Voicing Distinction of Obstruents in the Hangzhou Wu Chinese Dialect

Yang Yue[1], *Fang Hu*[2,1]

[1]Department of Linguistics, Graduate School, University of Chinese Academy of Social Sciences
[2]Institute of Linguistics, Chinese Academy of Social Sciences
kk24325535@qq.com, hufang@cass.org.cn

Abstract

This paper gives an acoustic phonetic description of the obstruents in the Hangzhou Wu Chinese dialect. Based on the data from 8 speakers (4 male and 4 female), obstruents were examined in terms of VOT, silent closure duration, segment duration, and spectral properties such as H1-H2, H1-F1 and H1-F3. Results suggest that VOT cannot differentiate the voiced obstruents from their voiceless counterparts, but the silent closure duration can. There is no voiced aspiration. And breathiness was detected on the vowel following the voiced category of obstruents. An acoustic consequence is that there is no segment for the voiced glottal fricative [ɦ], since it was realized as the breathiness on the following vowel. But interestingly, it is observed that syllables with [ɦ] are longer than their onset-less counterparts.

Index Terms: obstruent production, closure duration, breathy vowel, Hangzhou Wu Chinese dialect

1. Introduction

The tripartite distinction of stops and affricates into voiceless unaspirated, voiceless aspirated and voiced is a well-known characteristic that defines the Wu Chinese dialect family ([1], [2], [3]). Historically speaking, the Wu dialect family preserves the pattern of voicing distinction in Middle Chinese. However, there is a long debate on the phonetics and phonology of voicing in Wu dialects, since the obstruents are usually not truly voiced, especially when they are not in an intervocalic position. Citing Liu's acoustic experiments in 1920s, Chao proposed that the voiced obstruents in Wu dialect are actually voiceless, but with a voiced aspiration ([1]). For instance, the voiced stops /b d g/ should be transcribed as [pʱ tʱ kʱ] respectively.

More recently, phonetic studies revealed that the voiced obstruents were indeed voiceless in a number of Wu dialects, but there was no voiced aspiration as Chao had proposed. The perceived voicedness can be attributed to the low tone and breathy voice on the subsequent vowels, especially at the beginning part ([4], [5], [6], [7], [8]). It seems that the voiced category of obstruents is undergoing devoicing in Wu dialects. On the other hand, the devoicing process varies across different phonetic environments such as citation form, stressed position in running speech, and intervocalic position in disyllabic words ([9]). And in general, more truly voiced productions are phonetically realized for voiced obstruents in the intervocalic position than in the stressed position and citation form.

However, previous studies usually focused on stops and little attention has been paid to affricates and fricatives. Like stops, affricates in Wu dialects have a three-way distinction into voiceless unaspirated, voiceless aspirated, and voiced. And fricatives have a voiceless versus voiced distinction in Wu. The phonetics of voicedness in affricates and fricatives is supposed to be the same as the voiced stops in Wu dialects. However, the aerodynamic configuration for the production of voiced fricatives is very different from that for voiced stops. The difficulty for the production of voiced stop lies in the maintenance of a subglottal pressure under the condition of a complete blockage of airflow in the oral cavity. The production of voiced fricative involves a conflicting aerodynamic specification for the production of frication and of voicing, since the former requires an airflow strong enough to form an audible turbulence, while the latter requires a modulated airflow to facilitate the vibration of the vocal folds. [10] reported that a number of factors may affect the production of voiced fricative in Wu, and there are large intra- and inter-speaker variations. A general observation is that more voicing is detected in the production of non-sibilant voiced fricatives such as [v] and [ɦ] than in the production of sibilant voiced fricatives such as [z] and [ʑ]. In other words, strong voiceless frication dominates the production of sibilant fricatives, and voicing is only partially maintained in the production of weak non-sibilant fricatives. And this could also explain why voiced fricatives are typologically more common than voiced stops in Chinese dialects in general. For instance, [v] and [ʐ] are commonly detected in Mandarin dialects.

A related issue concerns the voiced glottal fricative [ɦ], as there is a distinction between [h] and [ɦ] in Wu dialects in general. [11] reported that /ɦɛdɛ/ in the Wenling Wu dialect sounds the same when it is played back reversely, and the implication is that [ɦ] is accompanying throughout the vowel [ɛ]. But how? According to previous phonetic studies, there is probably no visible segment for [ɦ], and [ɦ] could be manifested by the breathiness on the following vowel. If so, a consequent issue is whether the breathiness is a feature for the consonant, the vowel, or even tone or syllable?

Table 1: *Inventory of initial consonants in Hangzhou.*

	stop/affr.	nsl.	fri.	appr.
labial	p pʰ b	m		
labio-dental			f v	
dental/alveolar	ts tsʰ dz t tʰ d	n	s z	l
alveolo-palatal	tɕ tɕʰ dʑ		ɕ	
velar	k kʰ g	ŋ		
glottal			h ɦ	

This paper investigates the voicing distinction among the obstruents in the Hangzhou Wu dialect. As shown in Table 1, the stops in Hangzhou have three places of articulation, labial, dental/alveolar, and velar, and a three-way distinction of voicing into voiceless unaspirated, voiceless aspirated, and voiced; the affricates have two places of articulation, dental/alveolar and alveolo-palatal, and the same three-way distinction of voicing as the stops; the fricatives have four places of articulation, labio-dental, dental/alveolar, alveolo-palatal, and glottal, and a voiceless versus voiced distinction of voicing. Note that the voiceless alveolo-palatal fricative [ɕ] does not have a voiced counterpart [ʑ]. That is because [ʑ] has merged into [ɦ] diachronically.

2. Methodology

8 native speakers, 4 male and 4 female, provided speech data. All of them were born and raised up in Hangzhou, and had no reported history of speech disorders. Meaningful monosyllabic words were used as test words. Each target initial consonant has three test words, preferably with the vowel /i/, /u/ or /a/. Test words, preferably with a mid level tone, were placed in citation form, in carrier sentences, and in intervocalic position in disyllabic words. Audio sounds were recorded directly into a laptop PC with the DMX 6 Fire USB sound card through a SHURE SM86 microphone. Speakers were instructed to speak the test words in a natural way with a normal tempo. The sample rate is 22,050 Hz. Five repetitions were recorded.

Speech data were annotated and analyzed by PRAAT 6.0.19 ([12]) and VoiceSauce ([13]). Acoustic measures include durations and spectral properties. First, voice onset times (VOT) were measured to characterize voicing and aspiration in stops and affricates ([14]). Second, closure durations were measured to characterize the formation of stops and affricates, which is reported to be related to the voiceless versus voiced distinction in the literature ([15]). Third, ratios of the voiceless part of the voiced fricatives to the syllable duration were measured to quantify the degree of devoicing of voiced fricatives. Finally, in order to acoustically quantify breathiness, differences in amplitude between the first and second harmonics (H1-H2), between the first harmonic and the strongest harmonic in the first formants (H1-F1), and between the first harmonic and the strongest harmonic in the third formants (H1-F3) are measured in the 11 equidistantly distributed points on the vowel ([16]).

3. Results

3.1 VOT for stops and affricates

The 9 initial stops in Hangzhou /p p^h b/, /t t^h d/, and /k k^h g/ contrast in place of articulation and voicing. First, it is generally observed that target stops and affricates in intervocalic positions in disyllabic words retain the three-way distinction of voicing. Second, closure duration can only be measured from target segments in carrier sentence, not in citation forms. And there is no significant difference in VOT values between in citation forms and in carrier sentences. Therefore, the following discussion mainly focuses on the data in carrier sentences.

Figure 1 shows mean VOT values in seconds, and Table 2 shows the result of a two-way ANOVA. As mentioned above, place of articulation and voicing are the two factors. And data were averaged from all the 8 speakers, as there is no significant difference between male and female speakers. That is, it is a 3 × 3 design. Table 2 shows significant effects of place of articulation, voicing and the interaction. First, post hoc Tukey HSD confirmed significant difference between labials and alveolars /p b/ < /t d/ ($p < 0.0001$), and between labials and velars /p b/ < /k g/ ($p < 0.0001$). Second, there is significant difference between voiceless unaspirated and voiceless aspirated stops /p t k/ < /p^h t^h k^h/ ($p < 0.0001$), and between voiced and voiceless aspirated stops /b d g/ < /p^h t^h k^h/ ($p < 0.0001$). Third, there is no significant difference between alveolars and velars ($p = 0.1898$), and between voiceless unaspirated and voiced stops ($p = 0.0445$). In summary, VOT differentiates aspirated stops from the other categories, but fails to differentiate voiceless unaspirated stops from voiced stops.

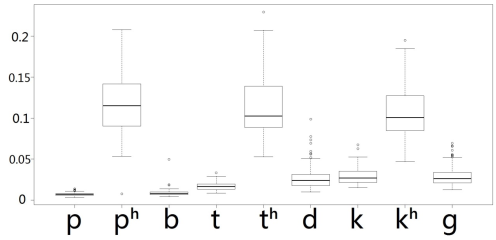

Figure 1: *VOT in seconds of Hangzhou stops.*

Table 2: *Two-way ANOVA for VOT of Hangzhou initial stops in carrier sentences.*

Factor	Mean Sq	F value	Pr (>F)
Place	0.01818	7.420	0.00656 **
Voicing	0.003108	1.268	0.26031
Place: Voicing	0.000145	0.059	0.80792

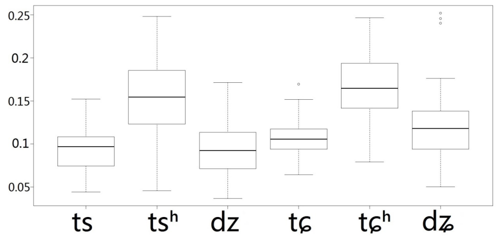

Figure 2: *VOT in seconds of Hangzhou affricates.*

Table 3: *Two-way ANOVA for VOT of Hangzhou initial affricates in carrier sentences*

Factor	Mean Sq	F value	Pr (>F)
Place	0.04755	47.170	< 0.0001 ***
Voicing	0.25976	257.701	< 0.0001 ***
Place: Voicing	0.00237	2.352	0.0959

VOT values in seconds for the 6 Hangzhou affricates, /ts ts^h dz/ and /tɕ $tɕ^h$ dʑ/ in carrier sentences were summarized in Figure 2. And Table 3 shows results of a two-way ANOVA. There are significant effects of place of articulation (/ts ts^h dz/ < /tɕ $tɕ^h$ dʑ/ respectively) and voicing, but no significant effect of interaction. Again, post hoc Tukey HSD confirmed that the differences between voiceless aspirated affricates and the other two categories are significant /ts^h $tɕ^h$/ > /ts dz tɕ dʑ/ ($p < 0.0001$). But there is no significant difference between the voiceless unaspirated and voiced affricates ($p = 0.1502$).

3.2 Closure durations for stops and affricates

It has been shown so far that voiceless and voiced stops and affricates could not be differentiated by VOT. That is, the difference in voicing is not manifested after stop release. Yet closure duration can serve as an indicator to differentiate the voicing distinction of stops and affricates, as closure duration reflects articulatory efforts in stop and affricate production before release.

Figures 3 and 4 show means in seconds of closure duration for the stops /p p^h b t t^h d k k^h g/ and affricates /ts ts^h dz tɕ $tɕ^h$ dʑ/ in carrier sentences in Hangzhou, respectively. And Tables 4 and 5 summarized results of two-way ANOVAs for the stops and affricates respectively.

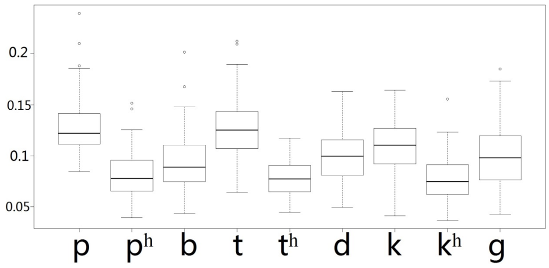

Figure 3: *Closure durations in seconds of Hangzhou stops.*

Table 4: *Two-way ANOVA for closure durations of Hangzhou initial stops.*

Factor	Mean Sq	F value	Pr (>F)
Place	0.00408	6.723	0.00125 **
Voicing	0.16033	264.443	<0.0001 ***
Place: voicing	0.00584	9.629	<0.0001 ***

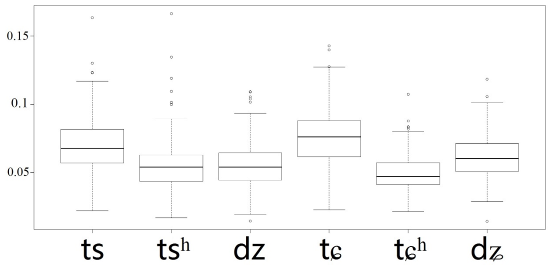

Figure 4: *Closure durations in seconds of Hangzhou affricates.*

Table 5: *Two-way ANOVA for closure durations of Hangzhou initial affricates.*

Factor	Mean Sq	F value	Pr (>F)
Place	0.001275	3.086	0.079411
Voicing	0.028584	69.170	< 0.0001 ***
Place: Voicing	0.003036	7.348	< 0.001 ***

Table 4 shows significant effects of place of articulation, voicing, and the interaction for stops; Table 5 shows no significant effect of place of articulation, but significant effects of voicing and the interaction for affricates. Post hoc Tukey HSD confirmed that (1) there are significant differences in closure duration between velar stops and the other two categories ($p < 0.01$), but no difference between labial and alveolar stops ($p = 0.9016$): /k k^h g/ > /p p^h b/ = /t t^h d/; (2) there are significant differences in closure duration between voiceless unaspirated, voiceless aspirated, and voiced stops and affricates ($p < 0.01$ for the pair of voiceless aspirated and voiced affricates, and $p < 0.0001$ for the other pairs): /p t k/ > /b d g/ > /p^h t^h k^h/ and /ts tɕ/ > /dz dʑ/ > /ts^h $tɕ^h$/.

The main point is that closure duration differentiated voicing distinctions among stops and affricates in Hangzhou. The results confirmed that voiceless unaspirated stops and affricates require significantly longer closure duration than their voiced counterparts to build up a greater articulatory effort. However, this kind of effort may not have an acoustic consequence, as it is not manifested in VOT.

3.3 Fricatives

There are three pairs of voiceless versus voiced distinction in fricatives in Hangzhou: /f v/, /s z/, and /h ɦ/. /ɕ/ does not have a voiced counterpart, so it is not included in the discussion.

As the aerodynamic requirements for the production of voicing and fricative is conflicting, devoicing of the voiced fricatives is common in Hangzhou. Our data show that the devoicing is gradient. That is, the voiced fricatives do retain voicing to a substantial degree. In order to quantify the degree of devoicing, occurrence frequencies of the fully voiced, partially voiceless, and fully voiceless were calculated for the voiced fricatives [v z]. The results were summarized in Table 6.

Table 6: *Realization of Hangzhou voiced fricatives.*

voice	v		z	
	occur.	%	occur.	%
full	102	85	63	52.5
Partial	10	8.3	12	10
zero	8	6.7	45	37.5

Table 7: *Realization of voiced fricatives in percentages in speakers (n=30 for each speaker).*

speaker	full	Partial	Zero
F1	73.4	23.3	3.3
F2	56.7	6.6	36.7
F3	80	0	20
F4	87.8	3.3	9.9
M1	60	16.7	23.3
M2	93.4	0	6.6
M3	26.7	16.7	56.6
M4	73.4	6.6	20

Table 6 demonstrates a clear difference in devoicing between /v/ and /z/. /v/ preserved 85% of fully voicing among 102 tokens in total, whereas /z/ retained 52.5% of fully voicing among 63 tokens in total. The results corroborate the report in previous studies ([10]) that weak fricative /v/ maintains more voicing than strong fricative /z/. However, results also suggest that there are inter-speaker variations, as shown in Table 7. In general, most speakers have a high percentage of fully voicing for the voiced fricatives in Hangzhou, but one speaker (M3) has more fully voiceless (56%) than fully voiced (26.7%) productions.

The data suggest that /ɦ/ could not be segmented from the following vowel in the acoustic signal. The question is: does /ɦ/ exist? Is there still an articulation for /ɦ/? In addition to possible breathiness on the following vowel that will be discussed in the next section, will a syllable with /ɦ/ have a longer duration than its corresponding syllable without it? As mentioned earlier, Hangzhou has /h/ versus /ɦ/ distinction. There is actually a three-way distinction of voicing if the onset-less syllables are also taken into the account. And it is common to Wu phonology in general that zero initial, /h/, and /ɦ/ form a three-way distinction of voicing ([1]).

Table 8: *One-way ANOVAs for durations in seconds of /i/ and /ɦi/ in carrier sentences and intervocalic positions in disyllabic words.*

carrier sent.	Mean	SD	P-value
/i/	0.22	0.002	<0.0001***
/ɦi/	0.28	0.006	
intervocalic	Mean	SD	P-value
/i/	0.28	0.006	0.1
/ɦi/	0.31	0.007	

Therefore, the minimal pair /i/ "clothes" versus /ɦi/ "to move" was chosen in the sampled database to explore whether there is a durational difference between /i/ and /ɦi/. Each has 40 data points (5 repetitions × 8 speakers). Table 8 summarized mean durations in seconds and standard deviations (SD) and P-values from one-way ANOVAs for /i/ and /ɦi/ in carrier sentences and intervocalic positions in disyllabic words. It can be seen from the table that /ɦi/ is 60 ms longer than /i/ in carrier sentences and /ɦi/ is 30 ms longer than /i/ in intervocalic positions in disyllabic words. And the difference is significant in carrier sentences, but is not in intervocalic positions in disyllabic words. In summary, /ɦ/ tends to have an articulatory effect on the syllable production in terms of duration on the one hand, and this kind of durational difference tends to disappear in unstressed positions in running speech.

3.4 Spectral properties

The data have shown voiced obstruents may have different acoustic phonetic manifestations in Hangzhou. The voicing is generally retained in the voiced fricatives /v z/, but is not observed in the voiced stops in terms of VOT. And the voiced glottal fricative /ɦ/ could not be segmented acoustically. Do they all have a similar effect on the following vowel by causing breathiness? Figure 5 shows the means in dB of spectral measures (H1-H2, H1-F1 and H1-F3) on the 11 equidistantly distributed data points of the target CV syllables. Due to space limit, discussion is focused on the /p-b/, /ts-dz/, /s-z/, and /zero-ɦ/ pairs, which are typical and representative for stops, affricates, fricatives, and the onset-less versus glottal articulations, respectively. Table 9 summarized P-values of one-way ANOVAs for the /p-b/, /ts-dz/, /s-z/, and /zero-ɦ/ pairs on the onset (10%), mid (50%), and offset (90%) data points of the target syllables in carrier sentences. Asterisks denoted significant levels, and P-values were given if no significance.

Table 9: *One-way ANOVAs for spectral measures on the onset, mid, and offset of the syllables with /p-b/, /ts-dz/, /s-z/, and /0-ɦ/ pairs in carrier sentences.*

		H1-H2	H1-F1	H1-F3
/p-b/	onset	***	***	***
	mid	***	***	***
	offset	***	***	***
/ts-dz/	onset	*	**	***
	mid	*	0.06	**
	offset	0.96	0.84	0.18
/s-z/	onset	0.1	0.12	*
	mid	0.06	0.13	*
	offset	*	**	0.5
/0-ɦ/	onset	**	***	***
	mid	0.23	***	***
	offset	**	0.41	0.7

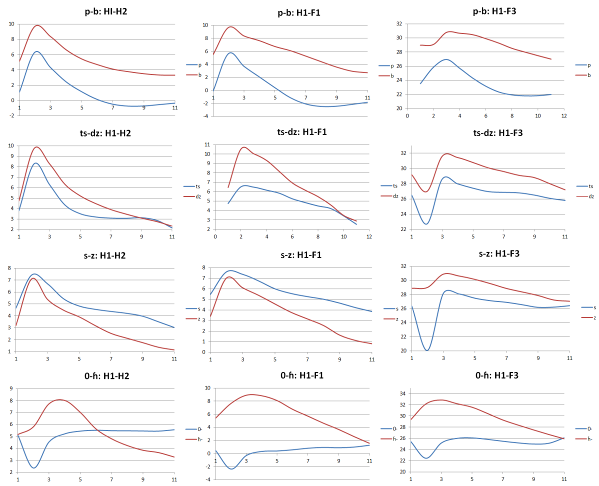

Figure 5: *Means in dB of spectral measures (n=120) for /p-b/, /ts-dz/, /s-z/, and /0-ɦ/ in carrier sentences.*

It can be seen from Figure 5 that the vowel following voiced obstruents generally has greater H1-H2, H1-F1 and F1-F3 values than the vowel following their voiceless or onset-less counterparts in Hangzhou. The results confirmed that the vowel following voiced obstruents is breathier than the vowel following their voiceless or onset-less counterparts. Meanwhile there are inter-category variations. As shown in Figure 5 and Table 9, the difference in breathiness is highly significant throughout the entire syllable for the syllables with stop pair /p-b/. The syllables with affricate pair /ts-dz/ exhibit significant difference in breathiness on the first half of syllable duration. The syllables with fricative pair /s-z/ seem to be exceptional: no significant difference in terms of H1-H2 and H1-F1 except for the offset, and weak significant differences on onset and mid points. This could probably be attributed to the fact that the voiced fricatives partially maintained voicing. Finally, syllables with the onset-less and /ɦ/ pair show highly significant difference in breathiness on the onset and mid points in terms of H1-F1 and H1-F3, but the significant difference is detected only on the onset in terms of H1-H2 (Note that the offset data point even showed an opposite significant difference).

4. Conclusions

This paper gives an acoustic phonetic description of the obstruent distinction in the Hangzhou Wu Chinese dialect. Results show that VOT cannot differentiate the voiced obstruents from their voiceless counterparts, but the silent closure duration can. And breathiness was detected on the vowel throughout the entire syllable following the voiced stops. In conclusion, Hangzhou generally maintained a three-way laryngeal distinction between voiceless unaspirated, voiceless aspirated, and voiced stops. Affricates also show a three-way laryngeal distinction, and the breathiness lasts for the first half of syllable duration in voiced affricates. And the onset-less initial, /h/, and /ɦ/ form a three-way laryngeal distinction, too.

The voiced stops /v z/ generally maintained voicing, though there are inter-speaker variations. And the weak fricative /v/ retained more voicing productions than the strong one /z/.

5. Acknowledgements

This work is supported by Chinese National Social Science Foundation (Project number: 15BYY073; P.I.: Hu Fang).

6. References

[1] Y. -R. Chao, *Studies in the Modern Wu-Dialects*. Beijing: Tsing Hua College Research Institute. 1928.

[2] Y. -R. Chao, "Types of Plosive in Chinese Dialects," *Bulletin of the Institute of History and Philology, Academia Sinica*, 4, pp. 515-520, 1935.

[3] Y. -R. Chao, "Contrastive Aspects of the Wu Dialects," *Language*, 43, pp. 92-101, 1967.

[4] J. Cao, "Articulatory characteristics of historical voiced obstruent initial consonants in the Changyinsha dialect," [in Chinese]. *Studies of the Chinese Language*, 167, pp. 273-287, 1982.

[5] J. Cao, "On the voiceless versus voiced distinction," [in Chinese]. *Studies of the Chinese Language*, 197, pp. 101-109, 1987.

[6] F. Shi, "Acoustic characteristics of voiced stops in Suzhou dialect," [in Chinese]. *Studies in Language and Linguistics*, 4, 49-83, 1983.

[7] N. Ren, *Phonation types and stop consonant distinctions—Shanghai*. Ph. D. dissertation, The University of Connecticut, 1992.

[8] J. Cao and I. Maddieson, "An exploration of phonation types in Wu dialects of Chinese," *Journal of Phonetics*, 20, pp. 77-92, 1992.

[9] Z. Shen, C. Wooters and S.-Y. Wang, "Closure Duration in the Classification of Stops: A statistical analysis," *OSU Working Papers in Linguistics*. 35, pp. 197-209, 1987.

[10] F. Ling, "On the voiced fricative initial consonants in the New variety of Shanghai Dialect," [in Chinese]. *Studies of the Chinese Language*, 381, 733-768, 2017.

[11] R. Li, "/ɦɛdɛ/ in Wenling Dialect sounds the same when it is play back reversely," [in Chinese]. *Studies of the Chinese Language*, 191, pp. 106, 1986.

[12] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," [Computer program]. Version 6.0.19, 2016.

[13] Y.-L. Shue, *The voice source in speech production: Data, analysis and models*. Ph. D. Dissertation, UCLA, 2010.

[14] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, 20, pp. 384-422, 1964.

[15] L. Lisker, "Closure duration and the intervocalic voiced-voiceless distinction in English," *Language*, 33, pp. 42-49, 1957.

[16] G. Fant, "Preliminaries to analysis of the human voice source". *Quarterly Progress Report*, 4, pp. 1-27, 1982.