottalization seldom appears in tone 1, tone 2 or tone 4, with the following exceptions:

1. Such phenomena can only be observed at the end (offset) of few second tones and a lot of fourth tones for some speakers. This has also been observed by Chao [5] that the trail end of the 4th Tone, for instances, trails to near grunt, since it touches or seems to try to go beyond the lowest limit of one's voice. This can be regarded as trivial for synthesis, because this part can be omitted by segmentation for inventory, or it can be absorbed in continuous speech.

2. It is also found in the beginning of the syllables with vowel initials, which was also reported in German and English [6]. This can also be regarded as normal, which can also be handled for synthesis by deleting this part in segmentation.

The analysis was thus focused on the middle part of the third tone, when the pitch dips down. According to the results, apart from the regular pitch period excitations, there are two kinds of irregular pitch variations:

1) Glottalization

The presence of irregular pitch periods is easily seen in the speech wave (Fig.1). Regular excitation can be observed only at the beginning and at the end of the signal. The disturbance of F0, which is characteristic of glottalization, lasts as long as 20 ... 40 ms for both female and male speakers, F0 drops as low as 20 ... 40 Hz for female speakers as well as for male speakers.

2) Voiceless Excitation

The glottalized voicing can sometimes go to extreme as to the absence of voicing or silence. Fig.2 shows a voiceless part in the middle of a voiced syllable.

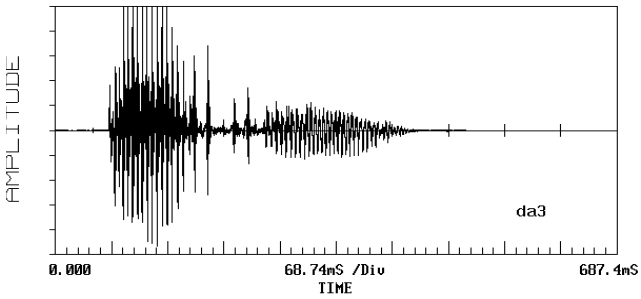

Fig. 1. Waveform of syllable "da3" with glottalization.

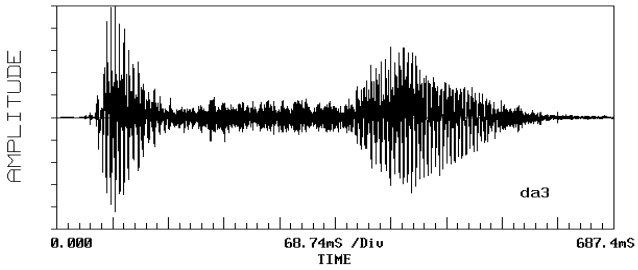

Fig. 2. Waveform of syllable "da3" with a voiceless part in the middle.

There was sometimes a mixture of these three categories. There were many regular excitations with very short period of glottalized voice for several ms, which can be regarded as normal speech. But there were also many glottalizations together with a long phase of voiceless part, which could be clearly perceived as an aspiration in this period. This kind was not regarded as an ordinary pronunciation. Many such cases can be found in the speech from Speaker 4.

2.1.5. Interpretation of Results

It seems that the local accent or sex difference are not the main reasons to give rise to glottalization. From the experience in labeling the phonemes, we also found that some vowels (particularly the low vowels) tend to be more glottalized than other vowels (such as the high vowels). This result also coincides with that in [7].:

1. *The appearance of the three different variants reveals that they are strongly speaker-dependent.*

   The kinds of excitation and the correspondent frequency of the investigated third tones are illustrated in Fig.3 by statistics. The difference of the behavior among these speakers can clearly be observed.

2. *The tendency of glottalization will essentially be reduced by shortened duration and tone coarticulation in the text.*

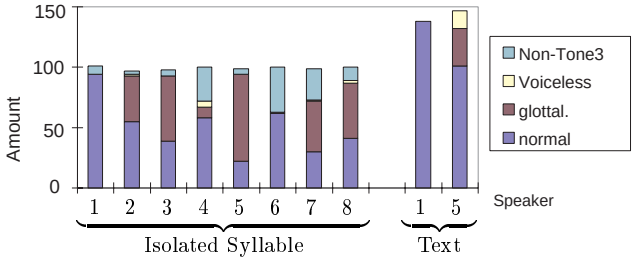

Fig. 3. The frequency of several variants of tone 3 in isolation and in text among different speakers (Non-Tone3 are mispronounced 3rd tones)

- In Fig.3, it is conspicuous that speaker 5 produced much less glottalized tone 3 in text than in isolated syllables. In the texts, the average syllable duration is about 300ms, while in isolated syllables about 600ms. In isolated syllables, the unnatural long time allows one to realize the irregular pronunciation, while in natural text, one does not have enough time to fulfill such changes. The glottalized part will be reduced in a shortened duration.
- In text, more tone 3 appear before another tone than before a pause, which are the environments for allophonic tone 3: A third tone changes either to a second tone (before another third tone) or it just goes down and does not rise again (before tone 1, tone 2, and tone 4). This means that the coarticulation effects in text also reduce the frequency of glottalized third tone.

2.1.6. Implications

1. Glottalization in the middle part of the syllable can pose a great difficulty to the prosodic manipulation. In order to avoid this problem, a speaker without or with little glottalization should be found.
2. In order to reduce the phenomenon with applicable methods, units for the inventory should be taken from fluent read sentences rather than from isolated monosyllabic words.

These conclusions were taken into consideration in the course of our inventory database preparation for Chinese.

2.2. English

Glottalization has been described as ”glottal stops” in English [1]. Some experts even made efforts to synthesize such effect in the synthesis of English [6]. Glottal stops typically occur during or immediately following the transition between adjacent vowels, as shown in Fig.4.

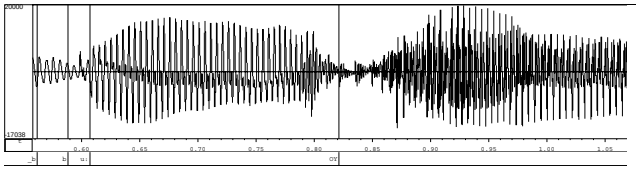

Fig. 4. Waveform of /uOI/ in word “bamboo-oil” with a glottalization between two vowels.

Because the speaker wanted to mark the differentiation of adjacent vowels when they occur between words, she pinched the vocal cords, so that the airflow was reduced, vocalization was diminished, and thus effected glottal stops.

In analyzing the speech database from 6 US-English speakers, we found it impossible to generalize a particular pattern for their occurrences. Actually, glottal stops can be found anywhere in an utterance. The frequency of the occurrence varies from speaker to speaker. We investigated the speech database of these 6 speakers, strong glottalization between vowels could only be found in the speech of two speakers. The other speakers did not demonstrate any glottalization in the vowel-vowel transition, or in other contexts. The only glottalization demonstrated by almost every speaker is before a word-initial vowel as shown in Fig.5, and sometimes at the word-final.

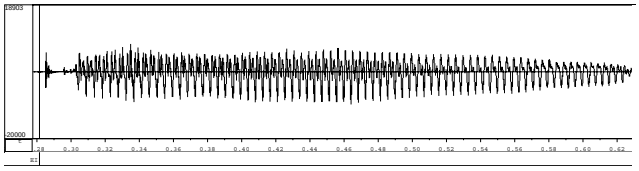

Fig. 5. Waveform of /eI/ in “aas” with glottalization before the word initial vowel.

By deleting the first part of glottalization at the word-initial, or embed the target diphones in the middle of the words, we can avoid the problem of glottalization in synthesis. Thus we tended to regard glottal stops as a nonphonemic acoustic phenomenon in American English, they were not considered in the design of phoneset. A speaker with less glottalization was chosen.

By the careful selection of inventory speaker and intelligent treatment of carrier words, the glottalization can be neglected.

2.3. German

Glottal stops have also been described in German [8]. In the database of our German speakers, who have least glottalization, we still find many occurrences of such discontinuity not only at the beginning of word-initial vowels, but also at the syllable-initial vowels. For example, in word ”verübeln (resent)” in Fig.6, the glottal stop marked as /Q/, shows a

significant drop in the amplitude before the beginning of vowel.

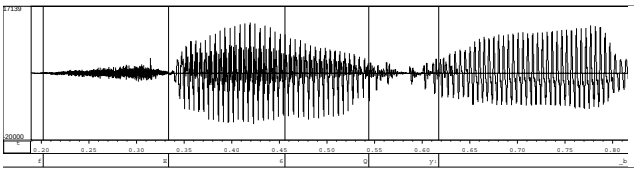

Fig. 6. Waveform of /fE6Qy:/ in "verübeln" with glottalization at the syllable initial vowel /y:/.

Because the glottis occurs much more frequent in German than in English, and it occurs not only at the word-initial or word-final, it also appears across syllables in the word. It is not possible to neglect their appearance in the diphone units. The best way to deal with it is, to label the glottalization in the inventory, to take it into consideration. In this way, we separate the discontinuous part in the course of prosody manipulation of the vowels and consonants. Therefore, a smooth concatenation of the diphones and a correct manipulation of the pitches can be resulted.

3. CONCLUSION

By the investigation of the phenomenon of glottalization in the database of Mandarin, American English and German, we come up with the result, that the glottalization is quite unpredictable and idiosyncratic for particular utterances and to individual speakers. But some characteristics can be summed up:

Characters, which are common in these languages are

- They are speaker dependent in these languages.
- It occurs frequently at the word-initial vowels, and sometimes at word-final vowels.
- It can disappear in continuous speech.

The special characters in each language:

- They are often found in the middle of tone 3 of isolated syllables, and sometimes at the end of tone 4 in Mandarin.
- They occur more frequently at the syllable-initials vowel in German than in English. Perhaps the particular syllable structure of English is usually CVC, which does not provide so many chances as in German for the occurrences of glottalization.

Different treatment of glottalization in each language

1. By selecting an inventory speaker with less glottalization, and deleting the glottalization part at word initials or finals, it is unnecessary for us take the glottalization as a phone in the phoneset in Mandarin and English. We can thus reduce the number of syllable units in Chinese or diphone units in English significantly.
2. But in German synthesis, glottalization is regarded as a separated phone in combining diphones, because the discontinuous part can hardly be eliminated from the phone-phone transition. By indicating its appearance, we can treat them differently in the prosodic manipulation, the quality of concatenative synthesis will not be degraded.

We have proposed our treatment of glottalization in different languages, so as to construct an inventory with least number of acoustic units, and at the same time to produce a natural speech. But if you want to synthesize the variety of natural speech, glottalization can be modeled, because it exists in the natural languages.

4. REFERENCES

[1] LADEFOGED, P. & MADDIESON, I. (1996): The Sounds of the World's Language. Blackwell Publishers

[2] BUNNELL, H. T. & YARRINGTON, D. & BARNER, K. E. (1994): Pitch Control in Diphone Synthesis. Second ESCA/IEEE Workshop on Speech Synthesis, p. 127-130

[3] CHAO, YUEN-REN (1956): Tone, Intonation, Singsong, Chanting, Recitative, Tonal Composition and Atonal Composition in Chinese. For R. Jackobson, p. 52-59. The Hague: Mouton

[4] DING, H. & HELBIG, J. (1996): Sprecher- und kontextbedingte Varianz des dritten Vokaltones in chinesischen Silben - Eine akustische Untersuchung. DAGA1996, Bonn, p. 514-515

[5] CHAO, YUEN-REN (1968): A Grammar of Spoken Chinese. University of California Press, Berkely

[6] PIERREHUMBERT, J.B. & FRISCH, S. (1997): Synthesizing Allophonic Glottalization. In van Santen, J.P.H. et al.(editors): Progress in Speech Synthesis. p. 9-26. Springer-Verlag, New York, Berlin.

[7] BELOTEL-GRENIE, A. & GRENIE, M.(2004): The Creaky Voice Phonation And The Organisation of Chinese Discourse. International Symposium on Tonal Aspects of Langauges 2004, Beijing, p.5-8

[8] KOHLER, K.J. (1994): Glottal Stops and Glottalization in German. Phonetica, Vol. 51, p.38-51