University of Pennsylvania Working Papers in Linguistics

Volume 10
Issue 2 *Selected Papers from NWAVE 32*

Article 12

1-1-2005

Acoustic Contributions to Sociolinguistics: Devoicing of /v/ and /z/ in Dutch

Mikhail Kissine

Hans Van de Velde

Roeland van Hout

Acoustic Contributions to Sociolinguistics: Devoicing of /v/ and /z/ in Dutch

Acoustic Contributions to Sociolinguistics: Devoicing of /v/ and /z/ in Dutch

Mikhail Kissine, Hans Van de Velde, and Roeland van Hout

1 Introduction

This paper has a threefold aim. First, we want to evaluate whether acoustic techniques are applicable to quantitative sociolinguistic studies of consonantal variation. Until now, sociolinguists have mainly used acoustic techniques for the study of vowels (Thomas 2002, Foulkes 2002). Second, we want to develop acoustic measurements that uncover the core voice characteristics of the Dutch alveolar and labial fricatives. Third, we want to gain more insight into the devoicing of /v/ and /z/ in standard Dutch.

Van de Velde, Gerritsen and Van Hout (1996) showed in a real-time study that devoicing of /v/ and /z/ is a change in progress in northern standard Dutch (as spoken by broadcasters in the Netherlands) between 1935 and 1993. In southern standard Dutch (as spoken by broadcasters in Flanders) weak devoicing of /v/ and /z/ was—surprisingly—observed in the speech of the 1990's. In a study of regional variation in contemporary standard Dutch pronunciation, Van de Velde and Van Hout (2001) showed that devoicing of /v/ and /z/ occurs throughout the language area, but that it is much more advanced in the Netherlands than in Flanders. In this paper we present a detailed acoustic analysis of the same data. Next to regional differences in the strength of devoicing, variation in its acoustic correlates is found. Furthermore, regional variation showed up in the realization of voiceless /f/ and /s/ (Kissine, Van de Velde and Van Hout 2003). Some of these variation patterns contradict the traditional assumptions about the /v/-/f/ and /z/-/s/ contrasts in Dutch.

2 Research method

2.1 Participants

The participants of this study are 160 Dutch language teachers, stratified for community (2), region (4), sex (2) and age (2), as can be seen in Table 1. They participated in a Flemish-Dutch research project on the pronunciation of standard Dutch (Van Hout et al. 1999).

		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

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

Participants were selected from middle-sized cities in four regions in both the Netherlands and Flanders. In the Netherlands these regions are: 1. Randstad, the economic and cultural center of the Netherlands, which is also the core area for ongoing changes in the standard language; 2. Middle, an intermediate zone in the South of the province of Gelderland, along the borders of the Great Rivers; 3. North, a peripheral area in Groningen and the North of Drenthe; 4. South, a second peripheral area in Limburg. In Flanders the four main dialect areas were covered: 1. Brabant, i.e. the economic and cultural center of the northern part of Belgium, which is the core area for ongoing changes in Dutch spoken in Belgium; 2. East-Flanders, an intermediate zone; 3. West-Flanders, a peripheral zone in the west; 4. Limburg, a second peripheral area in the east. At the time of data collection, participants 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.

2.2 Speech Material

The participants were instructed about the aim of the research project: a study of standard Dutch pronunciation. Part of the questionnaire 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 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 participants 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 interviews took place at the speaker's workplace or home. 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 separate sound files. The total number of tokens in this study is 1280: 160 speakers x 4 variables x 2 realizations per variable.

2.3 Auditory Transcription

Auditory transcriptions of the voice characteristics were made of all tokens of (f), (v), (s), and (z). Three variants were distinguished: fully voiceless, partially voiced, and fully voiced. Two trained transcribers made consensus transcriptions for (v) and (z): one a native speaker of Dutch (the second author), one with a very limited knowledge of Dutch. In cases of disagreement a third trained transcriber was consulted (the third author). The realizations of (f) and (s) were only transcribed by one judge (the second author).

2.4 Acoustics Measurements

2.4.1 Labeling Procedure

All acoustic measures were done using *Praat* (Boersma and Weenink). For the segmentation of the speech signal *Signal Segmentor 1.01* was used, a program developed by Alain Soquet. The program displays for each sound file a wave signal, a spectrogram and a 512-point FFT spectrum (25 ms Hamming window) corresponding to the cursor position on the spectrogram and the wave signal. The program displays formants (computed with standard LPC-binomial) and F_0. The durations of voiced and devoiced fricatives appear to be almost identical if computed between the points where F1 transition starts and finishes (Stevens et al. 1992). 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. The 20% border was fixed visually on the basis of the spectrogram and the FFT, and verified by means of an auditory control. By convention the boundaries were placed at zero-crossings.

2.4.2 Duration of the Fricative (DUR)

The duration was computed from the beginning to the end point of the fricative. Voiceless fricatives tend to be longer than their voiced counterparts in Dutch (Slis and Cohen 1969, Slis and Van Heugten 1989, Debrock 1977). Consequently, DUR is expected to be longer for (v) and (z) in the regions showing strong devoicing of these variables. It is expected that DUR will be a good predictor for the auditory transcription.

2.4.3 F_0 Extent in the Fricative (F_0E)

The major cue for the voiced/voiceless distinction in Dutch seems to be the presence or absence of vocal cord vibration in the fricative (Slis and Cohen 1969, Van den Berg 1989). The F_0E value (in Hertz) was computed with intervals of 10 ms using the auto-correlation method. We employed the default procedure in *Praat*, which evaluates the presence of periodicity between 75 and 500 Hz. Finally, the number of samples with F_0E was divided by the total number of samples, resulting in a relative F_0E index, ranging from 0 (no pitch at all) to 100 (pitch present during the whole fricative). F_0E was expected to be the main predictor for the auditory transcription of voice.

2.4.4 Mean Intensity of the Frication Noise (MIN)

Slis and Cohen (1969) for stops and Van den Berg (1989) for fricatives (in a two-obstruent cluster) showed that the intensity of the frication noise is a perceptual cue for the voiced/voiceless distinction in Dutch. The intra-oral pressure causing the generation of supra-glottal noise depends on the pressure drop at the glottal level. In order to produce vocal fold vibration and noise generation across an oral constriction simultaneously, the cross-sectional areas at both constrictions have to be about equal. Consequently, if the cross-sectional area at the glottis constriction increases, the intra-oral pressure—and the amplitude at the supra-glottal turbulence source—will increase (Stevens et al. 1992). In other words, if vocal cords are adducted, the intra-oral pressure decreases. Hence, the intra-oral pressure is higher for voiceless fricatives, which are produced with a larger glottal opening than their voiced counterparts. Therefore, if the devoicing of voiced fricatives is associated with a glottal opening, as predicted by Haggard (1978), one should expect the intensity of the frication noise to be higher for (partially) devoiced fricatives than for voiced ones. From the beginning to the end point of the fricatives 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 (IF) = $10 \log_{10} [10^{I/10}/(1+10^{H/10})]$[1]. Finally, the sum of IF for each sample was divided by the total number of samples, resulting in a Mean Intensity of Noise (= MIN) value for each token.

3 Results of Auditory Transcription

In Figure 1 the results of the auditory transcriptions are presented split up by region (index scores from 0 (voiceless) to 100 (fully voiced)). The analyses of variance showed a substantial community effect for both (v) and (z) and a regional effect for (v) (Van de Velde and Van Hout 2001:233). There is more devoicing of (v) and (z) in the Netherlands than in Flanders. In each community, there are regional differences for (v). Especially in N-N devoicing of (v) is very strong. For a discussion of these results we refer to Van de Velde and Van Hout (2001). (f) and (s) are—as expected—voiceless in all regions.

---

[1]The intensity of the frication is assumed to be equivalent to the amount of the aperiodic energy in the signal. H = $10\log_{10}$(periodic energy / aperiodic energy) and I = $10\log_{10}$(periodic energy + aperiodic energy).

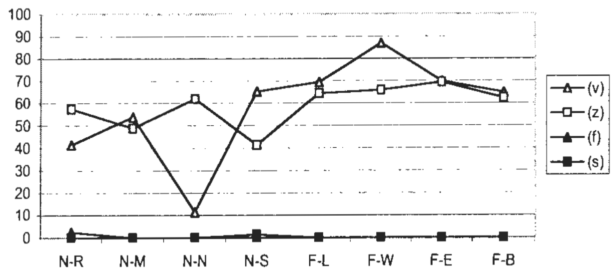

Figure 1: Auditory transcription of voice for (v), (z), (f) and (s) split up by region; 0 = voiceless, 100 = fully voiced

4 Evaluation and Refinement

4.1 Linear Discriminant Analysis (LDA)

A linear discriminant analysis was performed in order to predict the three auditory transcription categories (fully voiced, partially voiced or voiceless) using DUR, F_0E and MIN as predictors. A perfect match of actual auditory category and predicted auditory category was reached for 87% of the 1280 fricative realizations. The match was acceptable for 10% (a minimal difference of one category, e.g. fully voiced as partially voiced). In 3% of cases (N=39), the LDA resulted in a maximal mismatch (fully voiced as voiceless or vice versa). A closer examination of the mismatches revealed that seven of them were due to a coding error. In fifteen cases the F_0E value was unreliable because of a wrong detection of F_0 by *Praat*. However, seventeen mismatches between acoustic measurements and auditory coding remained unexplained. In order to remedy some of the mismatches, the calculation of F_0E and MIN were changed.

4.2 F_0E_{new}

Our data were recorded outside laboratory settings and the presence of some external periodic noise in the signal (e.g. due to a fan) cannot be excluded. If the fundamental frequency of such a noise was between 75 and 500 Hz, it was mistakenly analyzed as F_0 of the speech signal by *Praat*. To avoid this bias, the presence of periodicity was evaluated between two points: 1. the maximum F_0 during the silence preceding the utterance plus 10 Hz; 2. the maximum F_0 in the first third of the vowel plus 50 Hz. Nevertheless, eight tokens still contained an incorrect evaluation of F_0. These tokens were excluded from the database.

4.3 MIN$_{norm}$

In order to avoid a bias due to differences in recording quality, the mean noise intensity of the fricative was subtracted from the maximum intensity of the first third of the following vowel. This "normalized mean intensity of frication noise" (henceforth MIN_{norm}) measure is expected to be a better predictor of the auditory transcription.

4.4 Second LDA

A new LDA was performed: the predicted values were the three auditory transcription categories of the fricative realizations, and the predictors were DUR, F_0E_{new} and MIN_{norm}. There is a clear improvement: in 89.3% of the cases there is a perfect match and an acceptable one in 10.1%. The number of mismatches is reduced to 0.6%. There is a good match between auditory transcriptions and the acoustic correlates of devoicing we selected. Nevertheless, it is clear that additional acoustic measurements could perhaps improve the description of the process of devoicing. Four additional acoustic measurements were applied.

4.5 Additional Measurements

The duration of the vowel following the fricative (DUR_{vowel}) and the duration of the syllable ($DUR_{\surd}$) were measured. On the one hand, it is expected that syllable durations are almost identical for voiced and voiceless fricatives. On the other hand, vowels will be longer after phonologically voiced fricatives than after voiceless ones (Slis and Cohen 1969, Slis 1970, Stevens et al. 1992). Therefore, the devoicing of (v) and (z) should result in a shortening of DUR_{vowel}, while $DUR_{\surd}$ should remain constant.

Vowels following voiced fricatives show an F_0 contour that increases slowly, starting from low frequencies. Vowels following voiceless fricatives show a decreasing F_0 contour, starting from high frequencies. Moreover, the F_0 in the onset of the vowel following a voiceless fricative is higher than the F_0 in the vowel following a voiced fricative (Slis and Cohen 1969). The F_0 was computed at the first voiced period of the vowel and 30 ms later. Their mean (F_0V_{mean}) reflects the mean F_0 in the vowel onset. The difference between the last and the first value gives an indication of the pattern of the F_0

in the onset of the vowel (F_0V_{dif}): positive values represent an increasing F_0, negative values a decreasing F_0 and values near zero a steady F_0 pattern. These acoustic correlates are traditionally related to the tenseness of the vocal folds. Since /v/ and /z/ have often been described as tense, independently of their voicing (Debrock 1977, Slis 1970, Slis and Cohen 1969), it is expected that the devoicing of (v) and (z) does not yield important changes in F_0V_{mean} and F_0V_{dif}.

5 Devoicing of (v) in the Randstad Region (N-R)

In this section we will show how a careful quantitative analysis can improve the description of the phonetic facts. We focus on the devoicing of (v) in the Randstad region (N-R).

In N-R (v) is strongly devoiced (cf. Figure 1), which is confirmed by the low average F_0E_{new} in this region (cf. Figure 2). However, a surprising observation concerns the MIN_{norm} in N-R. As can be seen in Figures 1 and 2, the devoicing of (v) is stronger in N-R than in N-M. If the predictions of Stevens et al. (1992) were true, the intensity of the frication noise would be higher in N-R than in N-M. However, Figure 3 shows this not to be the case. The MIN_{norm} of (v) in N-M is higher than in N-R, despite a higher F_0E_{new}. A possible interpretation might be that in the latter region, the vocal cords remain adducted during the devoiced portions of [v], impeding the building up of intra-oral pressure. Thus, in N-R the devoicing of /v/ would not be consistently accompanied by a glottal opening, as predicted by Haggard (1978). In fact, it would confirm Abramson's (1967) claim that voiced fricatives may occur with either a closed or an open glottis, independently of the presence or absence of glottal pulsing.

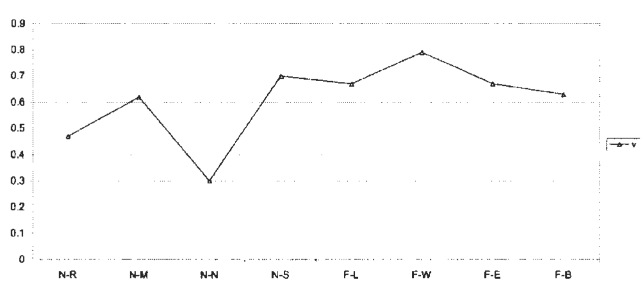

Figure 2: F_0E_{new} for (v) split up by region

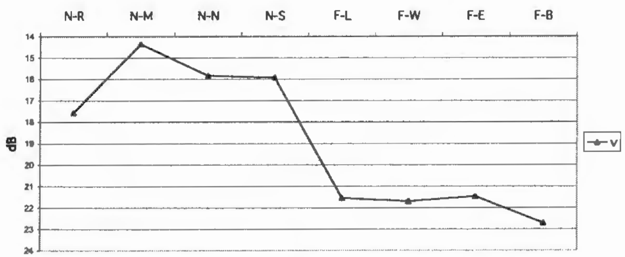

Figure 3: MIN_{norm} for (v) split up by region

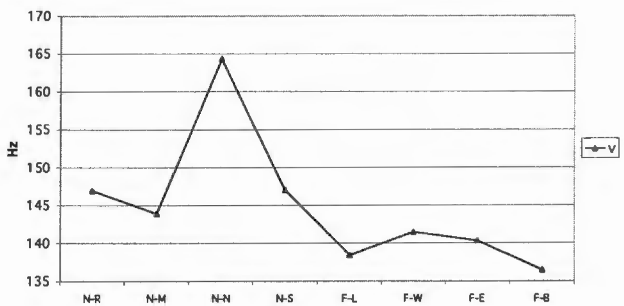

Figure 4: F_0V_{mean} for (v) split up by region

If we assume that the glottis remains closed during some of the devoiced parts of (v), we still have to explain the absence of glottal pulsing. One explanation is that the transversal adduction of vocal folds is associated with a longitudinal tension. Some supportive evidence was found. There is a significant positive correlation in N-R between F_0V_{dif} and F_0E_{new} for (v) (*r=.501, p<.01*). This means that vocal fold vibrations start with a higher F_0 in vowels that follow devoiced realizations of (v). A higher F_0 at the onset of the fricative reflects the stiffness of the vocal folds, which impedes glottal pulsing during the frication. On the other hand, there is no significant correlation for (v) between MVF_0 and F_0E_{new} in N-R. MVF_0 stays low in N-R despite the strong devoicing of (v), as can be seen in Figure 4. Moreover, MVF_0 and SF_0V do not correlate at a significant level in N-R.

These results have interesting phonetic implications. First, since (v) in N-R is realized with both tense and adducted vocal folds, it is not possible to describe the /v/-/f/ contrast in traditional "tense/lax" terms.

Second, low F_0 after [v] is not an automatic effect of the transversal glottal configuration, as predicted by Slis (1970), Slis and Cohen (1969), and Goldstein and Browman (1996). It can be analyzed as a "listener-oriented" optimality strategy, which counterbalances low voicing extent in order to maintain an auditory contrast between /v/ and /f/. Kingston and Diehl (1994) argue that F_0 depression next to [+voice] stops is the product of an "independently controlled articulation, whose purpose is to enhance the [voice] contrast." An identical effect might be at work here. In line with this interpretation, data provided by Collier et al. (1979) shows that activity patterns of the vocalis differ between voiced and voiceless Dutch fricatives only after the release of the supra-glottal constriction. Furthermore, these authors found that vocalis activity is identical to the activity of the cricothyroid muscle, which can be held responsible for the F_0 changes in the vowel following a consonant.

Another surprising result concerns durational patterns of (v) in N-R. In line with the literature, it has been predicted that the devoicing of phonologically voiced fricatives should lead to an increase of DUR (see Sections 2.4.2 and 4.5). However, as shown in Figure 5, the difference between the average duration of (f) and (v) is higher in N-R than in N-M, and the average DUR of (v) is shorter in N-R than in N-M. This might be a consequence of the aerodynamic difficulty to maintain sufficient pressure drop at the labio-dental constriction when adducted and tensed vocal folds allow only a small amount of airflow through the glottis. This articulatory constraint might enhance the /v/-/f/ distinction as well.

Furthermore, for (f) $DUR_{\surd}$ is almost equal in N-R and N-M, but for (v) $DUR_{\surd}$ is shorter in N-R than in N-M (Figure 6). This implies that the shortening of (v) in N-R is not compensated by the lengthening of the following vowel, as predicted by the literature. This exceptional pattern is confirmed by the correlation between fricative duration and syllable duration for (v): in comparison with the overall correlation for all regions together ($r=.788$, $p<.01$) it is strong in N-R ($r=.886$, $p<.01$). Moreover, DUR correlates in N-R stronger with $DUR_{\surd}$ ($r=.917$, $p<.01$) than with DUR_{vowel} ($r=.627$, $p<.01$). The shortening of the DUR for (v) in N-R might be due to an articulatory constraint that the combination of longitudinal tension and transversal adduction imposes on the duration of the frication period. However, the fact that this shortening is not compensated by vowel lengthening suggests that the duration of /v/ enhances the /v/-/f/ distinction in N-R.

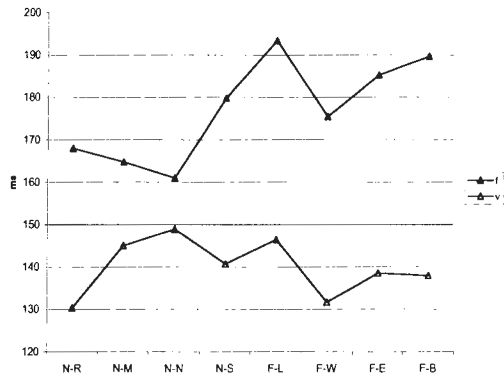

Figure 5: DUR for (v) and (f) split up by region

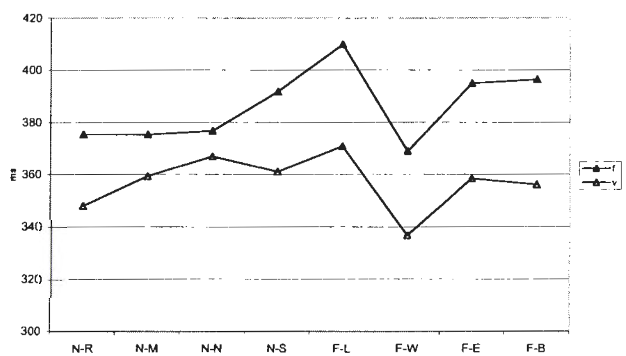

Figure 6: DUR_V for (v) and (f) split up by region

6 Conclusion

We were able to develop acoustic measurements that uncover the core voice characteristics of the labio-dental and alveolar fricatives in Dutch. Auditory transcriptions played a crucial role in the validation process, but the phonetic analyses resulted in a better and more detailed description of consonantal variation, providing more insight into the regional patterns of variation in standard Dutch. The acoustic correlates of the /v/-/f/ distinction in Dutch, such as the extent of glottal pulsing and the intensity of the frication noise, vary regionally. Quantitative sociolinguistic methods, applied to the devoicing of /v/, confirmed that phonological distinctions cannot be described either in terms of sets of features or by referring to production constraints (Kingston and Diehl 1994, 1995). Rather, auditory contrast between /v/ and /f/ in Dutch seems to be achieved through independent sets of controlled articulatory gestures. We can conclude that acoustic techniques can be used for the sociolinguistic study of consonantal variation. Furthermore, the wide-scale application of these techniques to speech collected outside the phonetics lab, taking into account sociolinguistic variation, may contribute to phonetic sciences as well.

References

Abramson, Arthur. 1967. On matching the speech signal to mechanisms of production. *Haskins Laboratories. Speech Research* 10:13-48.

Boersma, Paul, and David Weenink. See *www.praat.org*.

Collier, René, Leigh Lisker, Hajime Hirose, and Tatsujiro Ushijima. 1979. Voicing in intervocalis stops and fricatives in Dutch. *Journal of Phonetics* 7:357-373.

Debrock, Mark. 1977. An acoustic correlate of the force of articulation. *Journal of Phonetics* 5:61-80.

Foulkes, Paul. 2002. Current trends in British socio-phonetics. *Penn Working Papers in Linguistics 8.2: Selected papers from NWAV 30.*

Goldstein, Louis, and Catherine P. Browman 1986. Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics* 14:339-342.

Haggard, Mark. 1978. The devoicing of voiced fricatives. *Journal of Phonetics* 6:95-102.

Kingston, John, and Randy Diehl. 1994. Phonetic knowledge. *Language* 70(3):419-454.

Kingston, John, and Randy Diehl. 1995. Intermediate proprieties in the perception of distinctive feature values. In *Phonology and phonetic evidence: papers in laboratory phonology*, ed. Bruce Connell and Amalia Arvaniti, 7-27. Cambridge: Cambridge University Press.

Kissine, Mikhail, Hans Van de Velde, and Roeland Van Hout. 2003. An acoustic study of standard Dutch /v/, /f/, /s/ and /z/. In *Linguistics in the Netherlands 2003*, ed. Leonie Cornips and Paula Fikkert, 93-104. Amsterdam: John Benjamins.

Slis, Iman H. 1970. Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. *Phonetica* 21:193-210.

Slis, Iman H., and Antonie Cohen. 1969. On the complex regulating the voiced-voiceless distinction I & II. *Language and Speech* 12(2) & (3):80-112; 137-155.

Slis, Iman H., and M. van Heugten. 1989. Voiced-voiceless distinction in Dutch fricatives. *Linguistics in the Netherlands* 6:123-132.

Stevens, Kenneth N., Sheila E. Blumstein, Laura Glicksman, Martha Bruton, and Kathleen Kurowski. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal Acoust. Soc. Am.* 91:2979-3000.

Thomas, Eric K. 2002. Instrumental Phonetics. In *The handbook of language variation and* change, ed. Jack Chambers, Peter Trudgill and Nathalie Schilling-Estes, 168-200. Oxford: Blackwell.

Van de Velde, Hans, Marinel Gerritsen, and Roeland 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, Hans, and Roeland van Hout. 2001. The Devoicing of Fricatives in a Reading Task. In *Linguistics in the Netherlands 2001*, ed. Ton van der Wouden and Hans Broekhuis, 225-236. Amsterdam: John Benjamins.

Van den Berg, Rob. 1989. Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues. *Speech Communication* 8:17-25.

Van Hout, Roeland, Georges De Schutter, Erika De Crom, Wendy Huinck, Hanne Kloots, and Hans Van de Velde. 1999. De uitspraak van het Standaard-Nederlands: variatie en varianten in Vlaanderen en Nederland. In *Artikelen van de Derde Sociolinguïstische Conferentie*, ed. Erica Huls and Bert Weltens 183-196. Delft: Uitgeverij Eburon.

Mikhail Kissine
mk466@cam.ac.uk

Hans Van de Velde
UiL-OTS, Utrecht University
Trans 10, 3512 JK Utrecht
The Netherlands
hans.vandevelde@let.uu.nl

Roeland van Hout
CLS, Radboud University Nijmegen
P.O. Box 9103, 6500 HD Nijmegen
The Netherlands
r.v.hout@let.kun.nl