Estimation of second subglottal resonance based on F2 measurements and its application in consonant-vowel classification in Hungarian[a)]

Tamás Gábor Csapó[b)]
Laboratory of Speech Technology, Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary

(Dated: June 21, 2009)

It has been shown for several languages that subglottal resonances (SGRs) play a dividing role in the frequency space of consonants and vowels (e.g. vowels are separated into the back-front categories by the second subglottal resonance). Consonant-vowel transitions are characterized by a regression line (locus equation), and can be classified into distinct categories in the locus equation space, according to their place of articulation. Several attempts have shown that the dividing lines between these categories may be the SGRs. In this paper, the relation between CV transitions in the locus equation space and the separating role of the subglottal resonances are further investigated. Locus equation space of one native speaker of Hungarian is examined. Consonant-vowel transitions are classified based on SGRs estimated from the locus equations of a subset of CV sequences. The hit rates and false alarm rates of the classification are comparable to a baseline experiment where the subglottal resonances were measured from accelerometer signal.

PACS numbers:
Keywords: subglottal resonance, locus equation, Hungarian

I. INTRODUCTION

It has been shown for several languages that subglottal resonances (SGRs) play a dividing role in the frequency space of consonants and vowels (e.g. American English: Lulich (2009) and Hungarian: Csapó *et al.* (2009)). It seems that speakers try to avoid putting formants in the regions of SGRs, consequently vowels are divided into distinct categories (e.g. back vs. front by $Sg2$ and low vs. non-low by $Sg1$, see Csapó *et al.* (2009)).

Consonant-vowel transitions are characterized by a regression line (locus equation), which shows the correlation between the second formant at the onset of voicing in the consonant ($F2_{burst}$) and that of at the steady state of the vowel ($F2_{vowel}$), as described in Lulich (2008). When illustrating $F2_{burst}$-$F2_{vowel}$ together in the locus equation space, groups of CV transitions are distinguishable according to their place of articulation. Chen and Lulich (2009) showed that the boundaries between these distinct groups are related to subglottal resonances.

In this experiment, the relation between CV transitions in the locus equation space and the separating role of the subglottal resonances is further investigated based on the acoustic data of a native speaker of Hungarian. After that, an $Sg2$ estimation algorithm (Chen and Lulich, 2009) is applied and further improved for use in Hungarian. Subglottal resonance frequencies of six other speakers are measured and examined, in order to find correlation between $Sg1$, $Sg2$ and $Sg3$. For the first speaker, estimated SGR values are used in the clustering of CV transitions. Hit rates and false alarm rates of the classification of consonants and vowels are investigated.

II. METHODS

A. Accelerometer recordings

Acoustic data were collected from seven native speakers of Hungarian (referred as TB and S1-S6) in an anechoic chamber. While the speakers uttered several sentences read from a paper, voice and accelerometer recordings were done. The speech utterances were recorded using an EMC 100 condenser microphone at a distance of approximately 15 cm from the lips. The subglottal data were recorded using a K&K HotSpot accelerometer attached to the skin of the neck below the thyroid cartilage. The two signals were digitized at 8 kHz with a Terratec DMX 6 Fire USB external sound card and were recorded to separate channels using Wavesurfer (Sjőlander and Beskow, 2009). These voice recordings were unrelated for this experiment.

B. Subglottal resonance measurements

The first three subglottal resonances were measured manually 25 times from the accelerometer signal of speaker TB, using Wavesurfer. A sample FFT taken from the accelerometer signal can be seen in Fig. 1, which shows that this measurement is similar to reading off formants. A detailed description about measuring SGRs can be found in Lulich (2009) and Chi and Sonderegger (2007). For speakers S1-S6, the subglottal resonances were measured 10 times from their accelerometer signal.

The mean, median and standard deviation values of

[a)] Final project for "Speech Acoustics" course
[b)] Electronic address: `csapot@tmit.bme.hu`; URL: `http://speechlab.tmit.bme.hu/csapo`

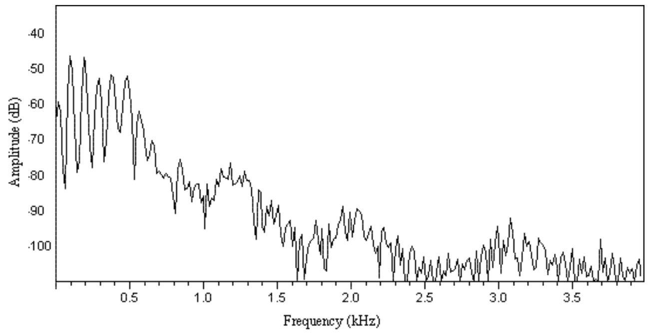


FIG. 1. Sample FFT from the accelerometer data of speaker TB. Spectral peaks (454 Hz, 1211 Hz, 2023 Hz and 3067 Hz) may correspond to subglottal resonances. However, $Sg1$ seems to be very low due to the coupling to the supraglottal tract.

subglottal resonances for speaker TB are shown in Table I. Table II shows the medians of SGRs, for speakers S1-S6. The measured resonance frequencies correspond to the values reported in the literature (Stevens, 1998).

C. Speech recordings

In a separate recording session, speaker TB uttered "ɔCVbɔ"-like nonsense words in an anechoic room. At the place of the first consonant, all 8 Hungarian stop consonants were included (labials: [b,p], alveolars: [d,t], velars: [g,k] and palatals: [ɟ,c]). All of the 14 Hungarian vowels were included in the second non-stressed syllable ([ɔ,aː,o,oː,u,uː,ɛ,eː,i,iː,ø,øː,y,yː]). The nonsense words were uttered in a randomized order, 10 times each, for a total of 1120 utterances. Voice data were recorded using an EMC 100 condenser microphone and digitized at 48kHz with a Terratec DMX 6 Fire USB external sound card using Wavesurfer.

TABLE I. *Mean, median and standard deviation of measured SGRs for speaker TB.*

	$Sg1$	$Sg2$	$Sg3$
Mean	545 Hz	1241 Hz	2027 Hz
Median	554 Hz	1244 Hz	2022 Hz
Std. dev.	60	42	145

TABLE II. *Median of measured SGRs for speakers S1-S6.*

	$Sg1$	$Sg2$	$Sg3$
S1	612 Hz	1498 Hz	2283 Hz
S2	572 Hz	1417 Hz	2265 Hz
S3	662 Hz	1561 Hz	2420 Hz
S4	577 Hz	1235 Hz	1994 Hz
S5	617 Hz	1412 Hz	2237 Hz
S6	582 Hz	1306 Hz	2070 Hz

TABLE III. *Medians of the measured $F2_{burst}$ values for speaker TB (all numbers in Hz). $F2_{burst}$ values were measured at the burst of the stop consonants.*

		Labial		Alveolar		Velar		Palatal	
		b	p	d	t	g	k	ɟ	c
Back	ɔ	1045	1435	1074	2022	1066	1001	1560	2058
	o	830	1304	843	1651	878	841	1514	1714
	oː	817	1374	853	1632	782	793	1499	1860
	u	805	1486	852	1703	807	789	1587	1899
	uː	825	1435	805	1690	784	798	1526	2035
Front	aː	1236	1655	1714	2001	1752	1266	1638	2106
	ɛ	1518	1726	2021	2181	2101	1542	1753	2179
	ø	1348	1661	1374	2076	1524	1390	1695	2018
	øː	1518	1726	1635	2055	1688	1525	1729	2007
	y	1594	1841	1730	2149	1809	1569	1860	2116
	yː	1708	1899	1796	2198	1961	1774	1934	2149
	eː	1769	1894	2112	2242	2299	1997	1880	2264
	i	1939	2022	2242	2225	2292	1956	1947	2244
	iː	2014	2025	2235	2217	2266	2308	1945	2309

TABLE IV. *Medians of the measured $F2_{vowel}$ values for speaker TB (all numbers in Hz). $F2_{vowel}$ values were measured at the midpoint of the vowels.*

		Labial		Alveolar		Velar		Palatal	
		b	p	d	t	g	k	ɟ	c
Back	ɔ	1056	1251	1114	1295	1095	1037	1197	1322
	o	797	978	875	1003	845	786	958	1036
	oː	651	691	675	720	661	633	674	703
	u	691	878	749	919	712	686	849	976
	uː	619	712	640	691	644	552	678	728
Front	aː	1478	1506	1593	1560	1564	1504	1527	1541
	ɛ	1678	1716	1798	1846	1795	1706	1678	1812
	ø	1433	1500	1475	1583	1477	1446	1500	1613
	øː	1659	1680	1600	1702	1621	1602	1703	1663
	y	1803	1904	1740	1909	1782	1824	1975	1881
	yː	1953	2002	1824	1927	1849	1878	1911	1848
	eː	2278	2302	2288	2306	2300	2308	2287	2296
	i	2209	2281	2300	2255	2258	2240	2274	2190
	iː	2317	2380	2409	2334	2357	2312	2358	2358

D. Formant measurements

Sound boundaries of the "aCVba" utterances were labelled automatically using a Hungarian speech recognition engine (Mihajlik *et al.*, 2002) in forced aligned mode. Second formant frequencies were measured automatically in Praat (Boersma and Weenink, 2008) at the burst of the stop consonant and at the midpoint of the second vowel. Doubtful F2 values were hand-corrected.

Tables III and IV show the medians of the measured formant frequencies for the burst of the stop consonants and the midpoint of the vowels, respectively.

III. RESULTS

A. Locus equation space

From speaker TB's measured $F2$ and SGR frequencies a locus equation space was drawn (Fig. 2). As the figure shows, the locus equation space is separated into distinct regions by the subglottal resonances. The $F2_{burst}$ - $F2_{vowel}$ pairs can be clustered according to the place of articulation of the consonants and vowels.

In the figure, six regions are indicated by numbers, each rectangle (surrounded by SGRs) corresponds to a different group of CV transitions:

1. Labial and velar consonants with back vowels
2. Alveolar and palatal consonants with back vowels
3. Alveolar, labial and velar consonants with front vowels, except [i, iː, eː]
4. Alveolar and labial consonants with unrounded non-low front vowels ([i, iː, eː])
5. Palatal consonants with front vowels, except [i, iː, eː]
6. Palatal and velar consonants with unrounded non-low front vowels ([i, iː, eː])

These regions differ partly from the ones reported for American English (Chen and Lulich, 2009). It seems that in English, velars before all front vowels have $F2_{burst}$ higher than $Sg3$. In this Hungarian experiment, velars before front vowels, except [i, iː, eː] are below $Sg3$ and only velars before [i, iː, eː] have $F2_{burst}$ higher than $Sg3$. Palatal consonants were not reported in Chen and Lulich (2009).

It seems that SGRs classify the locus equation space well into distinct regions. However, some smaller CV groups spread over these boundaries. Palatal consonants followed by back vowels occupy a large space in $F2_{burst}$ direction, some of them have $F2_{burst}$ values higher than $Sg3$. A well-defineable group of palatals lies in the crossing of vertical $Sg2$ and horizontal $Sg3$ (with $F2_{vowel}$ higher than $Sg2$). Most of the palatal consonants followed by front vowels have $F2_{burst}$ values higher than $Sg3$, whereas approximately a third of the "palatal-front, except [i, iː, eː]" CV transitions are below $Sg3$. Labial consonants followed by vowels [i, iː, eː] are mainly in region 4, but some parts extend to region 6. A few extreme cases of these are visible in the figure, with the highest $F2_{burst}$ values over all CV transitions. Alveolar, labial and velar consonants followed by unrounded non-low front vowels are distributed across regions 4 and 6.

B. Locus equations of various consonants

Linear regressions were estimated for the different groups of CV transitions (according to the place of articulation of the consonant). The equations and the Pearson's coefficient of regression are shown in Table V. The linear regressions show that the slopes (m) and y-intercepts (b) differ for the groups. Alveolars and palatals have slopes of about 0.3, while for labials and velars the steepness is closer to 1. Labials and velars have higher Pearcon's coefficient values, explaining their shape in Fig. 2. These CVs have rather linear relation of $F2_{burst}$ and $F2_{vowel}$.

TABLE V. *Linear regression coefficients and Pearson's coefficient of regression for the different clusters of the locus equation space.* $F2_{burst} = m \cdot F2_{vowel} + b$

	m	b	R^2
Alveolar	0.333	1184.35	0.768
Labial	0.732	301.22	0.915
Palatal	0.307	1552.82	0.628
Velar	0.912	179.195	0.936

C. Relation of SGRs

From the subglottal resonance measurements shown in Table II (speakers S1-S6), dependence of $Sg1$ and $Sg3$ on $Sg2$ were calculated, separetely. The relation between SGRs seems to be linear. Linear regression calculations resulted in the following equations:

$Sg1 = 0.226 \cdot Sg2 + 285.743$ ($R^2 = 0.631$)

$Sg3 = 1.266 \cdot Sg2 + 433.354$ ($R^2 = 0.964$).

D. Subglottal resonance estimation

The algorithm described in Chen and Lulich (2009) was used for calculating $Sg2$ from the CV dataset of speaker TB. The algorithm selects iteratively m, $m+1$, $m+2$, ... data points and calculates their first order locus equations (FOLE). After that, a second order locus equation (SOLE) is calculated, from which $Sg2$ can be derived. As m grows, the calculation approximates a value ($\widetilde{Sg}2$), and the deviation is smaller and smaller. The goal is to calibrate the algortihm so that the estimated $\widetilde{Sg}2$ value is close to the real $Sg2$.

Therefore, in our experiment the original method was modified. For the estimation of $\widetilde{Sg}2$ from the locus eqation space, only CV transitions with a labial or a velar consonant were used. Fig. 3 shows the calculated $\widetilde{Sg}2$ as the size of the subset (m, number of data points) grows. When m is small (2, 3 or 4), $\widetilde{Sg}2$ is much higher (at about 2000 Hz) than the real value. As m reaches 7, $\widetilde{Sg}2$ is very close to the measured $Sg2$ (approximately 1250 Hz). As m grows further, $\widetilde{Sg}2$ goes below the real $Sg2$ value and alternates in the region of 1050–1150 Hz.

For the calculation of $Sg1$ and $Sg3$ based on $Sg2$, the previously shown equations were used. When the size of the subset is greater than 10, this results in the estimated subglottal resonance frequencies of about 530 Hz, 1100 Hz and 1830 Hz (approximately 10% lower than the measured values).

E. Classification of CV sequences based on estimated SGRs

A modified version of the classifiction algorithm introduced in Chen and Lulich (2009) was used, in order to classify the CV transitions into different groups corresponding to their place of articulation. As discussed earlier, the locus equation space of the Hungarian speaker TB differs partly from the one reported for American English in Chen and Lulich (2009). Regions 1–6 shown in Fig. 2 were taken into consideration in the classification. The boundaries of these regions can be described in the form of inequalities between SGRs, $F2_{burst}$ and $F2_{vowel}$

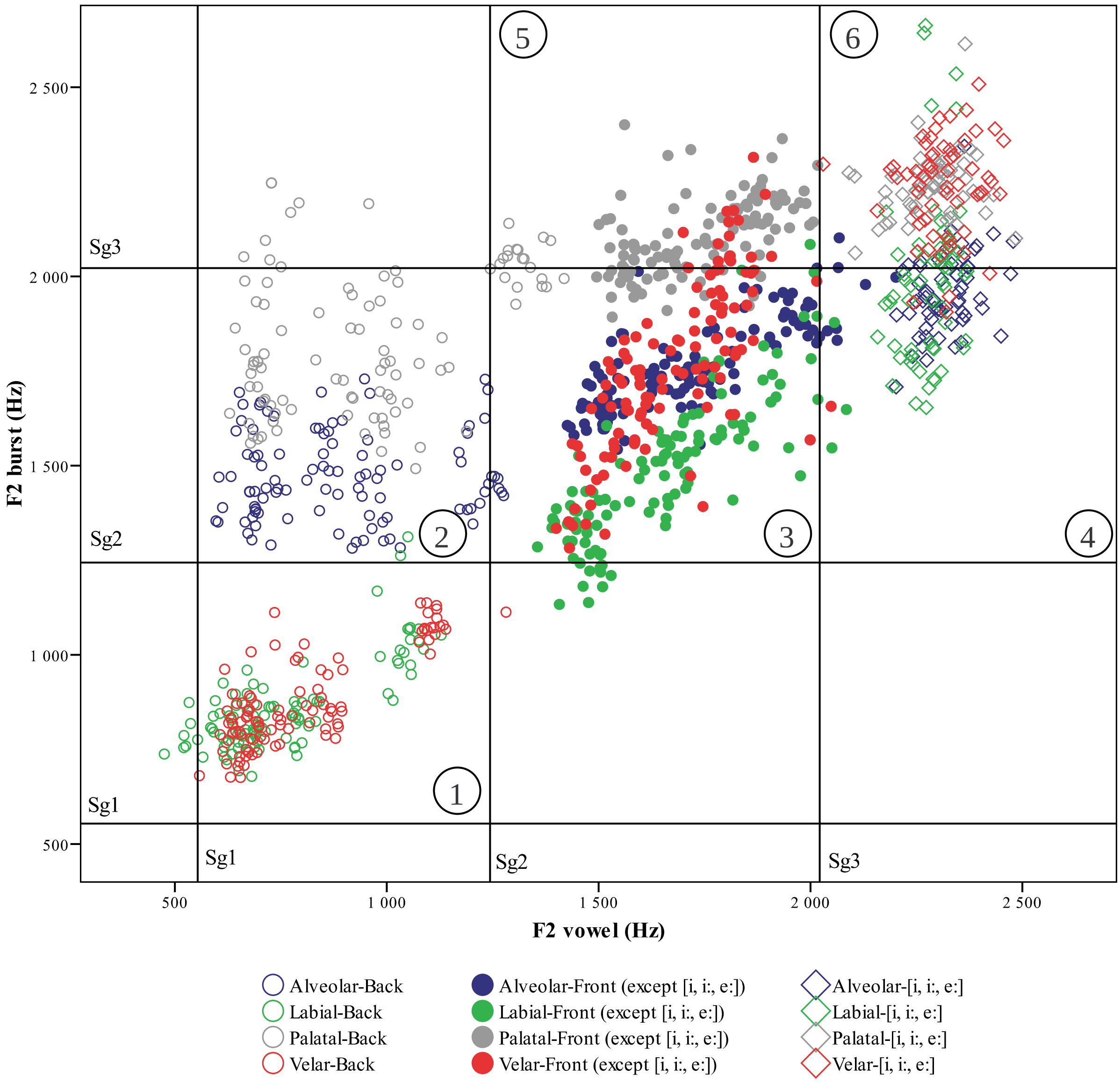


FIG. 2. Locus equation space of speaker TB. 1120 data points are shown, measured in CV transitions of nonsense words. Horizontal and vertical lines indicate subglottal resonances. Stop consonants with various place of articulation are denoted with different colors. Back, front (except [i, i:, e:]) and unrounded non-low front vowels ([i, i:, e:]) are denoted with different forms. The $F2_{onset}$-$F2_{vowel}$ pairs can be clustered according to the place of articulation of the consonants and vowels.

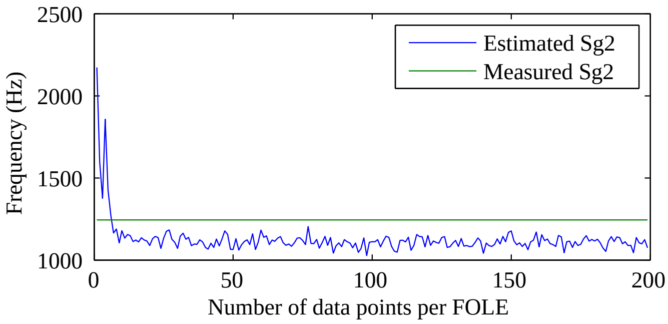


FIG. 3. Estimated value of the second subglottal resonance ($\widetilde{Sg}2$) as a function of the number of data points used in the calculation. $\widetilde{Sg}2$ reaches the measured $Sg2$, when the number of data points per FOLE is 7.

TABLE VI. *Inequalities describing the CV regions. (A stands for alveolar, L for labial, P for palatal and V for velar consonants. B stands for back, F for front except [i, i:, e:] and F* for front unrounded non-low vowels.)*

Region	CVs	Inequality 1	Inequality 2
1	LV-B	$Sg1 < F2_{burst} < Sg2$	$Sg1 < F2_{vowel} < Sg2$
2	AP-B	$Sg2 < F2_{burst} < Sg3$	$Sg1 < F2_{vowel} < Sg2$
3	ALV-F	$Sg2 < F2_{burst} < Sg3$	$Sg2 < F2_{vowel} < Sg3$
4	AL-F*	$Sg2 < F2_{burst} < Sg3$	$Sg3 < F2_{vowel}$
5	P-F	$Sg3 < F2_{burst}$	$Sg2 < F2_{vowel} < Sg3$
6	PV-F*	$Sg3 < F2_{burst}$	$Sg3 < F2_{vowel}$

values. Table VI shows the inequalities corresponding to each region.

TABLE VII. *Consonant hit rates and false alarm rates of a baseline classification based on measured SGRs. (C stands for all consonants summarized, A for alveolar, L for labial, P for palatal and V for velar consonants. B stands for back, F for front except [i, iː, eː] and F* for front unrounded non-low vowels.)*

	C	LV-B	AP-B	ALV-F	AL-F*	P-F	PV-F*
Hit rate	84	95	83	89	65	74	97
False alarm rate	2.4	0	0.2	5.5	1.3	2.9	4.3

TABLE VIII. *Vowel hit rates and false alarm rates of a baseline classification based on measured SGRs. (V stands for all vowels summarized.)*

	V	B	F	F*
Hit rate	97	93	97	100
False alarm rate	1.8	0	4.1	1.3

TABLE IX. *Consonant hit rates and false alarm rates of the classification based on estimated SGRs.*

	C	LV-B	AP-B	ALV-F	AL-F*	P-F	PV-F*
Hit rate	67	88	64	68	18	64	100
False alarm rate	5.3	0	0.5	3.8	2.8	5.7	19.3

TABLE X. *Vowel hit rates and false alarm rates of the classification based on estimated SGRs.*

	V	B	F	F*
Hit rate	87	85	75	100
False alarm rate	7.7	0	9.2	13.9

First, a baseline classification was carried out with speaker TB's measured subglottal resonances for the above-mentioned regions. The hit rates and false alarm rates are shown in Tables VII and VIII for consonants and vowels, respectively. For the consonants altogether, the hit rate is 84%, while the false alarm rate is 2.4%. For all of the vowels, the hit rate is 97% and the false alarm rate is 1.8%.

Our goal was to come close to the hit rates and false alarm rates of the baseline classification. The modified algorithm was run on the CV dataset of speaker TB. The algorithm runs with growing subset size, as described earlier. First it estimates the second subglottal resonance ($\widetilde{Sg}2$), after that it calculates the $\widetilde{Sg}1$ and $\widetilde{Sg}3$ on the basis of the linear relation between them. Using these estimated SGRs, a classification is carried out.

Fig. 4 shows the resulting hit rates and false alarm rates of the classification, showing the tendency while the subset size is growing. Hit rates are highest when the number of data points in each subset is small (m is about 7–8). This is the consequence of the $Sg2$ estimation method, Fig. 3 showed that $\widetilde{Sg}2$ was closest to the measured $Sg2$, when $m = 7$.

The overall hit rates and false alarm rates of the classification algorithm are shown in Tables IX and X (for subset size of 200). For the consonants, the overall hit rate is 67%. High hit rates were reached for most consonant types, except "AL-F*". This was caused by the overlapping dense data points in Regions 4 and 6. As the estimated SGRs are approximately 10% lower than the real values, Region 4 was moved lower in $F2_{burst}$ direction. The high false alarm rates of group "PV-F*" are due to the same shift. For the classification of the vowels, an overall hit rate of 87% was obtained. False alarm rates of unrounded non-low front vowels (F* group, [i, iː, eː]) are as high as 14%, because estimated SGRs shifted the boundaries of the classes lower in the $F2_{vowel}$ region.

IV. DISCUSSION AND CONCLUSIONS

In this paper, a series of experiments investigating locus equation space and the separating role of subglottal resonances were described. First, the locus equation space of the native speaker of Hungarian TB was examined, and six regions were defined according to the place of articulation of the consonant-vowel transitions. The locus equations were calculated for the various consonant types, and velar and labial consonants were chosen for use in an $Sg2$ estimation algorithm. The relation of $Sg1$, $Sg2$ and $Sg3$ was investigated in order to calculate the SGRs from the estimated $\widetilde{Sg}2$. The algorithm described in Chen and Lulich (2009) was modified, and applied for the CV dataset. From the estimated SGRs consonant and vowel classifications were run. These resulted in 10-15% lower hit rates than a baseline classification, in which the measured subglottal resonance frequencies of the same speaker were used. The results of the classification are similar to the one reported in Chen and Lulich (2009) for American English.

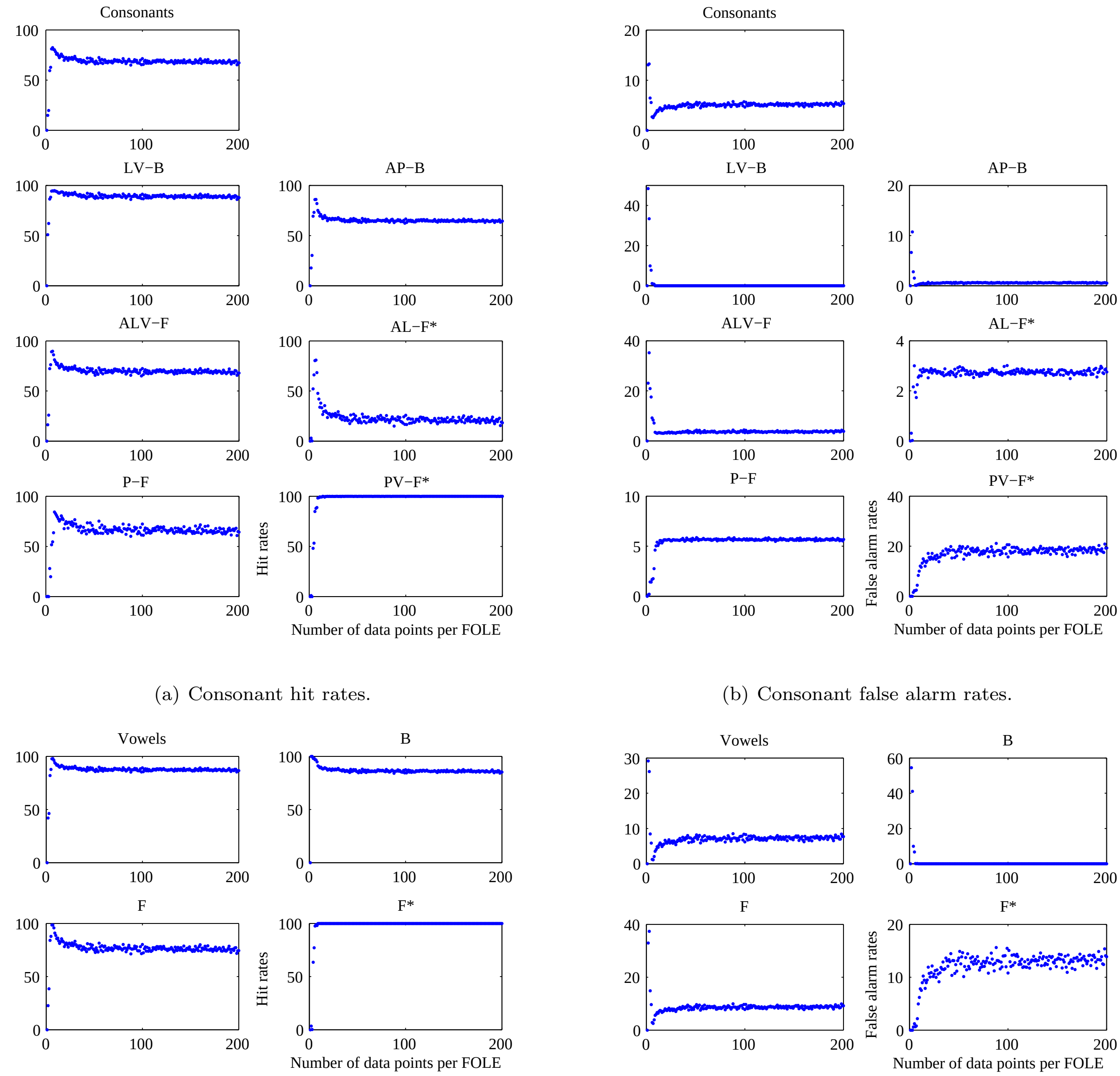
Future work includes the improvement of the $Sg2$ estimation algorithm. In this experiment, the estimated SGRs were approximately 10% lower than the measured SGRs, while in Chen and Lulich (2009) higher estimations were reported than the real subglottal resonance frequencies. If estimated SGRs were closer to the real ones, higher hit rates and lower false alarm rates could be reached in the classification of the CV sequences.

In this work only the data of one speaker was analyzed. A further experiment in this topic should investigate the classification of CV transitions with more speakers. It should be examined, if subglottal resonances play a similar separating role in the locus equation space of other speakers.

Acknowledgements

We would like to thank Steven M. Lulich for handing out this task. His lectures and explanations related to subglottal resonances helped a lot in understanding the role of SGRs, and contributed to the preparation of this final project.

We also thank Tamás Bőhm for recording the CV dataset, and speakers S1-S6 for their accelerometer recordings.



(a) Consonant hit rates.

(b) Consonant false alarm rates.

(c) Vowel hit rates.

(d) Vowel false alarm rates.

FIG. 4. Hit rates and false alarm rates for consonants and vowels. (A stands for alveolar, L for labial, P for palatal and V for velar consonants. B stands for back, F for front except [i, iː, eː] and F* for front unrounded non-low vowels.)

References

Boersma, P. and Weenink, D. (**2008**). *Praat [Computer program] (Version 5.0.20)*, URL `http://www.praat.org`, accessed on 11 Dec 2008.

Chen, N. F. and Lulich, S. M. (**2009**). "Context-free classification of consonant-vowel transitions based on subglottal resonances and the second formant", JASA **125**.

Chi, X. and Sonderegger, M. (**2007**). "Subglottal coupling and its influence on vowel formants", JASA **122**, 1735–1745.

Csapó, T. G., Bárkányi, Z., Gráczi, T. E., Bőhm, T., and Lulich, S. M. (**2009**). "Relation of formants and subglottal resonances in Hungarian vowels", in *Proceedings of Interspeech*, (in press).

Lulich, S. M. (**2008**). "On the relation between locus equations and subglottal resonances", JASA **124**, 2558.

Lulich, S. M. (**2009**). "Subglottal resonances and distinctive features", Journal of Phonetics Doi:10.1016/j.wocn.2008.10.006.

Mihajlik, P., Révész, T., and Tatai, P. (**2002**). "Phonetic transcription in automatic speech recognition", Acta Linguistica Hungarica **49**, 407–425.

Sjőlander, K. and Beskow, J. (**2009**). *Wavesurfer [Computer program] (Version 1.8.5)*, URL `http://www.speech.kth.se/wavesurfer`, accessed on 3 Apr 2009.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).