An acoustic study of standard Dutch /v/, /f/, /z/ and /s/

Mikhail Kissine, Hans Van de Velde and Roeland van Hout
Vrije Universiteit Brussel / Utrecht University /
Nijmegen University

1. Introduction

One of the main pronunciation differences between the varieties of standard Dutch as spoken in the Netherlands and Flanders, is to be found in the voice characteristics of /ɣ/, /v/ and /z/. The voiced fricatives are in a process of devoicing in northern standard Dutch and real time data showed that they are affected in the following order: /ɣ/ – /v/ – /z/ (Van de Velde, Gerritsen & Van Hout 1996). In southern standard Dutch /ɣ/ is often (partially) devoiced (Debrock 1977, Van de Velde et al. 1996), but Van de Velde & Van Hout (2001) found that /v/ and /z/ may be devoiced too by Flemish speakers. In this paper we present acoustic analyses of /v/ and /z/ and of their voiceless counterparts /f/ and /s/, having three aims in mind. First, we want to evaluate whether acoustic techniques are useful for quantitative sociolinguistic studies of consonantal variation. Until now acoustic techniques have mainly been applied to vowels (cf. Thomas 2002). Second, we want to develop measurements that uncover the core voice characteristics of the Dutch alveolar and labial fricatives (/s/, /z/, /f/, /v/). Third, we want to gain more insight in the process of fricative devoicing in standard Dutch, especially in the devoicing process of /v/ and /z/ observed in Flanders and the Limburg region in the Netherlands (Van de Velde & Van Hout 2001). To get a better understanding of this process, the voiceless consonants /f/ and /s/ were included in the analyses too. The sociolinguistic data used in this article are described in Section 2. Section 3 discusses the segmentation procedure, the acoustic measurements and the perceptually distinguished variants. The results are presented in Section 4. The conclusions of this study are summarized in Section 5.

Linguistics in the Netherlands 2003, 93–104.
ISSN 0929–7332 / E-ISSN 1569–9919

2. Subjects and speech material

The subjects are 160 Dutch language teachers, stratified for community (2), region (4), sex (2) and age (2), as can be seen in Table 1. The subjects were selected from schools in middle-sized cities. All subjects participated in a Flemish-Dutch research project on the pronunciation of standard Dutch (Van Hout et al. 1999).

Table 1. The corpus of Dutch language teachers, stratified for community, region, sex and age (N = 160)

		Core	Intermediate	Peripheral 1	Peripheral 2
The Netherlands		Randstad	Middle	North	South
Young	Male	5	5	5	5
	Female	5	5	5	5
Middle	Male	5	5	5	5
	Female	5	5	5	5
Flanders		Brabant	East-Flanders	West-Flanders	Limburg
Young	Male	5	5	5	5
	Female	5	5	5	5
Middle	Male	5	5	5	5
	Female	5	5	5	5

Subjects were selected in four regions in both the Netherlands and Flanders. In the Netherlands these regions are: 1. Randstad, i.e. the economic and cultural centre of the Netherlands, which also appears to be the core area for ongoing changes in the standard language (cities: Alphen aan den Rijn, Gouda); 2. Middle, i.e. an intermediate zone in the South of the province of Gelderland, along the borders of the Great Rivers (Culemborg, Ede, Elst, Tiel, Veenendaal); 3. North, a peripheral area in Groningen and the North of Drenthe (Assen, Veendam, Winschoten); 4. South, a second peripheral area, i.e. Limburg (Geleen, Roermond, Sittard). In Flanders we were able to cover the four dialect regions: 1. Brabant, i.e. the economic and cultural centre of the Flemish communities, which also appears to be the core area for ongoing changes in the standard language (Heist-op-den Berg, Lier); 2. East-Flanders, an intermediate zone (Oudenaarde, Zottegem); 3. West-Flanders, a peripheral zone in the west (Ieper and Poperinge); 4. Limburg, a second peripheral area in the east (Bilzen, Tongeren). At the time of data collection, subjects were living in the region, had lived there before their 8th birthday, and had been living there for at least eight years before their 18th birthday. Two age groups were distinguished: young (between 22 and 40) and middle (between 45 and 60). For sex, a biological distinction between male and female was made.

The subjects were instructed about the aim of the research project: a study of standard Dutch pronunciation in the Netherlands and Flanders. Part of the questionnaire used for this study aims at eliciting the best articulated realization of all phonemes of Dutch in a linguistic context which is the same for all vowels and consonants respectively. Therefore, the phonemes are embedded in a stressed syllable and put in a carrier sentence. For consonants in word initial position, a schwa-like environment is the most neutral context. The schwa is the most central vowel and is unrounded. However, it cannot occur in a stressed position. Therefore, the word initial consonant is preceded by a word ending in schwa (i.e. *de*) and followed by /œy/. For the variables (f), (v), (s) and (z) the carrier sentences are:

In de fuize horen we f	(in the 'fuize' we hear 'f')
In de vuize horen we v	(in the 'vuize' we hear 'v')
In de suize horen we s	(in the 'suize' we hear 's')
In de zuize horen we z	(in the 'zuize' we hear 'z')

The subjects were instructed to pronounce the single consonant as a combination of the consonant with schwa. However, only the target variable in the first part of the utterance is used in this study. 17 sentences covering the Dutch consonants in word initial position were presented with intervals of three seconds on the screen of a laptop computer. The subjects had to do the reading task twice, with an interval of about 20 minutes. Five random orders were used, each order occurring once in every cell (see Table 1). For the second task the items were presented in reverse order. The speech was recorded on digital audiotape with a portable TASCAM DA-P1 recorder and an AKG C420 headset microphone. The recordings were digitalized on computer and down-sampled to 16 kHz (16 bits). Then, the four target sentences were extracted from the database and saved as seperate sound files. The total number of tokens in this study is 1280: 160 speakers $\times 4$ variables $\times 2$ realizations per variable.

3. Segmentation, acoustic measurements and transcription

We used *Signal Segmentor 1.01* for speech segmentation, a program developed by Alain Soquet. The program displays for each sound file a wave signal, a spectrogram and a 512-point FFT spectrum corresponding to the cursor position on the spectrogram and the wave signal. It presets five boundaries that can be moved by means of the mouse to any place in the spectrogram or wave signal. The time points of the boundaries are automatically saved. The spectrogram can display formants (computed with standard LPC-binomial) and F_0. To enable a comparison of the duration of voiced and voiceless fricatives, we decided to include 20% of the F_1

transitions of the fricatives on each side.[1] The 20% border was fixed visually on the basis of the spectrogram and the FFT, and verified by means of an auditory control.

Three acoustic measurements were performed on each token of (f), (v), (s) and (z): the duration of the fricative in time (DURATION), the mean intensity of the friction noise (NOISE) and the pitch extent throughout the fricative (PITCH). We briefly motivate the choice for these measurements and explain how they are applied. All acoustic measurements were done with PRAAT.

Duration

Voiceless fricatives tend to be longer than their voiced counterparts in Dutch (Slis & Cohen 1969a, Slis & Van Heugten 1989, Debrock 1977). The duration (in ms.) was computed from the beginning to the end point of the fricative. Measurements of relative durations (fricative/syllable and fricative/utterance), correlate strongly with the absolute durations ($r = .844$ and .771). Therefore, in this paper the influence of voicing on duration is evaluated by means of the absolute durations. It is expected that (f) is longer than (v) and (s) longer than (z).

Noise

Slis & Cohen (1969a) for stops and Van den Berg (1989) for fricatives (in a two-obstruent cluster) showed that the intensity of the friction noise is a perceptual cue for the voiced–voiceless distinction. However, the total intensity (i.e. the sum of the intensity of the constriction-generated noise and the periodic, glottis-generated vibration intensity) has an instable relationship with the voiced (lenis) vs. voiceless (fortis) character of the fricative (Debrock 1977).

The prediction was that the mean noise intensity should be higher in voiceless ((f) and (s)) than in voiced ((v) and (z)) fricatives. The intra-oral pressure causing the supra-glottal noise generation depends on the pressure drop at the glottal level. In order to produce vocal cord vibration at the same time as a noise generation across an oral constriction, the two cross-sectional constrictions have to be about equal. Consequently, if the cross-sectional area of the glottis constriction increases, the intra-oral pressure, and therefore the intensity of the supra-glottal constriction noise, will increase (Stevens et al. 1992; Slis 1980; Clark & Yallop 1995). In other words, while the vocal cords are adducted, the intra-oral pressure will be lower. Slis (1970) presents transillumination data that show a closed glottis in all voiced fricative phonemes. The adduction of vocal folds reduces the amplitude of the noise generated by a supra-glottal constriction (Stevens et al. 1992).

From the beginning to the end point of the fricative an intensity noise measure (in dB) was calculated with intervals of 10 ms. The following formula was used, in which I stands for the intensity of the signal and H for the harmonicity or degree of periodicity: Intensity of friction $= 10 \log_{10} [10^{I/10}/(1+10^{H/10})]$.[2] Then a mean noise intensity was computed for each token.

Pitch

The major cue for the voiced/voiceless distinction in Dutch seems to be the presence or the absence of vocal cord vibration in the fricative itself (Slis & Cohen 1969a, b; Van den Berg 1998). Vocal cord vibration can be measured by computing pitch extent. A standard way to do this is a measurement of the first harmonic amplitude (Stevens et al. 1992). However, due to technical problems with the computation of those values for a large sample of tokens this method was rejected. As an alternative, the pitch value (in Hertz) was computed with intervals of 10 ms using the auto-correlation method. The presence of periodicity was evaluated between the minimum and maximum pitch values of the speaker within the specific utterance. Finally, the number of samples with pitch was divided by the total number of samples, resulting in a relative pitch extent index, ranging between 0 (no pitch at all) and 100 (pitch present during the whole fricative). These pitch values are expected to be higher in voiced than in voiceless realizations.

Voice

Auditory transcriptions of the voice characteristics were made of all tokens of (f), (v), (s) and (z). The results of the consensus transcriptions for (v) and (z) were presented in Van de Velde & Van Hout (2001). For (f) and (s) a different procedure was used: transcriptions were made by only one judge, the second author of this study. Three variants were distinguished for (f), (v), (s) and (z): fully voiceless, partially voiced, and fully voiced. Index scores were calculated ranging from 0 (fully voiceless) to 100 (fully voiced).

4. Results

The results for the dependent variables duration, noise, pitch and voice are summarized in Table 2. For duration the expected differences between voiced and voiceless fricatives show up: (f) and (s) are significantly longer than (z) and (v). There is no difference between (f) and (s), but the mean duration of (v) is about 14 ms shorter than (z). There is a difference in the noise intensitity between labio-dentals and alveolars, the latter being louder (see Ladefoged & Maddieson 1996). (s) is 2dB louder than (z), but the difference between (f) and (v) is very small. As expected (v) and (z) have a higher pitch extent value than (f) and (s), indicating that (v) and (z) are not completely devoiced. The presence of pitch in voiceless (f) and (s) is mainly due to co-articulation with the surrounding vowels. Unexpected is the higher pitch value for (f) than (s). The mean perceived voice for (v) and (z) is the same (see Van de Velde & Van Hout 2001). Both (f) and (s) were only once perceived as having voice.

Table 3 shows the Pearson correlation coefficients between duration, noise,

Table 2. Mean values of DURATION, NOISE, PITCH and VOICE of (f), (s), (v) and (z) (N = 160)

	(f)	(s)	(v)	(z)
DURATION	177.2	176.2	140.0	154.9
NOISE	63.2	67.9	62.8	65.9
PITCH	18.8	13.3	57.5	57.5
VOICE	0.3	0.2	57.7	59.4

Table 3. Pearson correlation coefficients between the dependent variables for the labio-dental and alveolar fricatives ($p < .05$)

	labio-dental (v) and (f)			alveolar (z) and (s)		
	NOISE	PITCH	VOICE	NOISE	PITCH	VOICE
DURATION	−.105	−.332	−.377	n.s.	−.438	−.384
NOISE		−.168	−.190		−.215	−.268
PITCH			.784			.884

PITCH and VOICE in labio-dental and alveolar fricatives. All correlations are significant ($p < .05$), except the correlation of NOISE with DURATION for the alveolar fricatives. There is a strong correlation between PITCH and VOICE. It confirms that pitch is the most important indicator for the identification of voiced fricatives in Dutch (Slis & Cohen 1969a & b, Van den Berg 1988) and provides acoustic support for the traditional auditory scoring techniques. The correlations of VOICE with DURATION and NOISE are negative and significant, as predicted. Both DURATION and NOISE systematically vary between voiced and voiceless fricatives, but their role is evidently smaller than the role of PITCH.

The results of the analyses of variance (anova) for the four dependent variables and the external factors community, region (for Flanders and the Netherlands seperately), sex and age are presented in Table 4. It shows that that there are effects of the factors community, region and sex for both voiced and voiceless fricatives, but that the patterns differ substantially. For age, only one out of sixteen variables is significant. In this heavily monitored speech material there are no differences in apparent time between the two age groups of this study, which might suggest that this change in progress has reached a period of more or less stable variation. Data from less monitored speech are needed to gain more insight in this issue. We will have a closer look at the geographical differences, which are visualised in Figures 1 to 4. The sex differences fall outside the scope of this paper and will not be discussed.

Van de Velde & Van Hout (2001) observed devoicing of (v) and (z) in Flanders, which is generally considered to have voiced realizations of /v/ and /z/. In the

Table 4. Anova results for the external factors community, region, sex and age, split up for (f), (s), (v) and (z). Significant effects are marked with * ($p<.05$), ** ($p<.01$) or *** ($p<.001$)

		community	region NL	region FL	sex	age
(f)	DURATION	***	*		***	
	NOISE	***	**	***	**	*
	PITCH				***	
	VOICE					
(s)	DURATION	***		**	***	
	NOISE	***	*	***		
	PITCH	***			***	
	VOICE					
(v)	DURATION		*		**	
	NOISE	***	**	***	*	
	PITCH	***	***			
	VOICE	***	***	**		
(z)	DURATION		***			
	NOISE	***	*	***	**	
	PITCH		*			
	VOICE	**				

Netherlands surprisingly strong devoicing of (z) showed up in the Limburg province (N-S), see also Figure 1. The mean scores for PITCH of (v) and (z) in Figure 2 and the statistical analyses confirm the observations based on auditory transcriptions. Van de Velde et al. (1996) showed that in standard Dutch as spoken by Dutch broadcasters between 1935 and 1993, (v) is systematically devoiced more than (z). Most of the Dutch speakers of that real-time study originate from the Randstad and are comparable with the N-R subjects of this study. However, this pattern is not found in all regions. In N-R and N-N the expected difference is found: more devoicing of (v) than (z). In N-M there is no difference between (v) and (z). But in N-S, F-L and F-W the opposite tendency is found: (z) is devoiced more than (v). In F-E and F-B (v) and (z) are equally devoiced, which is in line with the weak devoicing of (v) and (z) found in the speech of the Flemish 1993 broadcasters (Van de Velde at al 1996: 165).

On the basis of an auditory transcription of their voice characteristics we were not able to detect variation for (s) and (f). However, the acoustic analysis revealed that (f) has higher PITCH values than (s) and that the PITCH of (s) is significantly higher in the Netherlands than in Flanders.

Figure 3 reveals strong geographical differences for DURATION of (z), (s) and (f), but not (v). First, there are differences between the communities: (f) and (s) are

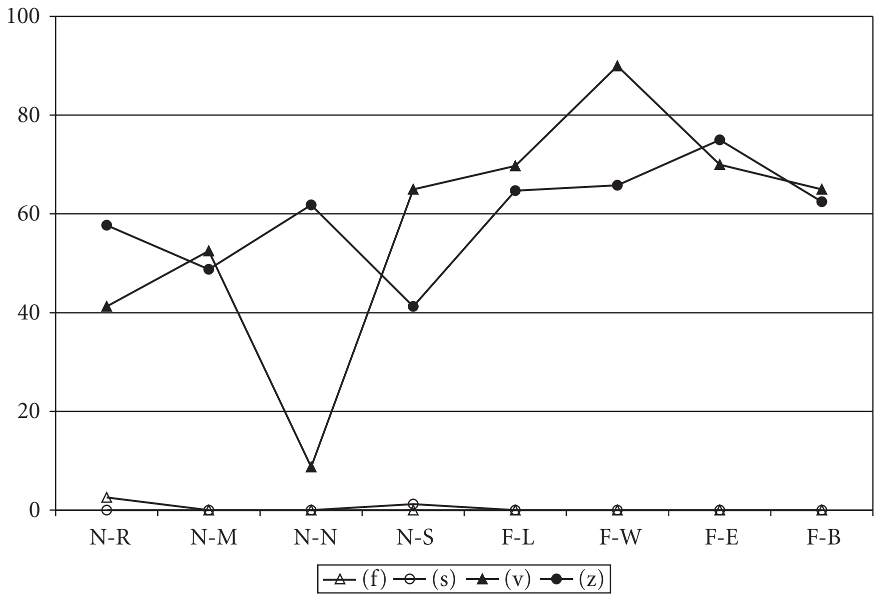

Figure 1. Mean voice for (f), (s), (v) and (z) split up by region.

shorter in the Netherlands than in Flanders, but (z) is shorter in Flanders than in the Netherlands. Second, there are obvious regional differences within each community: (s), (f) and (z) are the longest in the two Limburg provinces (N-S and F-L). Third, in all regions (z) is longer than (v), except in N-M and N-N, where no difference is found. Figure 4 shows the geographical differences for NOISE. First, all fricatives are louder in the Netherlands than in Flanders. Second, in Flanders F-B has the lowest values for all fricatives, F-L has higher values than the other regions for the alveolar pair.

Can we explain the geographical differences in devoicing of (v) and (z) by having a closer look, on the one hand, at the relationship between perceived voice and the three acoustic variables, and, on the other hand, at the relationship between the realizations of the fricative pairs (v)/(z) and (f)/(s)? We need to look first at Flemish and Dutch Limburg (F-L and N-S), where devoicing is surprisingly strong and where (z) is devoiced more than (v). Furthermore, Van de Velde & Van Hout (2001) observed large intra-speaker variation for (z) in N-S and F-L, suggesting linguistic insecurity. An explanation for the unexpected voice patterns might be found in the duration patterns. In F-L and N-S (f), (s) and (z) are longer than in the other regions, a pattern that does not apply to (v). A long articulation of a fricative may result in a subglottal pressure drop. Consequently, subglottal pressure may become too low to maintain vocal cord vibration during the entire fricative, but, at the same time, PITCH is the most salient cue in perceiving voice in Dutch fricatives.

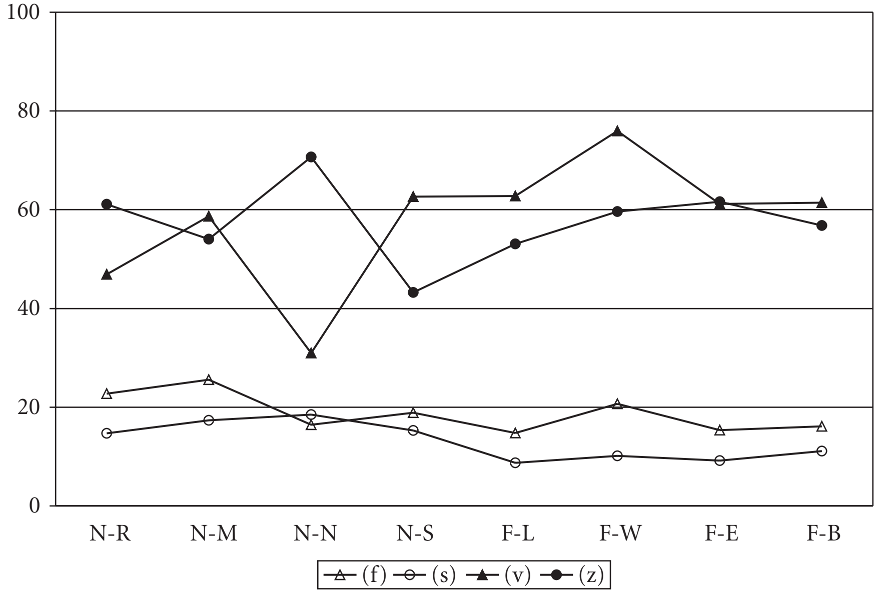

Figure 2. Mean pitch for (f), (s), (v) and (z) split up by region.

Figure 2 shows that in F-L and N-S the lowest pitch values for (z) are found in the respective communities. Consequently, the strong devoicing of (z) may be the concomitant phonetic effect of longer articulations of fricatives in these regions.

Table 5. Pearson correlations ($p < .05$) for the labio-dental fricatives in the two Limburg provinces

	N-S			F-L		
	noise	pitch	voice	noise	pitch	voice
duration	n.s	.404	−.492	.128	.390	−.517
noise		−.243	n.s.		−.243	−.236
pitch			.800			.848

But why do we not find longer durations for (v) in N-S and F-L? Table 5 shows correlations between DURATION, NOISE, PITCH and VOICE for the labio-dental fricatives in F-L and N-S. (v) is the shortest of all fricatives in this study, as can be seen from Table 2 and Figure 3. On the one hand the correlations between DURATION and VOICE are stronger in the Limburg provinces (r = −.494 en r = −517) than the overall mean (r = −.377, see Table 3). On the other hand, the correlations

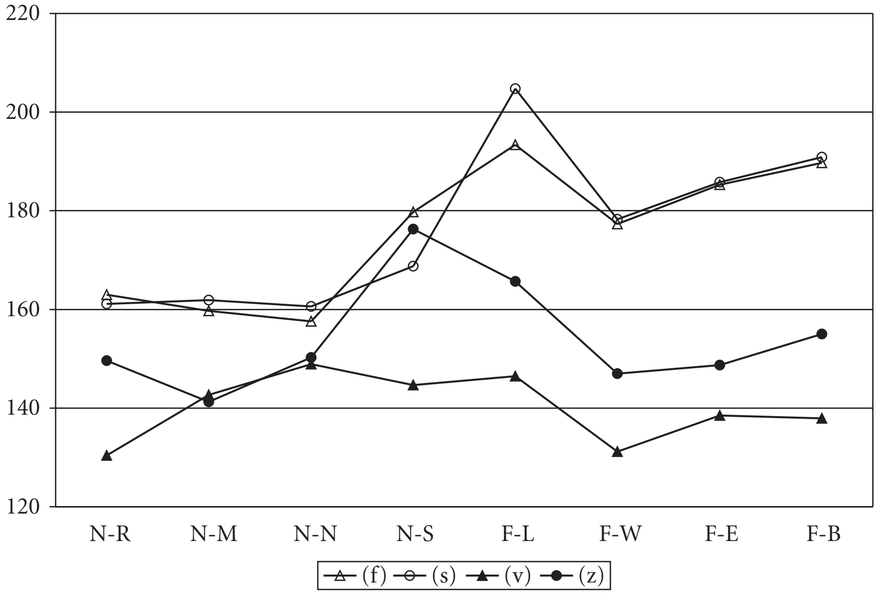

Figure 3. Mean duration for (f), (s), (v) and (z) split up by region.

between NOISE and VOICE are weaker in N-S and F-L than the overall correlation (see Table 3). Furthermore, from Figure 4 it can be seen that there is no difference in NOISE between (v) and (f). This supports the idea that in order to reach the target voiced pronunciation in this highly monitored reading task speakers tend to keep (v) short, which leads to the avoidance of a longer duration that might result in devoicing. Such a strategy needs not to be applied to (z), as there is a clear NOISE distinction between (s) and (z). Therefore (z) is articulated with a longer duration. This might explain why (z) is devoiced more than (v) in N-S and F-L, contrary to the patterns found in N-R and N-N and in broadcasters' speech (Van de Velde, Gerritsen & Van Hout 1996).

5. Conclusion

The first aim of this study was to evaluate the usefulness of acoustical techniques for quantitative sociolinguistic studies of consonantal variation. The high correlations between the acoustic measurements of pitch measurements and the perceptual transcription of voice, as predicted by the literature, support the validity of auditory transcriptions for tracing patterns of variation. However, the acoustic measurements give additional information. Duration and noise are concomitant characteristics in fricative realizations, both for voiced and voiceless variants. The acoustic

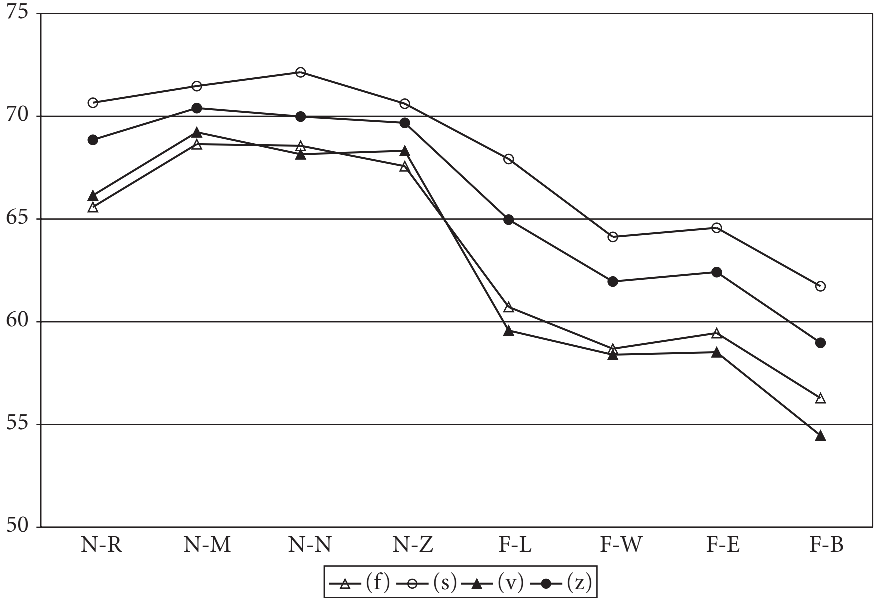

Figure 4. Mean noise for (f), (s), (v) and (z) split up by region.

measurements show systematic patterns of linguistic variation that are not mentioned in previous discussions on the devoicing of the Dutch fricatives. Measurements on duration and pitch may open the door for investigating the role of such a factor. However, the most striking result is that systematic patterns of variation were found for the voiceless fricatives too, especially for duration. These patterns are not mentioned in the sociolinguistic and phonological literature on Dutch. This relates to our second aim, uncovering the core voice characteristics of the Dutch alveolar and labial fricatives. We also showed that the measurements of the voiceless fricatives provide insight in the patterns of variation of the voiced fricatives, and help us understand the process of fricative devoicing in standard Dutch, especially in the devoicing process of /v/ and /z/ observed in Flanders and the Limburg region in the Netherlands (aim 3). The differences in the order of devoicing of /v/ and /z/ can be explained by differences in duration patterns.

Notes

1. Slis & Van Heugten (1989) and Stevens et al. (1992) showed that the durations of voiced and devoiced fricatives are almost identical if they are computed between the points where F1 transitions start and finish.

2. The intensity of the friction is assumed to be equivalent with the amount of the aperiodic energy of the signal. H = $10\log_{10}$ (periodic energy/aperiodic energy) and I = $10\log_{10}$ (periodic energy + aperiodic energy).

References

Berg, R. J. H. van den, (1989) 'Perception of voicing in Dutch two-obstruent sequences: covariation of voicing cues'. *Speech communication* 8, 17–25.

Berg, R. J. H. van den and I. H. Slis (1985) 'Perception of assimilation of voice as a function of segmental duration and linguistic context'. *Phonetica* 42, 25–38.

Clark, J. and C. Yallop (1995) *An introduction to phonectics and phonology*. Blackwell, Oxford. (2nd ed.)

Debrock, M. (1977) 'An acoustic correlate of the force of articulation'. *Journal of Phonetics* 5, 61–80.

Hout, R. van, G. De Schutter, E. De Crom, W. Huinck, H. Kloots and H. Van de Velde (1999) 'De uitspraak van het Standaard-Nederlands. Variatie en varianten in Vlaanderen en Nederland'. In E. Huls and B. Weltens (Eds.), *Artikelen van de Derde Sociolinguïstische Conferentie*. Eburon, Delft, 183–196.

Ladefoged, P. and I. Maddieson (1996) *The sounds of the world's languages*. Blackwell, Oxford.

Slis, I. H. (1970) 'Articulatory measurments on voiced, voiceless and nasal consonants: a test of a model'. *Phonetica* 21, 193–210.

Slis, I. H. (1983) 'Assimilation of voice as function of stress and word boundaries'. *Proceedings of the Institute of Phonetics of Nijmegen* 7, 74–82.

Slis, I. H. and A. Cohen (1969a) 'On the complex regulating the voiced-voiceless distinction I'. *Language and Speech* 12 (2), 80–112.

Slis, I. H. and A. Cohen (1969b) 'On the complex regulating the voiced-voiceless distinction II'. *Language and Speech* 12 (3), 137–155.

Slis, I. H. and M. van Heugten (1989) 'Voiced-voiceless distinction in Dutch fricatives'. *Linguistics in the Netherlands* 6, 123–132.

Stevens, K. N. et al., (1992) 'Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters'. *Journal Acoust. Soc. Am.* 91, 2979–3000.

Thomas, E. K. (2002) 'Instrumental Phonetics'. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change*. Oxford: Blackwell, 168–200.

Van de Velde, H., M. Gerritsen and R. van Hout (1996) 'The devoicing of fricatives in Standard Dutch: A real-time study based on radio recordings'. *Language Variation and Change* 8, 149–175.

Van de Velde, H. & R. van Hout (2001) 'The Devoicing of Fricatives in a Reading Task'. In T. van der Wouden & H. Broekhuis (eds.), *Linguistics in the Netherlands 2001*. Amsterdam: John Benjamins, 225–236.