COMPENSATORY STRATEGIES FOR VOICING OF INITIAL AND MEDIAL PLOSIVES AND FRICATIVES IN WHISPERED SPEECH IN DUTCH

D. J. van de Velde & V. J. J. P. van Heuven

Leiden University Centre for Linguistics, the Netherlands
d.j.van.de.velde@hum.leidenuniv.nl; v.j.j.p.van.heuven@hum.leidenuniv.nl

ABSTRACT

In this study, the compensatory strategies speakers adopt for conveying the voiced-voiceless distinction of consonants in whispered speech were investigated. Of 8 native Dutch speaking subjects each, 26 Dutch minimal word pairs containing voiced and voiceless obstruents, were recorded. Acoustic analysis with the software *Praat* revealed that both in whisper and in normal speech (1) preceding vowels are longer for voiced obstruents, (2) duration of the silent interval and (3) of the burst are longer in voiceless obstruents and (4) the burst intensity in voiceless obstruents is greater.

These recordings were subjected to a pool of Dutch listeners (N=16), judging for each of 1056 items if it was the voiced or the voiceless member of the minimal pair. This revealed no marked sensitivity of these contrasts in whispered speech. These results supports the Redundant Cue Hypothesis (RCH), stating that voicing perception in whisper depends mainly on the secondary voicing cues that remain in whisper.

Keywords: compensatory strategies, voicing, Dutch plosives and fricatives, whispered speech

1. INTRODUCTION

In whispered speech vocal fold vibration is absent. During this type of phonation, the glottis is closed with the exception of a small triangular opening between the arytenoids cartilages (whispering triangle), the same shape as during the production of voiceless [h] [6]. Therefore, the distinction between voiced and voiceless sounds cannot be conveyed by presence versus absence of peri-odicity (fundamental frequency) in the glottal source signal. Nevertheless, there are indications that listeners are able to discriminate between the voiced and voiceless members of a voicing contrast when exposed to whisper. There are two possible mechanisms that can be hypothesized as possible explanations for this ability on the part of the listener. First, it may be the case that the listener simply relies on concomitant, secondary cues that accompany the voicing contrast in normally phonated speech. We know, for example that voiced consonants have shorter closure duration, shorter or even negative voice onset time (VOT), shorter and low-intensity noise bursts, are preceded by longer vowels, with slower moving formant transitions and rise/ decay times into the surrounding vowels than voiceless obstruents, e.g. [3, 8, 9]. If the speaker maintains these secondary cues in whisper, the listener can use these in order to resolve the contrast even though there is no periodicity in the waveform. We will refer to this possibility as the redundant cue hypothesis (RCH). Second, the speaker may be subconsciously aware of the fact that whispered speech lacks periodicity, which may compromise the identification of the voicing feature. In order to remain intelligible, the speaker may apply a compensatory strategy by which he amplifies the normally redundant cues. In the latter case, which we will call the compensatory cue hypothesis (CCH), we may expect acoustic contrasts in the concomitant cues to be more clearly marked in whispered than in phonated speech. This would be in line with the Hyper&Hypo theory advanced by [7].

Although focus has been on vowels, a number of studies has centered on the production of whispered consonants. Jovicic and Saric [4] recorded nonsense syllables of the form /aCa/, contained in a carrier sentence, for 25 Serbian consonants from 6 speakers both in normal and whispered speech. They found that consonants were 10 percent longer in whispered than in normal speech, but the lengthening was smaller for unvoiced (5.8%) than for voiced (15.3%) ones. This lengthening was greater in sentence-initial and final than in medial position. There was no difference in VOT for voiceless plosives and affricates between the two modes of speaking, but

there was for voiced ones. Whispered consonant were on average 12 dB lower in intensity, but voiceless ones more so (maximally 3 dB attenuation) than the voiced (up to 25 dB attenuation). Jovicic and Saric concluded that whisperers maintain a high control of prosodic feature production in order to be intelligible despite lack of periodicity and lower overall intensity.

The present research tries to answer some of these questions concerning Dutch. Whereas previous research on the voicing distinction in whispered speech has been focusing on the quantitative question of how much certain acoustic features changed in comparison to normal phonation, this study at the same the qualitative question: which of these changes are relevant to listeners, i.e., which acoustic changes can be discerned as a voicing distinction by listeners.

2. METHOD

2.1. Subjects

Speakers (N=8; 4 female) were native Dutch speaking students of Leiden University between 19 and 30 years of age (of which 7 younger than 22). Apart from one speaker with a slight Brabant accent and another with a mild Rotterdam accent, they showed no marked dialectal accent and reported no articulatory problems. They were not paid. All listeners (N=16) were also native speakers of Dutch.

2.2 Stimulus materials

The Dutch consonant inventory consists of 17 consonants [2] along the dimensions of articulation place (labial/labiodental, alveolar, velar/uvular), articulation manner (plosive, fricative, nasal, liquid (and glide)) and voice (voiced, voiceless). Consonants can be either word initial, medial or final. The voicing distinction for obstruents occurs only word initially and medially. The opposition is neutralized in word-final (or even syllable-final) position such that only the unmarked (i.e. voiceless) member remains. The materials consisted of 26 Dutch minimal pairs from the standard Dutch lexicon. Low-frequency words were avoided. The target consonant differed along three binary factors: word position (initial or medial), articulation place (labial or alveolar) and articulation manner (plosive or fricative). This yielded the following set of distinctions: /t/-/d/, /p/-/b/, /s/-/z/ and /f/-/v/. Words were either monosyllabic or disyllabic. There was no restriction on the vowels surrounding the consonant but the target obstruent was never part of a cluster of consonants.

For the listening sessions, all items of all speakers were presented in isolated form, in a random order (the same order for all listeners). No resynthesis of the waveforms was performed.

2.3. Procedure

Recordings were made with professional audio equipment. Speakers were seated alone in a sound-attentuating recording booth. The subject's voice was recorded through a Sennheiser MKH-416 unidirectional condenser microphone directly onto a PC (22,050 Hz, 16 bit). Subjects were instructed to pronounce with a calm pace a list of words printed on a sheet before them, first in normal speech and then whispered. All subjects had the exact same list of words. Recordings were normalized for individual speaker volume.

In listening sessions (lasting approximately 50 minutes) the sound files were played back to listeners in small groups, using standard equipment. Subjects heard each of the 1056 items (66 items × 8 speakers times × 2 modes) and had to indicate for each them if they heard the unvoiced or the voiced variant. This was done by ticking the preferred option on a multiple-choice answer sheet containing only the relevant voiced and voiceless option for each item printed in normal Dutch orthography.

2.4. Analysis

Acoustic analysis was done with the *Praat* [5] speech processing software. We measured the duration of the occlusion and the noise burst of the plosives, the total duration of the plosives and fricatives, the duration of the preceding vowel (for medial consonants) and the mean intensity (in dB) of the plosive noise burst or friction portion.

3. RESULTS

Results of the duration measurements are summarized in Figure 1.

The results show that the voiced and voiceless members of the pairwise contrasts are acoustically distinct in the duration of the noise burst and/or the (inversely related) preceding vowel duration. It is not immediately clear from these results if the voiced-voiceless contrast is more clearly marked in whisper than in phonated speech, at least when we

limit the comparison to only the parameters that are shared in both modes.

Figure 1: stacked durations of (from left to right) the preceding vowel, the occlusion (or prevoicing) and the noise burst for medial obstruents in normal (panels A-B)) and whispered (panels C-D) mode of speech, separately for onset (panels A, C) and medial (panels B, D) positions.

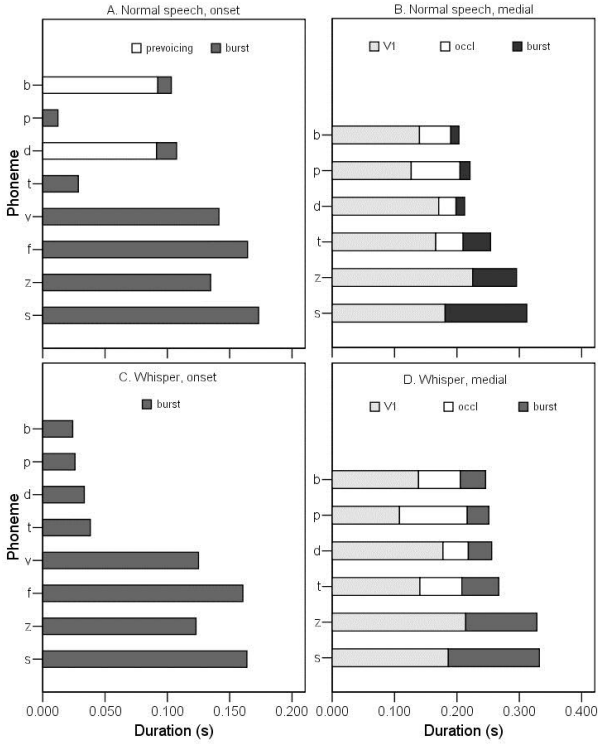

Figure 2 presents the intensity measurements. The intensity measurements show no reliable differentiation between voiced and voiceless counterparts in normal speech. In whisper, however, the voiceless member always has greater intensity than the voiced counterpart. Also, the range of intensities is restricted in the normally spoken items (between 70 and 75 dB) but intensities are more widely dispersed in whisper (between 60 and 75 dB).

In order to test the hypothesis that speakers compensate for the lack of periodicity-related cues in whisper, we ran Linear Discriminant Analyses [5] using the acoustic parameters (z-normalised within speakers, so as to eliminate speaker-individual differences) shared by phonated and whispered speech as predictors of voicing. If the CCH is true, the voicing feature should be classified correctly more often in whisper than in normal speech. In one LDA the durations of noise bursts of plosives and fricatives were used as one predictor (leaving the duration of the silent interval, which does not exist in fricatives) out of consideration. In a second LDA, the duration of silent interval in plosives was added to the burst duration, so that total consonant duration could be used as a predictor. The results are as in table 1.

Figure 2: Mean intensity (dB) of noise bursts in normally spoken (left panel) and whispered (right panel) voiced and voiceless obstruents. Onset and medial obstruents have been averaged.

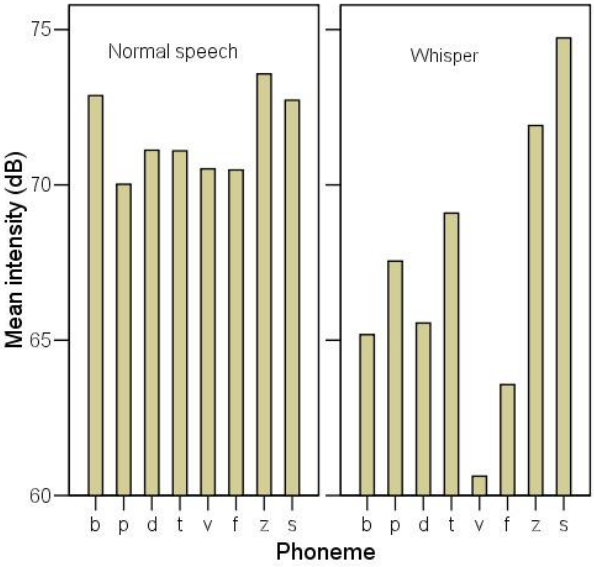

Table 1: Results of LDA. Percent correctly classified voicing for obstruents in onset and medial positions, in normal speech and in whisper. Correct classification in bold face.

	actual	% classified as		total % correct
		−voice	+voice	
onset	predictors: burst dur, intensity			
normal	−voice	**41**	59	47
	+voice	47	**53**	
whisper	−voice	**37**	63	53
	+voice	28	**72**	
medial	predictors: V1 dur, burst dur, intensity			
normal	−voice	**63**	37	76
	+voice	14	**86**	
whisper	−voice	**57**	43	66
	+voice	28	**72**	
	predictors: V1 dur, cons dur, intensity			
normal	−voice	**71**	29	84
	+voice	6	**94**	
whisper	−voice	**64**	36	77
	+voice	14	**86**	

Table 1 shows mixed results. Whispered obstruents are somewhat better classified for voicing than their normally spoken counterparts at the onset of utterances (53% versus 47% correct). However, in medial position voicing classification

is consistently better – by 7 to 10 percentage points – in normal speech than in whisper.

4. CONCLUSION

The results of the present experiments show that Dutch listeners are able to discriminate voiced from voiceless obstruents, not only in normally phonated speech but also in whisper, where periodicity is not available as a voice cue. However, the percentage of correct classification of voicing by human listeners was lower in whisper than in normal speech, which indicated that the lack of the periodicity cue is not fully compensated for in whisper.

Moreover, although the acoustic measurements revealed clear differences between voiced and voiceless counterparts in terms of preceding vowel duration, duration of silent interval, burst duration and burst intensity, both in normal speech and in whisper, we found no indications that the contrast was more clearly marked in whisper by non-periodicity-related parameters that are shared by normal speech and whisper.

We may conclude, therefore, that our speakers did not compensate (fully) for the lack of periodicity cues in whisper, so that the Compensatory Cue Hypothesis (CCH) is not supported. In fact, the poorer overall scores obtained in human perception and in automatic classification of voicing, suggest that voicing perception in whisper depends mainly on the secondary voicing cues that remain in whisper, thus lending credibility to the Redundant Cue Hypothesis RCH.

5. REFERENCES

[1] Boersma, P., Weenink, D. 1996. Praat: Doing phonetics by computer. *Report nr. 132, Inst. Phonetics* Un. Amsterdam.

[2] Booij, G.E. 1995. *The Phonology of Dutch*. Oxford: Clarendon.

[3] Debrock, M. 1977. An acoustic correlate of the force of articulation. *J. Phonetics* 5, 61-80.

[4] Jovicic, S., Saric, Z. 2008. Acoustic analysis of consonants in whispered speech. *J. of Voice* 22, 263-274.

[5] Klecka., W.R. 1980. *Discriminant Analysis*. Newbury Park CA: Sage publication.

[6] Laver, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

[7] Lindblom, B.E.F. 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle,W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer, 403-439.

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

[9] Slis, I.H., Cohen, A. 1969. *On the Complex Regulating the Voiced-voiceless Distinction I & II.. Language and Speech 12*, 80-102; 137-155.