Perception Of Final Consonant "Voicing" In Phonated And Whispered Speech

by

Yana D. Gilichinskaya

A dissertation submitted to the Graduate Faculty in Speech–Language–Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy
The City University of New York

2012

This manuscript has been read and accepted for the Graduate Faculty in Speech–Language–Hearing Sciences in satisfaction of the Dissertation requirement for the degree of Doctor of Philosophy

Winifred Strange, Ph.D. January 18, 2012

Chair of the Examining Committee

Glenis Long, Ph.D. January 18, 2012

Co-chair of the Examining Committee

Klara Marton, Ph.D. January 18, 2012

Executive Officer

Supervisory Committee:

Brett Martin, Ph.D.

Douglas Whalen, Ph.D.

The City University of New York

Abstract

Perception of Final Consonant "Voicing" in Phonated and Whispered Speech

by

Yana Gilichinskaya

Advisers: Dr. Winifred Strange, Dr. Glenis Long.

Our everyday verbal interaction does not always happen in the communication conditions optimal for the speaker or the listener. In order to increase intelligibility and to overcome interference of the communication medium, speakers may switch from a conversational rate of speaking to "clear" speech mode. In other situations, when fully phonated speech is not appropriate, speakers may deliberately depart from the optimal listening conditions to adopt yet another speaking mode — whispered speech. Despite acoustic differences between phonated and whispered speech, intelligibility of whispered speech remains relatively accurate. In perception of whispered consonants, the characteristic that is most susceptible to the absence of phonation is "voicing". At the same time, listeners identify "voicing" in whispered consonants far above chance. We hypothesized that perceptual discernment of whispered consonant "voicing" may be rooted in the maintenance and potential enhancement of time-based parameters that are known to contribute to consonant voicing: vowel duration, consonant duration and combined consonant-to-vowel duration-ratio. Acoustic parameters were compared between phonated and whispered conversational speech, and conversational and clear whispered speech. The results indicated that all time-based acoustic cues to "voicing" were maintained in whispered speech. In perception tests, overall consonant identification accuracy in conversational whispered speech was about

85% with errors on place or manner of articulation not exceeding 1%. The improvement in perception of clear whispered speech varied across speakers from -2% ... to 8%. The perceptual advantage afforded by two speakers was generally in the agreement with the predictions made on the basis of the production data. Logistic regression analysis indicated that after controlling for variability in speakers, consonant-pairs and vowels, consonant "voicedness" in both phonated and whispered speech could be accurately predicted by a single parameter — C/V duration-ratio. At the same time, correlation analysis between the magnitude of C/V duration-ratio contrast and consonant "voicing" intelligibility suggested that listeners did not fully utilize the distinctive information provided by the C/V duration-ratio. This study contributes to knowledge of factors responsible for intelligibility of whispered speech which may be valuable in designing speech processing algorithms for hearing-assistive devices and developing synthesis systems for speech reconstruction in voice-impaired patients.

Acknowledgment

This dissertation would not have been possible without the enthusiastic guidance of my supervisor, Dr. Winifred Strange. Throughout my graduate work, she provided support, patience and a lot of great ideas. It is her teaching that helped me think logically and has always made me want to learn more. I particularly appreciate her taking time from her retirement and coming to New York for my dissertation defense.

I owe my deepest gratitude to Dr. Glenis Long for her continuous and invaluable feedback throughout the dissertation process, for her general encouragement, support and occasional psychotherapeutic sessions. I would like to thank Dr. Brett Martin for always having me gear my research thinking towards clinical application. Additional thanks go Dr. Douglas H. Whalen and Dr. Klara Marton for their insightful comments and feedback on the dissertation.

I am grateful to Dr. James Jenkins for his valuable advice on the use of various statistical methods and tips to effective visual presentation, to Dr. Valerie Shafer for the immensely informative impromptu ERP-learning sessions. To Gary Chant for always being available to help with calibration and — even more so — with re-calibration of the sound equipment. Thank you, Loretta Walker and Linda Ashour for always being on my side in the fight against paperwork!

In my everyday Ph.D.-student life, I have been lucky to work (and play) with wonderful colleagues and friends. I am thankful to Franzo Law II for his help with editing and his moral support throughout all these years via countless alcoholic beverages. To Shari Berkowitz for being an inspirational source of creative ideas inside and outside of academia: the sheep throat cutting practicum, knitting, bread-making, theater-going, salsa dancing and many many more unforgettable activities. To Simon Henin for organizing a MATLAB workshop three years ago and making himself available for my bombarding him with questions. To Sue Thompson for

her enthusiastic support at the final stages of the dissertation and very helpful feedback on the presentation. To Puisan Wong for her encouraging e-mails that helped me see the light at the end of the dissertation tunnel. To Carol Tessel for making me balance my social and asocial lives. To Jason Rosas for entertaining and useful research discussions in the lab, and for teaching me socially important bits of colloquial Spanish. To Valeriy Shafiro for getting me involved in experimental research very early on in my graduate studies and for creating original MATLAB codes that later became the basis of my own acoustic analysis programs. To my fellow graduate and ex-graduate students, Miwako Hisagi, Kikuyo Ito, Yan Yu, Rebekkah Bucherri, Mieko Sperbeck and all of the Speech–Language–Hearing sympathizers for their camaraderie and support.

I would like to thank my friends for their palpable presence, entertainment and stress-relieving abilities. Elena Rykhlevskaia (Ry) for her longtime friendship, remote pep-up talks, and unsafe mountain descents and Zhenya Yubliler for all the hikes, drives, ski trips and her unsurpassed desire to start a fire in the woods. My friends here, Jamease Miles, Maria Obraztsova, Valeriy Grdzelishvili, and my friends back in Moscow, Anna Mukhina and Elena Solovieva, my brother Michael Gilichinsky, for supportive chats, good humor and advice.

I want to express my eternal love and gratitude to my grandparents, Pasha Rivkina, Charna Shoirif and Abram Gilichinsky. To my *dedushka* Mendel Rivkin who seventy years ago fought to save this world from terror and insanity. To him I dedicate this work.

Last and most important, I thank my parents, Elizaveta Rivkina and David Gilichinsky, for making me curious about things in the first place, for their love, care and support pre-, during and post-dissertation.

To my grandfather

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Statement of the problem

Everyday speech communication does not always occur in optimal conditions for the speaker or the listener. When speaking with a hearing-impaired person, a second-language learner, or a machine, many but not all speakers can improve acoustic signals by switching from the conversational speech mode to "clear speech". This increases speech intelligibility so that listeners can overcome recognition errors. In other situations, when fully phonated speech is not appropriate or the transmitted information is to remain private, speakers may deliberately depart from normal conversational or optimal speaking modes while striving to maintain a certain level of speech intelligibility by adopting yet another speaking mode — whispered speech.

Whispered speech may be the only speaking mode available to aphonic patients and might serve as a target (Hirahara et al., 2009) or a source speech mode (Tran et al., 2010, Morris & Clements, 2002) for speech reconstruction in voice-impaired patients. Unlike phonated speech, whispered speech is produced in the absence of vocal fold vibration. As vocal folds are held constantly apart, the whispered speech signal lacks periodicity and the vocal tract is excited solely by noise. As a result, the overall intensity of the resultant signal is greatly reduced compared to regular speech (Ito et al., 2005, Jovicic & Saric, 2006) but the *relative* intensity of conso-

nants to vowels is greatly enhanced. In addition, changes in the articulatory configuration of the larynx affect resonant characteristics of the vocal tract (Tsunoda et al., 1994). Consequently, the whispered speech signal is very different acoustically from phonated speech. Despite these differences, however, whispered speech remains intelligible: listeners identify whispered vowels with accuracy ranging from 65% (Kallail & Emanuel, 1985) to 85% (Tartter, 1991) — a 15-20% decrease relative to their phonated counterparts (Kallail & Emanuel, 1984b,a, 1985, Tartter, 1991); consonants are identified with 64% accuracy (Tartter, 1989, Munro, 1990). Relative intelligibility of whispered speech may be rooted in increased durations of the phonetic segments (Schwartz, 1972, Parnell et al., 1977, Jovicic & Saric, 2006) and naturally enhanced consonant/vowel amplitude-ratio. Exploring whispered speech as an acoustically impoverished yet intelligible speech signal may improve speech processing algorithms of hearing aids and cochlear implants and, thus, increase overall intelligibility of speech perceived through hearing assistive devices.

The goal of the present project was to explore the most consistent acoustic cues underlying "voicing" [1]contrasts in whispered speech and to investigate perception of consonant "voicing" in whispered speech. The majority of the whispered consonant identification errors is caused by incorrect perception of consonant "voicing" (72% accuracy); manner and the place of articulation cause less confusion — 86% and 91%, respectively (Tartter, 1989). In other words, although the dominant acoustic cue to voicing in phonated speech — periodicity in voiced consonants — is not present in whispered speech, listeners are able to perform above chance when asked to categorize "voiced"/"voiceless" consonants based on some secondary cues. Differences in the production patterns between whispered "voiced" and "voiceless" consonants at the level of the glottis (Weismer & Longstreth, 1980, Tsunoda et al., 1994)); false vocal folds (Tsunoda et al., 1994) and the lips (Higashikawa et al., 2003) suggest that the "voicing" contrast in whispered consonants may be maintained through relative duration and intensity.

[1]In the rest of the paper, "voiced" or "voiceless" is referred to the intended voicing property of the consonant independently of its acoustic characteristics

The current project is divided into two major parts, a production study and a perception study. Specific objectives of the production study were, first, to describe quantitatively acoustic cues to consonant "voicing" in phonated and whispered speech produced at conversational rates and to compare the differences in these cues between the two phonation modes. The second objective was to investigate changes in the "voicing" cues between whispered conversational and whispered clear speech. Acoustical analysis of both phonated and whispered speech samples was performed to obtain the necessary measurements. The perception study was concerned with whispered consonant identification and the effect of clear whispering on consonant identification accuracy. The first objective was to investigate the frequency and the types of errors made by listeners in perception of whispered consonants. The second objective was to explore changes in perception of consonants associated with clear whispered speaking style and how performance varies with individual speakers for different kinds of consonant-pairs: stops, fricatives and affricates. Finally, the third objective was to investigate relative intelligibility of different speakers and consonant-pairs in connection with those acoustic cues that were consistently present in whispered speech.

1.2 Production of whispered consonants

In phonated speech, adduction and abduction of the vocal folds underlie acoustic differences between voiced and voiceless consonants. In whispered speech, although the glottis is open, the vocal folds are adducted enough to create turbulence (Laver, 1994) and contrasting patterns in the production of whispered "voicing" cognates exist even at that level of the larynx. More specifically, production of whispered "voiced" and "voiceless" consonants is associated with different levels of activity in posterior cricoarytenoid (PCA), the sole abductor of the glottis. A relative decrease in PCA activity is recorded for "voiced" consonants /d/ and /z/ versus a relative increase in PCA activity for "voiceless" /t/ and /s/. At the same time, PCA's baseline activity in whispered speech is higher relative to phonated speech (Tsunoda et al., 1994). At the level of the false vocal

folds, there is increased activity of the thyropharyngeus muscle contributing to their adduction during whispering and, consequently, leading to a smaller area of supra-laryngeal aperture than in phonated speech (Tsunoda et al., 1994). Differences in the production of "voiced" versus "voiceless" whispered consonants in a kinematic study by Higashikawa et al (2003) revealed differences of lip motion. A video-based computerized tracking system was used to analyze lip movement by measuring displacement and velocity of the lips during oral opening and closing. While there was no difference between peak velocity of oral closing for phonated /b/ and /p/, the difference was significant for whispered productions, with greater velocities associated with /b/. Similarly, no significant differences in peak opening were found for phonated /b/ and /p/ phonemes while whispered /b/ had greater opening than /p/. In addition, /b/-/p/ differences in peak opening displacement were greater in whispered productions than they were in phonated ones. Finally, in whispered speech, there was a larger difference in the maximum lip separation for /b/ and /p/ (greater values for /b/ in whispered than in phonated stimuli). In summary, despite the absence of vocal fold vibration – an articulation gesture that forms the basis for the distinction between voiced and voiceless consonants in phonated speech — production of whispered "voiced" and "voiceless" consonants is associated with distinct articulation patterns at the level of the glottis, false vocal folds, and the lips (for labial stops). In the next section, the acoustic cues to consonant "voicing" that might be available to listeners in whispered speech as a result of these articulation differences is discussed.

1.3 Acoustics of whispered speech

The hallmark of whispered speech production is the absence of periodic vibration of the vocal folds. This entails certain articulatory adjustments along the vocal tract and determines specific acoustic properties of whispered speech. The open glottis provides weak coupling of the larynx with the subglottal space which is reflected acoustically by reduced energy in the first formant region. Constriction at the level of the false vocal folds decreases the size of the laryngeal cavity

and raises formant center frequencies of the vowels (Tsunoda et al., 1994). At the same time, the bandwidths of the formants are increased (Jovicic, 1998, Li & Xu, 2005). The absence of glottal pulses leads to a drop in the overall intensity and elimination of *f0*-based acoustic cues. Elements that are voiced in regular phonated speech — vowels and voiced consonants — undergo the greatest reduction of intensity (20-25 dB) while voiceless consonants remain virtually unaffected (Ito et al., 2005, Jovicic & Saric, 2006).

Compared to phonated speech, whispered "voiced" consonants have lower energy at frequencies up to 1.5 kHz and greater spectral flatness (Ito et al., 2005). Spectral flatness describes a deviation of a signal power spectrum from a flat spectral shape (Kostek, 2005, p89). In a white noise spectrum, spectral flatness approaches 1, whereas spectral flatness of a sinusoidal signal gravitates towards zero. In whispered consonants, greater spectral flatness is indicative of more homogenous distribution of energy across the frequency spectrum, reflecting the absence of the periodic energy source. At the same time, despite reduction in energy, whispered consonants overall have greater intensity relative to the vowels (Ito et al., 2005). Thus, comparde to phonated speech, there is a naturally enhanced consonant/vowel amplitude ratio — a unique property of whispered speech that may contribute to its generally high intelligibility, as discussed in the next section.

Whispered consonants (Schwartz, 1972, Parnell et al., 1977, Jovicic & Saric, 2006) and vowels (Parnell et al., 1977) increase in duration relative to phonated speech. Duration of fricative noise in /z/ and /s/ in VCV syllables was respectively 20 ms and 28 ms longer in whispered than in phonated speech (Parnell et al., 1977). Duration of stop closure in bilabial stops /b/ and /p/ increased relative to phonated speech by 16 ms and 19 ms, correspondingly (Schwartz, 1972). At the same time, for alveolar stops, an increase in stop closure duration was observed for whispered /t/ (7 ms) but not whispered /d/, which actually decreased in duration by 5 ms (Parnell et al., 1977). In a study with Serbian consonants, the increase in consonant duration was greater for "voiced" than "voiceless" counterparts (Jovicic & Saric, 2006). Whispered "voiced" stops,

fricatives, and affricates increased in duration by 16 ms, 18 ms and 28 ms, correspondingly, compared to phonated speech, while whispered "voiceless" stops, fricatives and affricates increased by 8 ms, 11 ms and 14 ms, correspondingly (Jovicic & Saric, 2006). The discrepancy between the three studies (Schwartz, 1972, Parnell et al., 1977, Jovicic & Saric, 2006) in the relative increase in consonant duration for "voiced" versus "voiceless" consonants may be attributed to the differences in the segmentation and measurements of whispered consonants. For instance, in Parnell et al. (1977), stop-closure duration was measured between the offset of the preceding vowel indicated by cessation of strong energy in *F2* and the onset of consonant burst; for similar stimulus materials — intervocalic syllable-medial consonants — measurements of stop duration in Jovicic & Saric (2006) included both the closure portion and the burst. In addition, position of the consonant within a syllable was different across studies — syllable-medial in Jovicic & Saric (2006) and Parnell et al. (1977), and syllable-initial in Schwartz (1972). Finally, the studies differed in the languages they were based on — Serbian in Jovicic & Saric (2006) and English in Parnell et al. (1977) and Schwartz (1972).

Although the aforementioned studies evaluated duration of whispered consonants relative to phonated consonants, none directly compared the duration *difference* between "voiced" and "voiceless" consonants in whispered and phonated speech. In phonated speech, duration of the stop-gap is one of the cues contrasting voiced and voiceless consonants with a longer closure for voiceless stops (Raphael, 1972). Similarly, voiceless fricatives are associated with longer frication noise segments (Kent & Read, 2002). A crude analysis of data reported in Schwartz (1972) reveals a trend for an increased duration difference between /b/ and /p/ in whispered speech. The difference is 29 ms and 26 ms for whispered and phonated "voicing" consonant-pairs, correspondingly. Similarly, the duration difference between /d/ and /t/ increased for whispered consonants, 29 ms compared to 17 ms for phonated consonants. At the same time, the trend is opposite for fricatives: a difference of 28 ms for whispered /z/-/s/ and 36 ms for their phonated counterparts (Parnell et al., 1977). Jovicic & Saric, (2006) found a trend for the difference between

members of the "voicing" cognate pairs for all consonant classes to be smaller in whispered than in phonated speech. The difference was 53 ms, 47 ms, 44 ms for whispered stops, fricatives and affricates, respectively, and 61 ms, 53 ms, 58 ms for phonated stops, fricatives and affricates, respectively. Thus, while there is agreement among the studies regarding the general increase in consonant duration in whispered relative to phonated speech, the question of relative duration differences between corresponding "voiced" and "voiceless" consonants in whispered and phonated speech still requires further investigation.

Similar to consonants, durations of vowels also increase in whispered compared to phonated speech. For example, vowels /i/ and /a/ in symmetrical VCV syllables increased by 28 ms and 13 ms, correspondingly (Parnell et al., 1977). To our knowledge, no study investigated the duration of preceding vowels as a cue to final consonant "voicing" in whispered speech. Differences in amplitude rise times have also been found between whispered /b/ vs. /p/ in syllable-initial position (Munro, 1990). In consonant-vowel syllables (CVs) truncated from original CVCs, there was a significant difference in mean amplitudes averaged across the first 10 ms of stimuli between /b/ and /p/ sets and a trend for greater mean amplitude for /b/ averaged across the first 20 ms of stimuli. Rise time slopes were calculated based on the time points from stimulus onset to the point at which amplitude reached 50% and 75% of its peak value. For whispered stimuli, these durations for /b/ were found to be 1.9 ms and 8.8 ms for 50% and 75%, correspondingly and for /p/, 7.9 ms and 12.4 ms, correspondingly, thus indicating a steeper rise time for /b/. The difference in rise times for /b/ vs /p/ was significant only at the 50% time point.

Tartter (1989) also noted differences between spectrographic representations of syllable-initial "voiceless" and "voiced" stops in whispered speech: "voiceless" stops appeared to have a double-spiked burst versus single burst in "voiced" stops. However, this measure was not assessed quantitatively and is therefore hard to compare across studies. She also observed that the first formant cutback was 58-183 Hz higher in "voiceless" than in "voiced" consonants although it was reliably present in whispered fricatives but not in whispered stops.

Despite being an acoustically impoverished signal, whispered speech shares certain characteristics with clear phonated speech. In everyday communication, clear speech is an intelligibility-enhancing speaking mode adopted to help listener's perceptual difficulties, be they due to a lack of proficiency with English, hearing problems, or interfering signals in the communication medium. Speakers also switch to clear speech after encountering recognition errors when communicating with interactive automatic speech recognition systems (Stent et al., 2008). In laboratory conditions, clear speech is elicited by instructing speakers to speak as though talking to a hearing-impaired person (Ferguson, 2004) or communicating in a noisy environment (Picheny et al., 1985); it may also be prompted by simulated recognition errors provided to the listeners as feedback to their original production by a computer system (Maniwa et al., 2008). Differences in the production strategies associated with clear speech improves perception of vowels and consonants compared to that in conversational speech (Picheny et al., 1985, Ferguson, 2004, Maniwa et al., 2008) and generally helps speech intelligibility (e.g., Bradlow et al., 1996).

Acoustic characteristics common to both whispered and clear speech involve timing-based and amplitude cues. Increased durations of consonants and vowels discussed above for whispered speech are also found in clear phonated speech consonants (Picheny et al., 1986, Smiljanic & Bradlow, 2008b,a, Maniwa et al., 2009) and vowels (Smiljanic & Bradlow, 2008b,a). In Picheny et al. (1986), duration of stops — defined as a sum of closure, frication and aspiration segments — increased in phonated clear relative to phonated conversational speech. The increase in closure duration consistently contributed to the overall increase in stop-consonant duration with less obvious and less consistent contribution from frication and aspiration segments. In fricatives, clear speech increased the duration of frication (Picheny et al., 1986, Maniwa et al., 2009). On other hand, decreased intensity of whispered consonants (Jovicic & Saric, 2006) raises the values of the C/V amplitude, also observed in clear speech (Krause & Braida, 2004).

To summarize, acoustic changes in whispered speech affect its spectral, temporal and intensity-related characteristics. Some of these properties — increased vowel and consonant durations and

enhanced C/V intensity-ratio — are similar to those found in clear phonated speech and, as such, might be beneficial in maintaining or enhancing otherwise reduced speech intelligibility.

1.4 Perception of whispered consonants

Two previous studies investigated the perception of whispered consonants extensively (Dannenbring, 1980, Tartter, 1989). In the Dannenbring study (1980), stimuli were isolated consonant-vowel (CV) syllables including 12 consonants (6 consonant "voicing"-pairs) /b-p, d-t, g-k, z-s, v-f, ð-ɵ/ followed by vowels /i/, /a/, or /u/. The experiment was designed so that listener's made a 2-alternative forced-choice decision on each trial; thus, only "voicing" errors were possible. Listeners' responses were scored using a composite measure combining the decision made with a confidence rating about that decision. For each pair of "voiced"/ "voiceless" consonants, these scores were then ranked together and converted to a D-score — a measure controlling for response bias — for each subject. Perfect discrimination of a pair of consonants would be reflected in D-scores equal to 1 while systematic incorrect judgments of two consonants would be expressed in D-scores equal to -1; near-zero D-scores would indicate random judgments. Mean D-scores computed across the listeners indicated that the /d-t/ pair was the easiest to categorize with the scores ranging from 0.83 to 0.95 across the three vowel contexts. Consonants in /ð-ɵ/ and /z-s/ pairs proved to be most challenging with D-scores ranging across vowels from 0.19 to 0.21 and from -0.80 to 0.25, correspondingly. Thus, despite the absence of the dominant cue to voicing in initial stop consonants — voice onset time (VOT) — the listeners were able to make correct judgments based on some secondary cues. However, the lack of acoustical data for the stimuli precludes the possibility of relating the accuracy of whispered consonant identification to the prominence of various secondary and/or alternative cues to voicing in the absence of VOT. A limitation of the Dannenbring study (1980) is that subjects were forced to make their decision solely within a "voicing" pair and, thus, the analysis of whispered-consonant perceptual confusion did not explore potential place of articulation and manner errors. Furthermore,

as the discriminability of the "voiced" vs. "voiceless" consonants was captured by a composite score, neither an overall measure of consonant identification accuracy nor individual "voicing"-pair identification accuracy was available. The use of D-scores (rather than more conventional d′ scores) based on the "voiced"-"voiceless" similarity ratings that were derived from the combined decision-plus-confidence measure further complicates the comparison of whispered consonants identification accuracy across studies.

In Tartter (1989), listeners identified 18 consonants [b, d, g, p, t, k, m, n, r, l, w, y, v, f, z, s, ʃ, ʒ] in /Cɑ/ (consonant-/ɑ/) syllables produced by two speakers. Results indicated that listeners made most errors on consonant "voicing" (72%). Although the identification test had twice as many "voiced" consonants as "voiceless" consonants, the listeners correctly identified "voiced" and "voiceless" consonants 68% and 80% of the time, respectively. In other words, even with a built-in presentation bias towards "voiced" consonants, listeners by and large opted for a "voiceless" consonant category when selecting their response. Accuracy of perception of consonant place and manner of articulation, albeit quite accurate, was not perfect, 91% and 86%, correspondingly. One shortcoming of the Tartter study (1989) is that the stimuli used were repetitions of 18 unique tokens from two speakers, not physically different productions of the stimuli which would have allowed for natural variation in the acoustic parameters. Thus, for a given speaker, a consonant category preceding each vowel was represented by a single token. Another limitation is that, similar to the Dannenbring study (1980) and unlike whispered vowel studies (Kallail & Emanuel, 1985, Tartter, 1991), there was no phonated condition to serve as a control, which could have been used to establish an overall and individual speakers' baseline of consonant identification. In addition, despite the range of consonants in question, identification accuracy of "voicing" was not analyzed for interaction with manner or place of articulation. Finally, neither Dannenbring (1980) nor Tartter (1989) related acoustical data with consonant identification accuracy, possibly because the limited number of speakers and tokens representing each consonant category made this difficult.

In a different study, Munro (1990) investigated the relationship in whispered /b/ vs. /p/ identification patterns and differences in the amplitude rise time of the syllable-initial consonant relative to the following vowel /æ, ɛ, i, u/. Amplitude rise time was defined as a time point from the stimulus onset to the point at which amplitude reached 50% of the mean vowel amplitude. The latter was calculated by averaging amplitude values over the segment of 30 ms between 30 and 60 ms from the stimulus onset. Consistent with Tartter (1989), the average correct identification of whispered consonants across 8 listeners was 63% with the range from 52% to 82%. Despite significant differences in the rise time between whispered /b/ and /p/, no correlation was found between the number of correctly identified /b/-tokens and small rise time values (steeper rise slope) or the number of correctly identified /p/-tokens and greater rise time values (more gradual rise slope).

To summarize, when perception of whispered consonants in syllable-initial position in isolated CV-syllables was investigated, listeners made more errors categorizing consonants based on "voicing" than on manner and place of articulation. At the same time, identification of whispered "voiced" and "voiceless" consonants remained better than chance, indicating that in the absence of the dominant periodicity cue to voicing in whispered speech listeners were able to make "voicing" judgments based on the remaining cues. Despite differences in spectrum, duration and intensity-related acoustic parameters between whispered "voiced" versus "voiceless" consonants, there is limited information so far regarding which acoustic properties are essential for listeners in identifying "voicing" in whispered consonants. In addition, published research on whispered consonant perception used a limited number of speakers and/or tokens and, thus, the effect of intra- and inter-subject variability in production on whispered consonant intelligibility is not known.

Intelligibility of whispered consonants may be rooted in increased durations of the phonetic segments (Schwartz, 1972, Parnell et al., 1977, Jovicic & Saric, 2006) and favorable consonant/vowel amplitude-ratio. Exploring those acoustic cues to consonant "voicing" that become

the primary source of differentiation of the "voiced" vs. "voiceless" consonants in whispered speech may help our general understanding of speech perception under non-optimal listening conditions by the normal hearing. In addition, it may contribute to our knowledge of acoustically impoverished speech such as speech perceived though cochlear implants and hearing aids by the hearing impaired. The current project is divided into two parts. The objective of the production study — the first part of this project — was to document the differences that exist in the acoustic cues for "voicing" in phonated and whispered syllable-final consonants produced in sentence medial position at a *conversational* rate by a set of four speakers. Parameters of interest included intensity and duration of the final consonant closure (interval of maximum restriction), and intensity, duration and frequency of the vowel preceding that consonant, i.e. the vocalic nucleus and offglide into the consonant closure. A detailed description of the segmentation markers and the parameters themselves is presented in the following section. The second objective was to assess the effect of clear whispered speech on the acoustic parameters associated with consonant "voicing". Additional acoustical analysis was conducted on whispered clear-speech words spoken by the same set of talkers and the measurements were then compared to the parameters obtained for their whispered conversational speech.

The second part of the project was concerned with perception of consonants in whispered compared to phonated speech and the effect of clear whispering on consonant identification accuracy. It is subdivided into two studies, Study 2A and Study 2B. The goal of Study 2A was twofold: (1) to establish the baseline of consonant perception accuracy in conversational phonated and whispered speech and to explore the distribution of errors on "voicing", place and manner of articulation; (2) to investigate the general effect of individual speakers and consonants on the accuracy of consonant perception in the two phonation modes. Study 2B had four specific goals. The first goal was to investigate the frequency and the types of errors made by listeners in perception of whispered speech. The second goal was to explore whether perception of consonants improved in clear whispered speech and if this potential improvement was uniform across speak-

ers and consonant-pairs. The third goal was to compare performance on consonant perception in whispered speech — overall and separately within each speaker and consonant-pair — with the predictions made by the logistic regression analysis. The final goal was to investigate associations between the magnitude of specific acoustic contrasts and intelligibility of whispered consonants.

Chapter 2

Production of "voicing" in whispered vs. phonated consonants

2.1 Materials and methods

2.1.1 Speakers

Speakers for this study were four female, monolingual speakers of American English, born and raised in the New York City metropolitan area with no history of speech or language disorders. They ranged in age from 22 to 30 years old. The speakers were recruited from the CUNY Graduate Center community and were compensated at the rate of $15 per hour for their participation. Prior to testing, all speakers signed a consent form approved by the IRB of the CUNY Graduate Center.

2.1.2 Stimuli

Tests words were nonsense disyllables *habVC* embedded in a carrier sentence 'I said _____ eight times'. The word following the test word was chosen to begin with a vowel in order to encourage speakers to produce a burst release of the final consonant – a potential acoustic cue to consonant

voicing. Each speaker recorded 8 lists of 9 sentences repeating each list four times in each of the three conditions: conversational phonated, clear whispered, conversational whispered. The lists were blocked by each of 8 American English vowels, /iː, ɪ, æː, ɛ, ɑː, ʌ, uː, ʊ/ within a condition. Only four of these eight vowels /æː, ɛ, ɑː, ʌ/ were used in the analysis of the "voicing" cues. Within a list, a particular vowel preceded each of the 8 consonants comprising four voicing pairs: stops /b-p, g-k/, fricatives /z-s/, and affricates, /dʒ-tʃ/. The order of the consonants within a list was randomized. The last item in each list was a repetition of one of the sentences in the list and was discarded to avoid the effect of list-final prosody.

2.1.3 Recording

During the recording session, each subject was seated in a sound-attenuated booth equipped with a dynamic Shure SM48 microphone with a pop-screen to prevent explosive breath sounds. During the recording of phonated speech, the microphone was placed at about 20 cm from the speaker's lips; during whispered speech production, this distance was reduced to 7 cm. Between sublists of stimuli, subjects were asked to re-measure this distance to ensure it remained constant. The overall input level was monitored on a computer screen to ensure that it remained relatively constant. Microphone output was routed to a preamplifier set to 54 dB gain during phonated speech production and readjusted to 60 dB during production of whispered speech. As it was empirically established, a gain of 60 dB allowed for maximum amplification of whispered speech signal without introducing any distortion while preserving the relative intensity difference between phonated and whispered speech modes. The output was then amplified using an Earthworks Microphone Preamplifier LAB101 and digitized at 16 bits, 22,050Hz, using Sound Forge 4.5 sound editing software. The program was installed on a Dell Dimension Pentium II XPS D233 using Microsoft Windows 98 platform and equipped with a Turtle Beach Montego II Sound Card. All the data were stored as *.wav* monaural audio files. A recording session lasted approximately 2 hours. Three speaking modes were recorded in the following order: conver-

sational phonated, clear whispered, conversational whispered. For the conversational phonated and whispered speech conditions, speakers were asked to read the sentences at close to their daily conversational rate, as if they were talking to a friend without any particular focus on the enunciation. For production of clear whispered speech, speakers were instructed to read the test sentences at a rate and with an enunciation effort suitable when talking to a hearing-impaired person or repeating a sentence after a normal-hearing listener had made a perceptual error in identifying the test word. The first set of lists in each condition served as practice for the speakers and was not used in the acoustical analysis or the perception test. During this practice set, the experimenter provided speakers with feedback on their manner of speaking and intensity level.

2.1.4 Acoustical analysis of the stimuli

Acoustical analysis was performed to quantify acoustic cues to final consonant "voicing" and to compare measurements across the same words produced in two phonation modes and speaking styles. The choice of the acoustic parameters was based on the acoustic cues known to contribute to the "voicing" distinction in phonated speech (Kent & Read, 2002, Port & Dalby, 1982, Raphael, 1972, Repp, 1979, Revoile et al., 1982). Acoustic parameters included preceding vowel duration, consonant duration, consonant-to-vowel duration-ratio, consonant-to-vowel amplitude intensity-ratio, difference in the first formant values at the preceding vowel offset and 75% point of the vocalic nucleus (*F1*-dynamics). Based on these direct measures, derived measures capturing an overall magnitude of the "voicing" contrasts were computed for each consonant-pair, separately for each individual vowel context and across all four vowel contexts. The five derived measures were: (1) difference in vowel duration between "voiced" and "voiceless" consonants (2) difference in consonant duration between "voiceless" and "voiced" consonants, (3) difference in the consonant-to-vowel duration-ratio between "voiceless" and "voiced" consonants (4) difference in $F1$-dynamics between "voiced" and "voiceless" consonants, (5) difference in the consonant-to-vowel amplitude intensity ratio between "voiced" and "voiceless" consonants.

2.1.4.1 Segmentation of the stiumuli

Segmentation of the stimuli was carried out according to the guidelines described below. In all three types of consonants — stops, fricatives and affricates — the "vowel" was defined as a segment between the offset of /b/ closure and the onset of the final consonant closure in *habVC*. Specifically, the beginning of the vowel was marked at the onset of the segment on the waveform following the burst release of the /b/ where both $F1$ and $F2$ were identified on the spectrogram. The offset of the vowel was marked at the point of abrupt intensity drop in the waveform, and cessation of energy and loss of structure in the upper formants in the spectrogram. In addition to the onset and offset marks, three time-points corresponding to 25%, 50% and 75% of the vocalic nucleus length were established within the vowel portion. The stop gap in stops was defined as a segment between the offset of the preceding vowel and the onset of the release burst characterized by the reappearance of intensity peaks in the waveform, and the onset of the wide band of energy in the spectrogram. The offset of the burst was marked at the point of the waveform following burst-related peaks and abrupt energy drop in the spectrogram. In affricates, the stop-gap was measured from the offset of the vowel to the onset of high frequency noise corresponding to the frication portion. Fricative duration was defined as the segment between the offset of the vowel and the offset of high-frequency noise. For the affricates, the frication portion was considered a segment between the onset and the offset of high frequency noise. The acoustic parameters extracted based on the above markings are summarized below.

2.1.4.2 Measurement and calculation of the acoustic parameters

Vowel duration was measured as a segment between the vowel onset and offset markers as described above. Measurement of consonant duration was conducted differently depending on the consonant type: in stops, consonant duration was considered to be the combination of stop gap and the burst; in fricatives, it was defined as the duration of fricative noise; in affricates, consonant duration was measured as a combination of stop gap and fricative noise durations.

Consonant-to-vowel duration-ratio (CVDR) was obtained by dividing the length of the consonant segment by the length of the vowel segment. The $F1$-dynamics measure was computed as a difference in the first formant values between 75% time-point of the vocalic nucleus and the vowel offset. Finally, consonant-to-vowel intensity-ratio was obtained by *subtracting* amplitude intensity of the consonant-segment from the amplitude intensity of the vowel-segment. In addition to these direct measures, five derived measures capturing an overall magnitude of the "voicing" contrasts were computed for each consonant-pair. The five derived measures were: (1) difference in vowel duration between "voiced" and "voiceless" consonants (2) difference in consonant duration between "voiceless" and "voiced" consonants, (3) difference in the consonant-to-vowel duration-ratio between "voiceless" and "voiced" consonants (4) difference in $F1$-dynamics between "voiced" and "voiceless" consonants, (5) difference in the consonant-to-vowel amplitude intensity ratio between "voiced" and "voiceless" consonants. Difference measures were obtained for each consonant-pair for each vowel in each speaker. They were computed by subtracting the average value of the direct measure for a particular "voiceless" consonant — averaged over two tokens of each type for each speaker — from the average value of the "voiced" consonant. For instance, for the /hɑbɛb/-/hɑbɛp/ pair, duration of /ɛ/ before /p/ in /hɑbɛp/ was subtracted from the duration of /ɛ/ before /b/ in /hɑbɛb/ (4) The "voiceless"/"voiced" consonant duration-ratio was obtained using stop-gap duration for stops, fricative noise duration for fricatives, and the combination of stop-gap and fricative noise for affricates. For example, for a /hɑbɛz/-/hɑbɛs/ pair, fricative noise duration of /z/ was subtracted from the duration of fricative noise of /s/ in /hɑbɛs/. The analysis was carried out using MATLAB-based software incorporating COLEA code for formant tracking Loizou (1999) and parts of the interface and algorithms of "CVCZ" program developed by Valeriy Shafiro.. The original program allowed selecting a segment of interest on the waveform/spectrogram, setting 25%, 50% and 75% time-point markers and performing FFT/LPC analyses at these points using a Hamming window of 25 ms and 24 LPC coefficients. The program was modified to include the following: measurements of formant values at the on-

set and offset of the vowel, additional segmentation of vowels and consonants, calculation of RMS and peak intensity values, re-access to previously logged measurements and an extended interface allowing for an online display of the measurements.

To explore the effect of whispered speech on the absolute acoustic measures, three-way mixed ANOVAs were performed on each acoustic parameter for the effects of Speaker as the between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors (repeated measures). Of particular interest was the Phonation-mode x Voicedness interaction. In the situations when this interaction was significant, the analysis was supplemented with *t*-test for related samples conducted on the difference in the acoustic parameter between the "voiced" and the "voiceless" consonants, i.e. acoustic contrast, in phonated conversational and whispered conversational speech. In the situations of the significant three-way Phonation-mode x Voicedness x Speaker interaction, additional two-way repeated measures ANOVA was conduced for the effects of Voicedness and Phonation-mode in each speaker. When necessary, additional analysis investigated the effect of whispered speech on individual consonant-pairs. In such cases, three-way mixed ANOVAs were performed on the difference in an acoustic parameter between the "voiced" and the "voiceless" consonants, i.e. acoustic contrast, for the effects of Speaker as a between-subject factor and Consonant-pair and Phonation-mode as within-subject factors. To explore significant effect of Consonant-pair (when available), post-hoc multiple comparisons using *t*-tests with Bonferroni correction were performed.

Similarly to the analysis described above, investigation of the effect of clear speaking-style on the absolute acoustic measures in whispered speech was conducted using three-way mixed ANOVAs. The ANOVAs were performed on each acoustic parameter for the effects of Speaker as a between-subject factor and Consonant-pair and Speaking-style as within-subject factors. Additional tests and post-hoc comparisons followed the same structure as outlined above for Phonated conversational vs. Whispered conversational speech analysis.

2.2 Results

2.2.1 Differences in acoustic cues to "voicing" in conversational phonated vs. conversational whispered speech

The goal of this section was to explore the effect of whispered speech on the variability of acoustic parameters associated with consonant "voicing". This was performed by comparing the effect of phonation-mode (phonated vs. whispered) on the magnitude of the "voicing" contrasts.

2.2.1.1 Speaking rate

To explore natural differences in speaking rates between speakers, speaking rates were calculated for each phonation condition in individual speakers. The entire stimulus phrase "*I said habVC eight times*" was used to estimate individual speaking rates. Although in whispered speech two speakers, Speaker 1 and Speaker 4, slowed down, they both spoke at a faster rate than the other two speakers in both phonation modes (see Fig. 1).

2.2.1.2 Vowel duration

Average vowel durations in two "voicing" contexts in phonated and whispered speech modes are shown in Fig. 2. The average vowel durations of individual speakers in each phonation mode for the "voiced" and "voiceless" contexts are given in Table 1 (for further details, see Table A1)

Three-way mixed ANOVA was conducted on the vowel duration for the effects of Speaker as the between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors (see summary in Table A2 in Appendix A). The results indicated that all three main effects were significant: Phonation-mode, $F(1, 124) = 117.94$, $p < 0.01$, Voicedness, $F(1, 124) = 860.60$, $p < 0.01$ and Speaker, $F(3, 124) = 13.99$, $p < 0.01$. There were two significant two-way interactions was significant: Phonation-mode by

Speaker, $F(3, 124) = 33.29$, $p < 0.01$ and Voicedness by Speaker, $F(3, 124) = 3.04$, $p < 0.05$. The three-way interaction of Phonation-mode by Voicedness by Speaker was also significant, $F(3, 124) = 18.57$, $p < 0.01$.

These results showed that vowels in the "voiced" context (M = 170.64, SD = 33.82) were longer than vowels in the "voiceless" context (M = 139.02 , SD = 32.73) and that vowels increased in duration in whispered speech (M = 161.34, SD = 34.45) relative to phonated speech (M = 148.91, SD = 38.31). Also, the four speakers differed in average vowel duration. Furthermore, the significant two-way Phonation-mode by Speaker interaction suggested that effect of whispered speech on the vowel duration was different in the four speakers. A series of pairwise comparisons (t-test for related samples) was conducted on vowel durations between whispered and phonated speech for each speaker. It can be seen in Fig. 3, that vowel durations were substantially longer in whispered than in phonated speech only for Speaker 1 and Speaker 4. The results of the comparisons confirmed this impression (for details, see Table A3 in Appendix A). The significant two-way interaction of Voicedness by Speaker indicated that vowel duration "voicing" *contrast* was generally different in the four speakers. This was further investigated within the analysis of the three-way interaction of Voicedness by Speaker by Phonation-mode below.

The absence of the significant Phonation-mode by Voicedness interaction indicated that overall, vowels in both "voiced" and "voiceless" contexts were lengthened to a similar degree and that the "voiced" vs. "voiceless" vowel duration contrast remained equivalent in the two phonation modes. On the other hand, the significant three-way interaction of Phonation-mode by Voicedness by Speaker suggested the *magnitude* of the vowel duration "voicing" contrast was differently affected by whispered speech in the four speakers. To investigate this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Phonation-mode in each speaker. In Speakers 1 and 4, only two main effects were significant (Speaker 1: Voicedness: $F(1, 31) = 321.02$, $p < 0.01$, Phonation-mode: $F(1,$

31) = 142.63, p <0.01; Speaker 4: Voicedness: $F(1, 31) = 145.09$, $p < 0.01$, Phonation-mode: $F(1, 31) = 87.48$, $p < 0.01$. In Speaker 2, there was a significant main effect of Voicedness, $F(1, 31) = 265.93$, $p < 0.01$, and a significant Phonation-mode by Voicedness interaction, $F(1, 31) = 59.25$, p <0.01. Similarly, in Speaker 3, there was a significant main effect of Voicedness, $F(1, 31) = 226.26$, $p < 0.01$, and a significant Phonation-mode by Voicedness interaction, $F(1, 31) = 17.47$, p <0.01. The presence of the significant Voicedness x Phonation-mode interaction in Speakers 2 and 3 indicated that only in these two speakers, whispered speech had an effect on the magnitude of the vowel duration "voicing" contrast. It can be seen from Fig. 3, that in Speaker 2, vowel duration contrast increased whereas in Speaker 3, it decreased in whispered conversational relative to phonated conversational speech. To explore these impressions statistically, t-test for related samples was conducted to compare the vowel duration *contrast* (i.e. difference in the vowel duration between the "voiced" and the "voiceless" consonants) in phonated and whispered speech conditions in these two speakers. The results confirmed a significant increase in the contrast magnitude of Speaker 2 in whispered speech ($M = 38.87$, $SD = 9.88$) relative to phonated speech ($M = 19.84$, $SD = 10.65$), $t(15) = 7.69$, p <0.01, and a significant decrease in the contrast magnitude of Speaker 3 in whispered speech ($M = 29.31$, $SD = 16.08$) relative to phonated speech ($M = 44.75$, $SD = 11.88$), $t(15) = -4.35$, p <0.01.

To summarize, in whispered speech, vowels lengthened in both "voiced" and "voiceless" consonant contexts so that the same magnitude of the vowel duration contrast was maintained in both phonation modes. However, the pattern of changes in vowel duration between phonated and whispered speech was not consistent across speakers.

2.2.1.3 Consonant duration

Average durations of "voiced" and "voiceless" consonants in phonated and whispered speech modes are shown in Figure 4. The average "voiced" and "voiceless" consonant durations of individual speakers in each phonation mode are given in Table 2 (for further details, see Table

A4)

Three-way mixed ANOVA was conducted on the consonant duration for the effects of Speaker as a between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors (see summary of the results in Table 13 in Appendix A). All three main effects were significant: Phonation-mode, $F(1, 124) = 130.40$, $p < 0.01$, Speaker, $F(3, 124) = 2.83$, $p < 0.01$ and Voicedness, $F(1, 124) = 738.04$, $p < 0.01$. There were also significant interactions: Phonation-mode x Speaker $F(3, 124) = 20.24$, $p < 0.01$ and Voicedness x Speaker, $F(3, 124) = 7.69$, $p < 0.01$. The three-way interaction of Phonation-mode x Speaker x Voicedness was also significant, $F(3, 124) = 4.11$, $p < 0.01$.

These results showed that "voiceless" consonants (M = 109.28, SD = 23.20) were longer than "voiced" consonants (M = 82.32, SD = 18.22) and that consonants lengthened in whispered speech (M = 101.70, SD = 22.82) compared to phonated speech (M = 89.89, SD = 21.11). In addition, four speakers differed in the average consonant duration. As with the analysis of vowel durations reported above, the absence of the significant Phonation-mode by Voicedness interaction indicated that in whispered speech both "voiced" and "voiceless" consonants lengthened to a similar degree and that the duration contrast between them was equivalent in the two phonation modes. On the other hand, Phonation-mode x Speaker interaction suggested that changes in the consonant durations in whispered relative to phonated speech were different in the four speakers. It can be seen from Fig. 5, that consonant durations in Speakers 2, 3, 4 increased in whispered speech. At the same time, consonant durations in Speaker 1 were virtually the same in the two phonation modes. These impressions were confirmed by pairwise comparisons (t-test for related samples) of the consonant durations between whispered and phonated speech conducted for each speaker. The length of consonants increased in whispered relative to phonated speech in all speakers except Speaker 1 (for more details, see Table A6 in Appendix A).

The significant Voicedness x Speaker interaction suggested that the magnitude of the consonant duration contrast between "voiced" and "voiceless" consonants was generally different

in individual speakers. This was further investigated within analysis of the three-way interaction of Voicedness by Speaker by Phonation-mode. The latter suggested that the *magnitude* of the consonant duration "voicing" contrast was differently affected by whispered speech in the four speakers. To investigate this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Phonation-mode in each speaker. In Speakers 2 and 3, both main effects and the interaction were significant (Speaker 2: Voicedness, $F(1, 31) = 306.23$, $p < 0.01$, Phonation-mode, $F(1, 31) = 217.68$, $p < 0.01$, Voicedness x Phonation-mode, $F(1, 31) = 5.74$, $p < 0.01$; Speaker 3: Voicedness: $F(1, 31) = 120.85$, $p < 0.01$, Phonation-mode, $F(1, 31) = 26.60$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 7.38$, $p < 0.05$). In Speaker 1, only the main effect of Voicedness was significant, $F(1, 31) = 106.97$, $p < 0.01$. In Speaker 4, both main effects but not the interaction were significant: Voicedness, $F(1, 31) = 280.09$, $p < 0.01$, Phonation-mode, $F(1, 31) = 46.36$, $p < 0.01$. The presence of the significant Voicedness x Phonation-mode interaction in Speakers 2 and 3 indicated that only in these two speakers, whispered speech had an effect on the magnitude of the consonant duration "voicing" contrast. It can be seen from Fig. 5, that in Speaker 2, consonant duration contrast increased whereas in Speaker 3, it decreased in whispered conversational relative to phonated conversational speech. T-test for related samples was conducted to confirm these impressions: the results showed a significant increase in the contrast magnitude of Speaker 2 in whispered speech ($M = 33.05$, $SD = 11.74$) relative to phonated speech ($M = 25.60$, $SD = 9.45$), $t(15) = 2.43$, $p < 0.05$, and a significant decrease in the contrast magnitude of Speaker 3 in whispered speech ($M = 19.66$, $SD = 16.43$) relative to phonated speech ($M = 29.92$, $SD = 9.92$), $t(15) = -2.59$, $p < 0.01$.

To recap, in whispered speech, both "voiced" and "voiceless" consonants lengthened relative to phonated speech so that the magnitude of the consonant duration contrast was equivalent in both phonation modes. Independently of the phonation mode, the magnitude of the contrast differed significantly across speakers.

2.2.1.4 Consonant-to-vowel duration-ratio

Average consonant-to-vowel duration-ratios (CVDR) in phonated and whispered speech are shown in Fig. 6 for the two "voicing" contexts. The average values of CVDR of individual speakers in each phonation mode in the "voiced" and "voiceless" contexts are given in Table 3 (for more information, see Table A7).

Three-way mixed ANOVA was conducted on the CVDR for the effects of Speaker as the between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors(see ANOVA summary in Table A8 in Appendix A). Two main effects were significant: Speaker, $F(3, 124) = 7.14$, $p < 0.01$ and Voicedness, $F(1, 124) = 635.03$, $p < 0.01$; significant interactions were Phonation-mode x Speaker, $F(3, 124) = 39.36$, $p < 0.01$, Voicedness x Speaker, $F(3, 124) = 3.91$, $p < 0.05$, and Voicedness x Phonation-mode x Speaker, $F(3, 124) = 11.52$, $p < 0.01$. These results showed that CVDR was significantly larger in the "voiceless" contexts (M = 0.84, SD = 0.28) than in the "voiced" contexts (M = 0.51, SD = 0.17). Four speakers also differed in the average CVDR. Furthermore, the significant Voicedness by Speaker interaction indicated that the magnitude of the CVDR "voicing" contrast was generally different in the four speakers. Since we were primarily interested in the changes in the contrast magnitude in whispered relative to phonated speech, this interaction was further investigated within the three-way Voicedness x Phonation-mode x Speaker interaction below.

The absence of the main effect of Phonation-mode suggested that overall CVDRs were similar in the two phonation modes. Furthermore, the absence of the Voicedness x Phonation-mode interaction indicated that CVDR "voiced" vs. "voiceless" *contrast* had similar magnitudes in phonated and whispered speech. On the other hand, the Phonation-mode x Speaker interaction indicated that the effect of whispered speech on CVDR was different in the four speakers. In Fig. 7, it can be seen that across speakers, CVDR changed inconsistently between phonated and whispered modes. A pairwise comparison of CVDR in whispered speech vs. CVDR in phonated speech (t-test for related samples) was conducted for each speaker. The results of com-

parisons determined that CVDR significantly decreased in Speaker 1 and significantly increased in Speaker 2 in whispered speech (for details, see Table A9 in Appendix A).

The significant three-way Voicedness x Phonation-mode x Speaker interaction indicated that whispered speech affected the magnitude of the CVDR "voicing" contrast differently in the four speakers. To explore this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Phonation-mode in each speaker. In all speakers with an exception of Speaker 1, both main effects and the interaction were significant (Speaker 2: Voicedness: $F(1, 31) = 216.39$, $p < 0.01$, Phonation-mode: $F(1, 31) = 122.25$, $p < 0.01$, Voicedness x Phonation-mode, $F(1, 31) = 29.65$, $p < 0.01$, Speaker 3: Voicedness: $F(1, 31) = 121.81$, $p < 0.01$, Phonation-mode: $F(1, 31) = 4.56$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 12.15$, $p < 0.01$, Speaker 4: Voicedness: $F(1, 31) = 139.34$, $p < 0.01$, Phonation-mode: $F(1, 31) = 5.67$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 5.47$, $p < 0.05$). In Speaker 1, only two main effects were significant (Voicedness: $F(1, 31) = 213.54$, $p < 0.01$, Phonation-mode: $F(1, 31) = 26.90$, $p < 0.01$). The presence of the significant Voicedness x Phonation-mode interaction in three speakers reflected the effect of whispered speech on the CVDR "voicing" contrast. It may be seen from Fig. 7, that in Speaker 2, CVDR "voicing" contrast increased whereas in Speakers 3 and 4, CVDR "voicing" contrast decreased in whispered relative to phonated speech. A series of pairwise comparisons using t-test for related samples was conducted to confirm these impressions statistically: results determined a significant increase in the contrast magnitude in Speaker 2 in whispered speech ($M = 0.39$, $SD = 0.16$) relative to phonated speech ($M = 0.23$, $SD = 0.10$), $t(15) = 4.58$, $p < 0.01$, a significant decrease in the contrast magnitude in Speaker 3 in whispered speech ($M = 0.22$, $SD = 0.16$) relative to phonated speech ($M = 0.33$, $SD = 0.12$), $t(15) = -3.590$ $p < 0.01$, and a significant decrease in the contrast magnitude of Speaker 4 in whispered speech ($M = 0.34$, $SD = 0.24$) relative to phonated speech ($M = 0.45$, $SD = 0.17$), $t(15) = -2.15$ $p < 0.05$.

To summarize, the results of CVDR analysis indicated that CVDR and the CVDR-contrast

in "voiced" vs. "voiceless" consonant contexts remained equivalent in the two phonation modes. At the same time, across individual speakers, the changes in CVDR between phonated and whispered speech were inconsistent.

2.2.1.5 First formant offset dynamics

As described in the previous chapter, $F1$-offset dynamics were expressed as Bark difference between first formant values at the offset of the preceding vowel and the 75% time-point of the vocalic nucleus. Average values of $F1$-offset dynamics in phonated and whispered speech are shown in Fig. 8 for the two "voicing" contexts. The average values of $F1$-offset dynamics of individual speakers in both phonation modes for the "voiced" and "voiceless" contexts are given in Table 4 (for further details, see Table A10 in Appendix A).

Three-way mixed ANOVA was conducted on $F1$-offset dynamics for the effects of Speaker as a between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors (see summary of the results in Table A11 in Appendix A). All three main effects were significant: Speaker, $F(3, 124) = 6.83$, $p < 0.01$, Voicedness, $F(1, 124) = 86.00$, $p < 0.01$, Phonation-mode, $F(1, 124) = 56.50$, $p < 0.01$. All three two-way interactions were also significant: Speaker x Voicedness, $F(3, 124) = 6.96$, $p < 0.01$, Phonation-mode x Speaker, $F(3, 124) = 10.56$, $p < 0.01$ and Phonation-mode x Voicedness, $F(1, 124) = 11.96$, $p < 0.01$. The three-way Phonation-mode x Voicedness x Speaker interaction was also significant $F(1, 124) = 13.28$, $p < 0.01$. These results indicated that $F1$ was significantly more dynamic (i.e. larger difference between the two points around vowel offset) in the "voiced" consonants ($M = 1.14$, $SD = 0.80$) than in the "voiceless" consonants ($M = 0.56$, $SD = 0.79$). Furthermore, $F1$-offset was more dynamic in phonated speech ($M = 1.07$, $SD = 0.95$) than in whispered speech ($M = 0.63$, $SD = 0.65$).

The interaction of Speaker x Voicedness indicated that average $F1$-offset dynamics contrast between "voiced" and "voiceless" contexts was different in the four speakers. Since we were

primarily interested in the changes in the contrast magnitude between two phonation modes rather than overall differences in the contrast magnitude between speakers, this was investigated in the analysis of the three-way Phonation-mode x Voicedness x Speaker interaction below.

The interaction of Phonation-mode x Speaker indicated that changes in $F1$-offset dynamics in whispered relative to phonated speech were not uniform across speakers. In Fig. 9, it can be seen that the decrease in $F1$-offset dynamics in whispered relative to phonated speech was particularly pronounced in Speakers 2 and 3. Pairwise comparisons of $F1$-offset dynamics between whispered and phonated speech (t-test for related samples) was conducted for each speaker to investigate these relationships. The results confirmed that in Speakers 2 and 3, the $F1$-offset dynamics significantly decreased in whispered compared to phonated speech. At the same time, in Speakers 1 and 3 the trend for decreased $F1$-offset dynamics was not significant (see details in Table A12).

The Phonation-mode x Voicedness interaction suggested that whispered speech affected the magnitude of the $F1$-offset dynamics contrast between the "voiced" vs. "voiceless" consonants. It can be seen in Fig. 8, that the difference between $F1$-offset dynamics in the "voiced" and "voiceless" contexts was larger in phonated speech. The pairwise comparison of the "voiced"$-$"voiceless" difference, conducted separately for phonated and whispered speech, indicated that in both phonation modes, it was significant (see details in Table A13). Furthermore, the pairwise comparison of the difference in the $F1$-offset dynamics *contrast* between phonated and whispered speech, revealed that the contrast was significantly smaller in whispered speech relative to phonated speech (for details, see Table A14).

We next investigated the three-way Phonation-mode x Voicedness x Speaker interaction which suggested that the effect of whispered speech on the $F1$-offset "voicing" contrast was different in the four speakers. To explore this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Phonation-mode in each speaker. In Speaker 1, the main effect of Voicedness was significant $F(1, 31) = 21.91$, $p <$

0.01; the Voicedness x Phonation-mode interaction was also significant, $F(1, 32) = 10.81$, $p < 0.01$. In Speaker 2, main effect of Phonation-mode was significant, $F(1, 31) = 33.25$, $p < 0.01$; effect of Voicedness was marginally significant, $F(1, 31) = 4.06$, $p = 0.05$, the interaction was not significant. In Speaker 3, both main effects were significant (Voicedness: $F(1, 31) = 104.05$, $p < 0.01$; Phonation-mode: $F(1, 31) = 32.55$, $p < 0.01$). The Voicedness x Phonation-mode interaction was also significant, $F(1, 31) = 42.13$, $p < 0.01$. Finally, in Speaker 4, only the main effect of Voicedness was significant $F(1, 31) = 17.32$, $p < 0.01$. These results indicated that in Speakers 2 and 4, the magnitude of the $F1$-offset dynamics contrast between the "voiced" and "voiceless" contexts was equivalent in both phonation modes whereas in Speakers 1 and 3, whispered speech affected the magnitude of the $F1$-offset dynamics contrast (see Fig. 9). T-test for related samples confirmed this pattern. In Speaker 1, the contrast of $F1$-offset dynamics in whispered speech ($M = 0.02$, $SD = 0.66$) was significantly smaller compared to phonated speech ($M = 0.69$, $SD = 0.44$), $t(15) = -2.75$, $p < 0.01$. Similarly, for Speaker 3, the contrast of $F1$-offset dynamics in whispered speech ($M = 0.35$, $SD = 0.57$) was significantly smaller compared to phonated speech ($M = 1.71$, $SD = 0.79$), $t(15) = -5.71$, $p < 0.01$.

In short, overall $F1$-offset was less dynamic in whispered speech. Although the difference in $F1$-offset dynamics between "voiced" and "voiceless" contexts was generally smaller in whispered speech, in some speakers, the contrast was maintained.

2.2.1.6 Consonant-to-vowel amplitude-ratio

As explained in the previous chapter, consonant-to-vowel amplitude-ratio (CVAR) was computed as a difference between amplitude of a consonant and amplitude of a preceding vowel. Average CVAR in phonated and whispered speech are shown in Fig. 10 for the "voiced" and "voiceless" consonant contexts. The average CVARs of individual speakers in each phonation mode across the "voiced" and "voiceless" contexts are given in Table 5 (for more details, see A16 in Appendix A).

Three-way mixed ANOVA was conducted on the CVAR for the effects of Speaker as a between-factor and consonant Voicedness ("voiced" vs "voiceless") and Phonation-mode (phonated vs. whispered) as two within-subject factors (see Table A17 for ANOVA summary). Only two out of the three main effects were significant: Phonation-mode, $F(1, 124) = 939.14$, $p < 0.01$, and Voicedness, $F(1, 124) = 31.81$, $p < 0.01$. Effect of Speaker was not significant. The significant interactions were Phonation-mode x Speaker, $F(3, 124) = 32.16$, $p < 0.01$, and Phonation-mode x Voicedness, $F(1, 124) = 74.54$, $p < 0.01$. The three-way interaction Voicedness x Phonation-mode x Speaker was also significant, $F(3, 124) = 2.76$ $p < 0.05$.

These results indicated that CVAR was significantly larger in phonated ($M = -16.62$, $SD = 5.40$) than in whispered ($M = -4.10$, $SD = 7.66$) speech and in the "voiceless" ($M = -9.53$, $SD = 7.89$) than in the "voiced" consonants ($M = -11.19$, $SD = 10.14$).

The significant Phonation-mode x Speaker interaction suggested that whispered speech affected CVARs of the fours speakers to a different extent. Pairwise comparison of the CVARs between phonated and whispered speech conducted for each speaker (t-test for related samples) determined that the differences in CVARs between the two phonation modes were significant in all speakers. The magnitude of the difference was different in the four speakers (see details of the comparison in Table A20 in Appendix A).

The significant Phonation-mode x Voicedness interaction indicated that the magnitude of the CVAR contrast between "voiced" and "voiceless" consonants was different in phonated and whispered speech. Pairwise comparisons between the "voiced" and the "voiceless" consonants in each phonation mode indicated that the *difference* in C/V amplitude-ratio in "voiced" vs. "voiceless" consonants was significant only in phonated speech (see Table A19 for details).

We next investigated the three-way Phonation-mode x Voicedness x Speaker interaction which suggested that the effect of whispered speech on the CVAR "voicing" contrast was different in the four speakers. To explore this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Phonation-mode in each

speaker. Both main effects and the interaction were significant in all four speakers (Speaker 1: Voicedness: $F(1, 31) = 4.71$, $p < 0.05$, Phonation-mode: $F(1, 31) = 269.39$, $p < 0.01$, Voicedness x Phonation-mode, $F(1, 31) = 25.09$, $p < 0.01$, Speaker 2: Voicedness: $F(1, 31) = 13.84$, $p < 0.01$, Phonation-mode: $F(1, 31) = 329.15$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 11.49$, $p < 0.01$, Speaker 3: Voicedness: $F(1, 31) = 4.99$, $p < 0.05$, Phonation-mode: $F(1, 31) = 277.46$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 13.71$, $p < 0.01$, Speaker 4: Voicedness: $F(1, 31) = 11.25$, $p < 0.01$, Phonation-mode: $F(1, 31) = 129.04$, $p < 0.05$, Voicedness x Phonation-mode, $F(1, 31) = 26.84$ $p < 0.01$). The presence of the significant Voicedness x Phonation-mode interaction in all speakers indicated that whispered speech affected CVAR "voicing" contrast in all speakers. It may be seen from Fig. 11 that in all speakers, CVAR "voicing" contrast was neutralized in whispered speech. T-test for related samples conducted on the CVAR contrast between phonated and whispered speech for each speaker confirmed this statistically (see Table A15).

In addition to the analysis above, *differences* in CVAR between the "voiced" and the "voiceless" consonants was investigated for the effect of Consonant-pair in both phonation modes.[1] Figure 12 shows CVAR contrast between "voiced" and "voiceless" consonants summed over four vowels for each consonant-pair. It can be seen in the figure, that whilst the magnitude of the contrast is greatly reduced in whispered speech, in some consonant-pairs, particularly b/p, it may be available. To test this pattern, three-way mixed ANOVA was conducted on the CVAR contrast for the effects of Consonant-pair (b/p, g/k, z/s, ʤ/ʧ) and Phonation-mode (phonated vs. whispered) as within-subject factors and Speaker as a between-subject factor (see summary of the results in Table A18 in Appendix A). Two main effects were significant: Phonation-mode, $F(1, 12) = 90.07$, $p < 0.01$; Consonant-pair, $F(3, 36) = 10.14$, $p < 0.01$. Main effect of Speaker was marginally significant, $F(3, 12) = 0.83$, $p = 0.05$. Significant interactions were Consonant-pair x Phonation-mode, $F(3, 36) = 16.82$, $p < 0.01$, Speaker x Phonation-mode, $F(3, 12) = 3.55$, $p <$

[1] Similar analysis were conducted for all acoustic measures considered in the present study. Effects of Phonation-mode and Speaker were similar to those reported above for the absolute measures. The effect of Consonant-pair and/or Consonant-pair by Phonation-mode interaction was significant only for C/V amplitude-ratio.

0.05. Three-way interaction of Consonant-pair x Phonation-mode x Speaker was also significant $F(9, 36) = 2.89$, $p < 0.05$. Significant main effect of Voicedness indicated overall reduction of the CVAR contrast in whispered relative to phonated speech as discussed above. Significant main effect of Consonant-pair pointed to the overall differences in CVAR between consonant-pairs. To explore if these differences existed in both phonation modes, the interaction of Phonation-mode x Consonant-pair was investigated further for the simple main effect of Consonant-pair in each mode using one-way repeated measures ANOVA. Simple main effect of Consonant-pair was not significant in neither speech modes. However, as suggested by the significant Consonant-pair x Phonation-mode x Speaker interaction, the effect of Phonation-mode on the CVAR contrast in individual consonant-pairs was mediated by Speaker.

To sum up, C/V amplitude-ratio decreased in whispered relative to phonated speech so that overall C/V amplitude-ratio *contrast* between "voiced" vs. "voiceless" consonants was present only in phonated mode.

2.2.1.7 Summary

Durations of phonetic segments increased in whispered speech. The increase in duration of vowels and consonants was such that the magnitude of all three duration based contrasts — difference in vowel duration, consonant duration and C/V duration-ratio between "voiced" and "voiceless" contexts — was equivalent in both phonation modes. The spectral cue of $F1$-offset dynamics, albeit reduced in whispered relative to phonated speech, was nevertheless generally available. Finally, C/V amplitude-ratio contrast between "voiced" vs. "voiceless" consonants was effectively neutralized in whispered speech. Changes in the acoustic cues associated with whispered speech were generally independent of a type of consonant-pair although the latter had some effect on C/V amplitude-ratio contrast.

2.2.2 Differences in acoustic cues to "voicing" in whispered conversational vs. clear speech

The goal of this section was to explore the effect of clear whispered speech on the variability of acoustic parameters associated with consonant "voicing" across speakers by comparing the effect of speaking-style (clear vs. conversational) on the magnitude of the "voicing" contrasts. All five parameters were analyzed in whispered clear vs. conversational speech. However, because there were no substantial changes to *F1*-offset dynamics and C/V amplitude-ratio parameters, only the analysis of the three time-based parameters — vowel duration, consonant duration and C/V duration-ratio — is presented below. The results of the analysis of *F1*-offset dynamics and C/V amplitude-ratio are given in Appendix A.

2.2.2.1 Speaking rate

Speaking rates for individual speakers in both speaking styles are shown in Fig. 1; corresponding numerical values are given in the table below the figure. All four speakers spoke at a slower speaking rate in clear speech in the range of 2.44–3.93 spc vs 3.13–4.84 spc in conversational speech.

2.2.2.2 Vowel duration

Average vowel durations in "voiced" and "voiceless" contexts in whispered conversational and clear speaking styles are shown in Fig. 13. The average vowel durations in each speaking style for "voiced" and "voiceless" contexts are given in Table 6 (for more details, see Table B1 in Appendix B).

Three-way mixed ANOVA was conducted on the vowel duration for the effects of Speaker as a between-factor and consonant Voicedness ("voiced" vs "voiceless") and Speaking-style (whispered conversational vs. whispered clear) as two within-subject factors (for summary of the results see Table B2 in Appendix B). The analysis found that all three main effects were significant:

Voicedness, $F(1, 124) = 871.47$, $p < 0.01$; Speaker, $F(3, 124) = 14.73$, $p < 0.01$; Speaking-style, $F(1, 124) = 240.20$, $p < 0.01$. All two-way interactions were significant: Speaker x Speaking-style, $F(3, 124) = 18.75$, $p < 0.01$, Speaker x Voicedness, $F(3, 124) = 9.70$, $p < 0.01$, and Speaking-style x Voicedness, $F(1, 124) = 49.08$, $p < 0.01$. The three-way Speaker x Speaking-style x Voicedness interaction was not significant. These results indicated that vowel durations in the "voiced" consonant context (M = 193.86, SD = 39.67) were significantly longer that in the "voiceless" consonant context (M = 152.07, SD = 37.31). Also, vowel durations in whispered clear speech (M = 185.17, SD = 48.62) were significantly longer than in whispered conversational speech (M = 160.76, SD = 34.34).

The significant Speaking-style x Voicedness interaction suggested that the magnitude of the vowel duration "voicing" contrast was not equivalent across the two whispered speech speaking-styles. It may be seen in Fig. 13 that the difference between the "voiced" and the "voiceless" consonants in vowel duration is larger in clear than in conversational whispered speech. T-test for related samples conducted on the vowel duration contrast in whispered clear and whispered conversational speech confirmed this impression: vowel duration *contrast* was significantly larger in clear whispered speech ($M = 51.47$, $SD = 18.06$), compared to conversational whispered speech, ($M = 32.11$, $SD = 14.37$), $t(63) = 7.34$, $p < 0.01$ (Table B4).

The significant interaction of Speaker x Speaking-style indicated that the effect of clear speech on the vowel duration was different in the four speakers. It can be seen in Table 6 that in Speaker 1, unlike in other speakers, average vowel durations in whispered conversational and whispered clear speech are close together. T-test for related samples was conducted on the vowel durations in whispered clear and whispered conversational speech for each speaker. The results determined that vowel durations in whispered clear speech significantly increased in all speakers except Speaker 1 (for detailed comparison results, see Table B3).

The significant interaction of Speaker x Voicedness suggested that in whispered speech overall, the difference in the vowel duration between the "voiced" and "voiceless" consonants was

different in the four speakers. Bonferroni multiple comparisons of the vowel duration *contrast* (i.e. the difference between the "voiced" and "voiceless" consonants) across the four speakers, indicated that in Speaker 2, the vowel duration contrast was significantly larger than in Speakers 3 and 4, at the $p < 0.01$ level.

Finally, we wanted to investigate if the changes in the magnitude of the vowel duration "voicing" contrast in whispered clear speech were associated with particular consonant-pairs. In Fig. 15, the vowel duration differences between "voiced" and "voiceless" contexts are summarized for each consonant-pair in each speaking-style in the four speakers. It can be seen in the figure, that vowel duration contrast is relatively uniform across consonant-pairs in the four speakers. To confirm this statistically, three-way mixed ANOVA was conducted on the vowel duration *contrasts* for the effects of Consonant-pair (b/p, g/k, z/s, ʤ/ʧ) and Speaking-style (whispered conversational vs. whispered clear) as within-subject factors and Speaker as the between-subject factor (for details, see Fig. B5). Only two main effects were significant (Speaking-style: $F(1, 12) = 41.23$, $p < 0.01$; Speaker, $F(3, 12) = 11.45$, $p < 0.01$). The main effect of Consonant-pair was not significant. There were no significant interactions. These results indicated that the effect of clear speech on the vowel duration "voicing" contrast was independent of the consonant-pair.

To summarize, vowel durations in clear whispered speech were longer relative to conversational whispered speech. At the same, the difference between vowel durations in "voiced" and "voiceless" consonant contexts was not increased in clear whispered speech.

2.2.2.3 Consonant duration

Average consonant durations in "voiced" and "voiceless" contexts in whispered conversational and clear speaking styles are shown in Fig. 16. The consonant durations in each speaking style for "voiced" and "voiceless" contexts are given in Table 7 (for more details, see Table B6 in Appendix B).

Three-way mixed ANOVA was conducted on the consonant duration for the effects of Speaker

as the between-factor and consonant Voicedness ("voiced" vs "voiceless") and Speaking-style (whispered conversational vs. whispered clear). See summary of the results in Table B7 in Appendix B. The analysis determined that all three main effects were significant: Voicedness, $F(1, 124) = 184.42$, $p < 0.01$, Speaker, $F(3, 124) = 39.05$, $p < 0.01$, and Speaking-style, $F(1, 124) = 109.92$, $p < 0.01$. There was also one significant interaction of Speaking-style x Speaker, $F(3, 124) = 71.59$, $p < 0.01$. The three-way Speaker x Speaking-style x Voicedness interaction was also significant, $F(3, 124) = 2.73$, $p < 0.05$. The results revealed that "voiced" consonants ($M = 99.49$, $SD = 36.15$) were shorter than the "voiceless" consonants ($M = 121.97$, $SD = 35.83$). Consonants were longer in whispered clear ($M = 120.53$, $SD = 45.93$) than in whispered conversational ($M = 100.85$, $SD = 23.17$) speech. The absence of the significant Speaking-style by Voicedness interaction indicated that consonants in both "voiced" and "voiceless" contexts were lengthened to a similar degree and that the "voiced" vs. "voiceless" consonant duration contrast remained equivalent in the two whispered speech speaking-styles. At the same time, the significant two-way Speaking-style x Speaker interaction suggested that clear speaking-style affected consonant duration differently in different speakers. Pairwise comparisons of the consonant durations between whispered clear and whispered conversational speech (t-test for related samples) was conducted for each speaker. The comparisons determined that consonants lengthened in clear speech only in Speakers 1 and 2 (for details, see Table B8).

The significant three-way interaction of Speaking-style x Voicedness x Speaker suggested that the effect of clear speech on the consonant duration "voicing" contrast was different in the four speakers. To explore this interaction more closely, separate two-way repeated measures ANOVAs were conducted for the effects of Voicedness and Speaking-style in each speaker. In Speaker 1, both main effects but not the interaction were significant (Voicedness: $F(1, 31) = 31.37$, $p < 0.01$, Speaking-style: $F(1, 31) = 45.74$, $p < 0.01$). In Speaker 2, both main effects and the interaction were significant (Voicedness, $F(1, 31) = 63.07$, $p < 0.01$; Speaking-style, $F(1, 31) = 108.27$, $p < 0.01$, Voicedness x Speaking-style: $F(1, 31) = 7.73$, $p < 0.01$). In

Speaker 3, only the main effect of Voicedness was significant, $F(1, 31) = 77.48$, $p < 0.01$. In Speaker 4, main effect of Voicedness and the interaction were significant (Voicedness: $F(1, 31) = 64.09$, $p < 0.01$, Voicedness x Speaking-style: $F(1, 31) = 9.39$, $p < 0.01$). The presence of the Voicedness by Speaking-style interaction in Speakers 2 and 4 indicated that clear whispered speech had an affect on the consonant duration "voicing" contrast in these two speakers. It may be seen from Fig. 17 that consonant duration contrast is smaller for Speakers 2 and 4 in whispered clear speech relative to whispered conversational speech. T-test for related samples confirmed the differences in the consonant duration contrast in two whispered speech conditions for these two speakers. In Speaker 2, the contrast of consonant duration in whispered clear speech ($M = 17.01$, $SD = 27.01$) was significantly smaller compared to whispered conversational speech ($M = 33.05$, $SD = 11.74$), $t(15) = -2.37$, $p < 0.05$. In Speaker 4, the contrast of consonant duration in whispered clear speech ($M = 17.37$, $SD = 16.09$) was significantly smaller compared to whispered conversational speech ($M = 27.44$, $SD = 14.94$), $t(15) = -3.13$, $p < 0.01$.

Thus, in whispered clear speech, consonants were longer compared to whispered conversational speech. However, the changes in consonant durations did not increase the difference between "voiced" and "voiceless" consonants in whispered clear speech.

2.2.2.4 Consonant-to-vowel duration-ratio

Average consonant-to-vowel duration-ratios (CVDR) for "voiced" and "voiceless" contexts in whispered conversational and clear speaking styles are shown in Fig. 18. The average CVDR in each speaking style for "voiced" and "voiceless" contexts are given in Table 8 (for more details, see Table B9 in Appendix B).

Three-way mixed ANOVA was conducted on the CVDR for the effects of Speaker as the between-factor and consonant Voicedness ("voiced" vs "voiceless") and Speaking-styl(see summary of the results in Table B10 in Appendix B). The analysis revealed significant main effects of

Voicedness, $F(1, 124) = 393.25$, $p < 0.01$, Speaker, $F(3, 124) = 19.73$, $p < 0.01$, and Speaking-style, $F(1, 124) = 10.76$, $p < 0.01$. There were significant two-way interactions of Speaker x Speaking-style, $F(3, 124) = 54.22$, $p < 0.01$, and Speaker x Voicedness, $F(3, 124) = 8.21$ $p < 0.01$. The three-way Speaker x Speaking-style x Voicedness interaction was also significant, $F(3, 124) = 10.85$, $p < 0.01$. These results revealed that CVDR was contrastive in "voiced" vs. "voiceless" consonant contexts: in the "voiced" context ($M = 0.53$, $SD = 0.20$), it was significantly smaller than in the "voiceless" context ($M = 0.86$, $SD = 0.35$). Also, CVDR was generally larger in clear ($M = 0.72$, $SD = 0.38$) than in conversational ($M = 0.68$, $SD = 0.28$) whispered speech. Furthermore, the results indicated that average CVDR varied between the four speakers. The absence of the significant Voicedness x Speaking-style suggested that clear speaking-style had a similar effect on CVDR in the two "voicing" contexts and that overall distinction in CVDR between "voiced" vs. "voiceless" contexts was equivalent in both speech modes.

The significant Speaker x Speaking-style interaction indicated that the effect of clear speaking-style on CVDR was different in the four speakers. It can be seen from Fig. 19 that in Speakers 1 and 2, CVDR increased whereas in Speakers 3 and 4, CVDR decreased. Pairwise comparisons confirmed that these differences were significant (see details in Table B11 in Appendix B).

The significant interaction of Speaker x Voicedness revealed that the CVDR contrast between "voiced" and "voiceless" consonant contexts was generally different in the four speakers. Furthermore, significant three-way interaction of Speaker x Voicedness x Speaking-style suggested that this distinction was differently affected by clear speech in the four speakers. To investigate this interaction further, a series of two-way repeated measures ANOVAs was conducted for the effects of Voicedness and Speaking-style in each speaker. Except Speaker 3, both main effects and the interaction were significant in all speakers (Speaker 1: Voicedness, $F(1, 31) = 100.38$, $p < 0.01$; Speaking-style, $F(1, 31) = 37.74$, $p < 0.01$; Voicedness x Speaking-style, $(1, 31) = 14.29$, $p < 0.01$; Speaker 2: Voicedness, $F(1, 31) = 185.80$, $p < 0.01$; Speaking-style, $F(1, 31) = 42.04$, $p < 0.01$, Voicedness x Speaking-style, $(1, 31) = 4.29$, $p < 0.05$; Speaker 3: Voicedness,

F(1, 31) = 98.87, $p < 0.01$; Speaking-style, F(1, 31) = 63.28, $p < 0.01$; Speaker 4: Voicedness, F(1, 31) = 69.55, $p < 0.01$; Speaking-style, F(1, 31) = 138.19, $p < 0.01$, Voicedness x Speaking-style, F(1, 31) = 11.38, $p < 0.01$). Thus, with the exception of Speaker 3, the magnitude of the CVDR "voicing" contrast was affected by clear speech speaking-style (see Fig. 19). T-test for related samples confirmed that in Speaker 1, the CVDR contrast in whispered clear speech ($M = 0.50$, $SD = 0.23$) was significantly larger compared to whispered conversational speech ($M = 0.30$, $SD = 0.15$). Similarly, in Speaker 2, the CVDR contrast in whispered clear speech ($M = 0.46$, $SD = 0.18$) was significantly larger compared to whispered conversational speech ($M = 0.39$, $SD = 0.16$). At the same time, in Speaker 4, the CVDR contrast in whispered clear speech ($M = 0.21$, $SD = 0.13$) was significantly *smaller* compared to whispered conversational speech ($M = 0.34$, $SD = 0.24$). See test results in Table B13.

Next, in each speaking style, we wanted to investigate relative difference in the CVDR distinction in "voiced" vs "voiceless" contexts between speakers. To explore this, two separate repeated measures ANOVAs were conducted for the effects of Voicedness and Speaker in whispered conversational and whispered clear speaking styles. In conversational whispered speech, both main effects were significant but the Voicedness x Speaker interaction was not: Voicedness, F(1, 247) = 135.02, $p < 0.01$; Speaker, F(3, 247) = 5.92, $p < 0.01$; Voicedness x Speaker, F(3, 247) = 1.79 (ns). In clear whispered speech, both main effects and the Voicedness x Speaker interaction were significant: Voicedness, F(1, 247) = 119.67, $p < 0.01$; Speaker, F(3, 247) = 64.70, $p < 0.01$, Voicedness x Speaker, F(3, 247) = 5.86. The absence of the Voicedness x Speaker interaction in whispered conversational speech suggested that only in clear speech, there were differences between speakers in the CVDR distinction between "voiced" and "voiceless" consonant contexts. In Fig. 19, it can be seen that in clear speech of Speakers 1 and 2, CVDR in "voiced" vs. "voiceless" contexts are further apart than in Speakers 3 and 4. Bonferroni-corrected multiple comparisons of the CVDR contrasts across speakers established that in clear speech, CVDR contrast was significantly larger in Speaker 1 (M = 0.50, SD = 0.23) and Speaker

2 (M = 0.46, SD = 0.18) than in Speaker 3 (M = 0.22, SD = 0.09) and Speaker 4 (M = 0.21, SD = 0.13). For more details on the comparisons, see Table B12 in Appendix B.

Finally, we wanted to investigate if the changes in the magnitude of the CVDR "voicing" contrast in whispered clear speech were associated with particular consonant-pairs. In Fig. 20, CVDR differences between "voiced" and "voiceless" contexts are summarized for each consonant-pair in each speaking-style in the four speakers. It can be seen in the figure, that although the contrast magnitude was different in the four speakers, within an individual speaker, it was relatively equivalent. To confirm this statistically, three-way mixed ANOVA was conducted on the CVDR *contrasts* for the effects of Consonant-pair (b/p, g/k, z/s, ʤ/ʧ) and Speaking-style (whispered conversational vs. whispered clear) as within-subject factors and Speaker as the between factor (see Table B14 for details). The results determined that effect of Consonant-pair was significant, $F(3, 9) = 7.58$ $p < 0.01$. There was also a significant interaction of the Speaker x Speaking-style, $F(3, 12) = 6.51$, $p < 0.01$. Bonferroni-corrected multiple comparisons of the CVDR contrasts across consonant-pairs indicated that CVDR contrasts in ʤ/ʧ (M = 0.43, SD = 0.23) and b/p (M = 0.36, SD = 0.20) were significantly greater than in z/s (M = 0.24, SD = 0.15). The absence of Speaking-style x Consonant-pair interaction suggested that effect of clear speaking-style was independent of the consonant-pair identity.

To summarize, C/V duration-ratio in clear speech did not change consistently across speakers. In conversational whispered speech, the magnitude of the C/V duration-ratio "voicing *contrast* was equivalent in all speakers, in clear whispered speech, the contrast was significantly more prominent in two speakers than in the other two. At the same time, overall magnitude of the C/V duration-ratio " contrast remained equivalent in both speaking-styles. However, in one speaker, the C/V duration-ratio contrast was enhanced in clear whispered speech.

2.2.2.5 Summary

Three duration-based acoustic parameters — vowel duration, consonant duration and C/V duration-ratio — changed in clear whispered relative to conversational whispered speech. Although changes in vowel durations were more consistent across speakers than changes in consonant durations, overall, both parameters increased in clear speech. At the same time, overall C/V duration-ratio was equivalent in both speaking styles. Regarding the effect of clear speech on the magnitude of the acoustic contrasts, clear speech led to the overall enhancement of the vowel duration contrast but not consonant duration or C/V duration-ratio contrasts.

2.3 Discussion

Considering the absolute measures in conversational whispered vs. phonated speech first, our results showed that durations of phonetic segments in both vowels and consonants generally increased in whispered speech. At the same time, first-formant offset-dynamics and consonant-to-vowel amplitude ratio were generally reduced in whispered speech. In whispered clear speech, further lengthening of vowel and consonant phonetic segments was observed. At the same time, no particular changes in the patterns of $F1$-offset and CVAR were noticed. Comparison of the relative measures, i.e. differences in the acoustic parameters between "voiced" and "voiceless" consonants in conversational phonated and whispered speech revealed that overall duration-based acoustic contrasts – vowel duration, consonant duration and consonant-to-vowel duration-ratio were equivalent in both phonation modes. Non-durational acoustic contrasts were generally less pronounced in conversational whispered speech: $F1$-offset dynamics contrast was produced by two out of four speakers (Speaker 2 and Speaker 4) whereas CVAR contrast was effectively neutralized. However, the consonant-pair b/p proved to be a notable exception. In whispered clear speech, the vowel duration contrast was further enhanced, whereas consonant duration and CVDR remained at the same level as in whispered conversational speech. Speaker identity was

not a factor in maintaining vowel and consonant duration contrast in conversational whispered vs. conversational phonated speech. However, relative prominence of the combined consonant and vowel duration measure, CVDR, between the two phonation modes was different in the four speakers. Specifically, CVDR contrast was enhanced in Speaker 2 in conversational whispered speech relative to conversational phonated speech whereas in the other three speakers, it remained at the same level. Similarly, in clear conversational speech,the increase in vowel duration contrast and maintaining of the consonant duration contrast was equivalent in the four speakers. At the same time, relative prominence of the CVDR contrast in conversational vs. clear whispered speech was different in the four speakers. In Speaker 1, CVDR contrast was significantly enhanced. Speaker 2 demonstrated a non-significant trend for CVDR contrast increase in clear whispered speech. For the other two speakers, CVDR contrast remained at the same level in the two whispered speech conditions. Considering the relative prominence of the acoustic cues in different phonation modes (phonated vs. whispered) discussed above, speaking-styles (conversational and clear whispered speech) and speakers, the following predictions regarding perception of consonant "voicing" were made: (1) mode: perception accuracy of consonant "voicing" identification will decrease in whispered relative to phonated speech due to the reduction or complete elimination of non-durational cues, $F1$-offset dynamics and CVAR. At the same time, despite the overall decrease, relatively accurate perception of consonant "voicing" in whispered speech should be possible due to retention of the duration-based cues; (2) speaking-style: in clear whispered speech, perception of consonant "voicing" was expected to be more accurate as the listeners might make use of the enhanced vowel duration cues and, in some cases, CVDR cues; (3) speakers: in conversational whispered speech, perception of consonants produced by Speaker 2 was expected to be more accurate due to the enhanced (relative to phonated speech) CVDR cue. Furthermore, in clear whispered speech, listeners should benefit from the enhanced CVDR cues in Speakers 1 and 2 and, thus, on their perception of Speaker 1 and 2's consonants might be more accurate than on the consonants of Speakers 3 and 4.

Chapter 3

Perception of "voicing" in whispered vs. phonated consonants

3.1 Materials and methods

3.1.1 Study 2A: Perception of consonants in conversational phonated and whispered speech

3.1.1.1 Listeners

Listeners were 5 normal hearing adults, ranging in age from 18 to 45 years old. They were monolingual speakers of American English born and raised in New York City metropolitan area. They reported no history of speech or hearing disorders. Listeners were paid $10 per hour for their participation. Prior to testing, all listeners signed a consent form approved by the IRB of the CUNY Graduate Center and successfully underwent a hearing screening at octave intervals from 500–4000 Hz at 25 dB HL. The listeners were recruited from the CUNY Graduate Center community and through Craigslist — an online classified advertisement.

3.1.1.2 Stimuli

The stimuli analyzed in Study 1 served as test materials in this study. Thus, the stimuli were nonsense disyllables *habVC* embedded in a carrier sentence 'I said _____ eight times' produced by 4 speakers in two conditions: conversational phonated and conversational whispered speech. The V (vowel) and the C (consonant) in the test word *habVC* were one of 4 vowels /ɛ, æ:, ɑ:, ʌ/ followed by one of 8 consonants /b, p, g, k, z, s, ʤ, ʧ/. This provided 32 vowel-consonant combinations each represented by two different tokens. The consonants included two stop pairs /b-p, g-k/, one fricative pair /z-s/ and one affricate pair /ʤ-ʧ/. The vowels constituted four contrasting pairs that are spectrally adjacent in $F2$-$F1$ vowel space but contrastive in intrinsic duration. The choice of vowels was based on their relatively high first formant, $F1$, to allow $F1$-related information in judging final consonant "voicing" in whispered speech to be more readily available.

3.1.1.3 Procedure

Five listeners were tested on their perception of both conversational phonated and whispered speech in a repeated measures design. The testing was conducted in a sound-treated booth, in which listeners were seated in front of a computer monitor and a computer mouse. Paradigm© software (Tagliaferri, 2005) was used to present stimuli and collect listeners' responses. The stimuli were presented binaurally via Sennheiser HD565 Ovation headphones at 75 dB SPL.

Familiarization The experiment began with three familiarization blocks using stimuli produced by a monolingual female speaker of American English not included in the test materials. The first familiarization block presented phonated conversational tokens with per-trial feedback and consisted of 32 trials covering all combinations of consonants and vowels. In the second familiarization block, 32 whispered clear-speech trials were presented, also with per-trial feedback. Finally, in the last familiarization block, the same 32 whispered trials were presented but feedback was provided only at the end of the block to mimic the design of the main experiment.

In all three familiarization blocks and subsequent experimental blocks, subjects were asked to identify consonants by clicking the mouse on the appropriate response. The list of the response options consisted of 8 possible final consonants [b, p, g, k, z, s, ʤ, ʧ] arranged in two columns. Each row presented a "voicing" pair with the "voiced" consonant appearing in the left column and the "voiceless" consonant in the right column. Affricate consonants [ʤ, ʧ] were written as "dge" and "tch", correspondingly and were accompanied with an example word, e.g. "*dge* as in judge" or "*tch* as in church".

Testing Subjects listened to 2 tokens — arbitrarily labeled token 1 and token 2 — of each of 32 stimuli, i.e. 4 vowels combined with 8 consonants, from each speaker. A token represented one of the two physical instances of a particular vowel-consonant combination. Each token was presented to a listener 3 times in each condition, i.e. a total of 192 tokens from each speaker. This constituted a total of 768 tokens per condition across 4 speakers giving 1536 trials in the entire test. Trials were blocked by speaker into 16 blocks, 96 trials each. The experiment was conducted in two sessions of 8 blocks. The first session consisted of phonated blocks only, while in the second session only whispered blocks were presented. Presentation of blocks was randomized within each session. Each session lasted about 60–70 minutes. Subjects were required to take a break after 4 blocks and between the sessions; they were also encouraged to take short breaks, if needed, between the blocks. Both sessions were conducted on the same day with a 20 minute break between them. The entire test took about 2.5–3 hours.

Data analysis To establish consonant "voicing" perception accuracy, percentage of correct responses was computed across the consonants and speakers for each condition: conversational phonated and whispered speech. In addition, accuracy scores were obtained for each consonant-pair and speaker individually. The responses were tallied as follows: each consonant in the combination with a particular vowel was judged by a listener 6 times based on responses to 3 repetitions of the 2 tokens per speaker in each condition. Thus, combined together, each

consonant-voicing pair yielded 12 judgments. Across all four vowels, each consonant-voicing pair received 48 judgments including 24 judgments of the consonants preceded by intrinsically short vowels /ɛ, ʌ/ and 24 judgments of the consonants preceded by intrinsically long vowels /æ:, ɑ:/. To investigate the pattern of confusions in consonant identification beyond "voicing", a confusion matrix was generated for each condition. The scores were computed across all the consonants similar to those described above but this time taking into account errors on place and manner of articulation in addition to consonant "voicing" errors.

3.1.2 Perception of consonants in whispered conversational and clear speech

3.1.2.1 Procedure

The test procedure of this study was similar to Study 2A described above with two exceptions: the number of listeners and the layout of the main experiment.

Listeners Fifteen listeners were normal hearing adults, ranging in age from 18 to 45 years old. They were tested on their perception of consonants in two conditions: conversational whispered and clear whispered speech.

Testing Presentation of experimental stimuli was blocked by speaker and condition with blocks randomized within each of the two sessions. The sessions differed only in the tokens that constituted the stimuli. As mentioned above, the tokens were assigned to the sessions arbitrarily. Thus, in the initial session, all stimuli were repetitions of token 1 from all 4 speakers; in the second session, they were all repetitions of token 2.

Data analysis The tally of responses was calculated the same way as described above for Study 2A. In addition, because of the expected bias towards "voiceless" consonants in whispered condition (Tartter, 1989), the accuracy scores were transformed into $A^{'}$-scores for each consonant. To compute $A^{'}$, all responses were first scored dichotomously based on correct perception of

"voicing" independently of manner and place of articulation. Using the terminology of Signal Detection Theory, responses that correctly detected the presence of "voicing" or correctly detected its absence were categorized as "hits" and "correct rejections", correspondingly while the responses that detected "voicing", when it was absent and the responses that failed to detect "voicing", when it, in fact, was present, were classified as "false alarms" and "misses, correspondingly. $A^{'}$-scores were computed according to the formula:

$$A^{'} = 0.5 + \frac{(H - FA)(1 + H - FA)}{4H(1 - FA)}$$

where H is proportion of "hits" and FA is proportion of "false alarms". An $A^{'}$-score of 0.5 indicated that subjects did not "discriminate" between "voiced" and "voiceless" members of a "voicing"- pair. Also, to allow for comparison of the results with other studies, the scores were transformed to rationalized arcsine units, or RAU (Studebaker, 1985). Transformation to RAU allows to linearize the data in relation to variance. To calculate RAU-scores, the number of correct responses and the total number of responses were arcsine-transformed:

$$ARC = \arcsin\left(\sqrt{\frac{C}{T+1}}\right) + \arcsin\left(\sqrt{\frac{C+1}{T+1}}\right)$$

where C is the number of correct responses and T is the total number of responses. The ARC value was then used to compute the RAUs:

$$RAU = \frac{146}{\pi(ARC)} - 23$$

The RAU-scale operates between approximately -23 and 123. Between 15–85, RAU scores are equivalent to percent correct scores; below and above 15 and 85, correspondingly, RAU-scale expands beyond percent-correct range.

Finally, to explore the effect of clear whispered speech on the perception of consonant "voicing", a three-way repeated measures ANOVA was performed on perception accuracy — ex-

pressed as A'-scores — with Speaking-style, Speaker, and Consonant-pair as independent variables.

3.2 Results

3.2.1 Consonant perception in conversational phonated and whispered speech

Overall perception of consonants in phonated conversational speech was highly accurate, 98.17%. Table 14a shows confusion matrix of consonant perception in phonated speech summed over four speakers. The scores are given as numbers of actual responses along with corresponding percentages. The numbers on the diagonal correspond to correct responses. As it can be seen from the table, the errors were generally limited to incorrect perception of "voicing" and were mostly due to misperception of the b/p pair: 8.59% of the time /b/ was misclassified as /p/. A few errors comprising less than 1% were also observed for g/k and ʤ/ʧ pairs. Errors were spread rather evenly across the four speakers' productions (see Table 14b).

In whispered conversational speech, consonants were identified with overall accuracy of 83.56% accuracy. Table 15a shows error distributions of consonants in conversational whispered speech summed over all speakers. Similar to phonated speech, most errors were misidentifications of bilabial stops, /b/ vs. /p/. The pattern of responses to stop consonants indicated the expected bias of the listeners towards "voiceless" consonant response alternatives in whispered speech, e.g. listeners incorrectly classified /b/ as /p/ more often than vice versa. However, for fricatives and affricates, this pattern was reversed: "voiceless" consonants were categorized as "voiced" more often. Errors made outside "voicing" pairs, i.e. errors on place and manner of articulation were non-systematic and did not exceed 1% for any consonant. Also, distributions of errors across speakers was quite uniform (Table 15b).

3.2.2 Consonant perception in conversational and clear whispered speech

Results of consonant perception in conversational whispered speech (Study 2B, N=15) generally followed the pattern described in Study 2A for the smaller set of listeners. Overall perception of consonants in conversational whispered speech was 85.34%. A confusion matrix in Table 17a displays accuracy of consonant identification in conversational whispered speech expressed as percentage of responses across all speakers. Confusion matrices summarizing the results for individual speakers in whispered conversational and clear speech are given in Table C1 and Table C2, correspondingly, in Appendix C. Listeners made most errors on bilabial stops, 25.90% and 13.33%, for /b/ and /p/, correspondingly. The response bias towards "voiceless" consonants was observed across all consonant-pairs except affricates where the trend was the opposite: "voiceless" consonants caused more "voiced" responses than vice versa. As it can be seen from Table 17a, errors on place and manner of articulation did not exceed 1% for any consonant.

For further analysis, perception scores were converted into A'- scores on consonant-pairs. In addition, to allow for comparison of the results with other studies, the scores were transformed to rationalized arcsine units (RAU). The average consonant intelligibility scores (A'- scores) are plotted in Fig. 21 separately for four speakers and the four consonant-pairs in each speaking-style. Corresponding numerical values along with the scores for individual consonant-pairs and absolute (clear$-$conv) and proportional gain ((clear$-$conv)/conv) are given in Table 18a. It can be seen in the table that there was substantial variability across the speakers, both in the level of intelligibility in conversational whispered speech and in the amount of benefit provided by clear whispered speech. In line with the predictions made based on the production data, Speakers 1 and 2 were not only generally more intelligible but, unlike Speakers 3 and 4, they also afforded intelligibility benefit in clear speech. A three-way repeated measures ANOVA was conducted for the effects of Speaker, Speaking-style and Consonant-pair on consonant "voicing" intelligibility expressed as A'-scores (see ANOVA summary in Table C3 in Appendix C). The results revealed that the accuracy of consonant "voicing" identification was significantly influenced by Speaking-

style $F(1, 14) = 17.19$, $p < 0.01$, Speaker, $F(2.47, 34.63)$[1] = 35.68, $p < 0.01$, and Consonant-pair, $F(1.45, 20.30) = 8.42$, $p < 0.01$. All two-way interactions were significant: Speaking-style x Consonant-pair, $F(3, 42) = 19.77$, $p < 0.01$, Speaking-style x Speaker $F(3, 42) = 51.44$, $p < 0.01$ and Speaker x Consonant-pair $F(9, 126) = 18.85$, $p < 0.01$. The three-way interaction of Speaker x Speaking-style x Consonant-pair was also significant, $F(9, 126) = 14.70$, $p < 0.01$.

These results indicate that consonant "voicing" was perceived better in clear whispered speech ($M = 0.93$, $SD = 0.072$) than in conversational whispered speech ($M = 0.91$, $SD = 0.072$) although the amount of benefit provided by clear speech was small. The analysis also revealed that the accuracy of "voicing" perception was affected by speaker identity and that it was different in the four consonant-pairs considered in the present study. Significant Speaking-style x Speaker interaction suggested that the clear speech benefit differed across speakers. It can be seen in Fig. 21a, that in conversational speech, intelligibility of the four speakers was relatively uniform whereas in clear speech, Speakers 1 and 2 appear to be more intelligible than Speakers 3 and 4. To support these impressions statistically, two one-way repeated measures ANOVAs were conducted for simple main effect of Speaker in each speaking style. Although, there was a simple main effect of Speaker in conversational whispered speech, $F(2.42, 142.58) = 3.14$, $p < 0.05$[1], neither of the pairwise comparisons with Bonferroni correction was significant. However, there was a non-significant trend for Speaker 2 to yield higher perceptual scores than the other three speakers. In clear whispered speech, main effect of Speaker was also significant, $F(1, 59) = 160.513$, $p < 0.05$. Post hoc tests confirmed the pattern in Fig. 21a: perceptual scores on the productions of Speaker 1 and Speaker 2 were significantly better than those on productions of Speakers 3 and 4.

The significant Speaking-style x Consonant-pair interaction indicated that the clear speech benefit differed across the four consonant-pairs. It can be seen in Fig. 21b, that in conversational speech, b/p caused most perceptual confusion whereas in clear speech, z/s was most challenging. Two one-way ANOVAs were conducted for simple main effect of Consonant-pair separately in

[1]Assumption of sphericity was not tenable; Greenhouse-Geisser correction for degrees of freedom was applied

conversational and clear speech. The Consonant-pair effect was significant in both speaking styles (conv: $F(2.47, 145.73)^1 = 12.31$, $p < 0.01$; clear: $F(2.34, 138.214)^1 = 20.73$, $p < 0.01$). Multiple comparisons confirmed that consonant-pair b/p caused significantly more confusion than g/k, z/s, and ʤ/ʧ pairs in conversational whispered speech. In clear conversational speech, perception of the z/s pair was significantly worse than all other pairs. All other differences were non-significant.

Finally, the three-way Speaking-style x Consonant-pair x Speaker interaction suggested that effect of clear speech on intelligibility of different consonant-pairs was not uniform across the four speakers. Figure 22 shows intelligibility scores for the four consonant-pairs in conversational and clear whispered speech of each speaker. Corresponding numerical values are given in Table 18b. It can be seen from the figure, that changes in the intelligibility of consonant-pairs associated with clear speech were non-systematic across speakers. However, it appears that perception of stop-pairs b/p and g/k improved whereas perception of z/s decreased in most speakers. To further investigate this relationship, a series of two-way repeated measures ANOVAs was conducted for each speaker for the effects of Consonant-pair and Speaking-style. In Speaker 1, 2 and 4, both main effects and the interaction were significant [Speaker 1: Speaking-style, $F(1, 14) = 106.62$, $p < 0.01$, Consonant-pair, $F(3, 42) = 42.31$, $p < 0.01$, Speaking-style x Consonant-pair, $F(3, 42) = 29.94$, $p < 0.01$; Speaker 2: Speaking-style, $F(1, 14) = 131.54$, $p < 0.01$, Consonant-pair, $F(3, 42) = 5.40$, $p < 0.01$, Speaking-style x Consonant-pair, $F(3, 42) = 15.87$, $p < 0.01$. Speaker 4: Speaking-style, $F(1, 14) = 13.35$, $p < 0.01$, Consonant-pair, $F(3, 42) = 7.11$, $p < 0.01$, Speaking-style x Consonant-pair, $F(3, 42) = 9.66$, $p < 0.01$. In Speaker 3, only one main effect and the interaction were significant: Consonant-pair, $F(3, 42) = 12, 32$, $p < 0.01$, Speaking-style x Consonant-pair, $F(3, 42) = 10.97$, $p < 0.01$. A series of pairwise comparisons (t-test for related samples) was conducted for each speaker to establish if individual consonant-pairs were perceived better in clear whispered relative to conversational whispered speech. The results revealed that for Speaker 1, only perception of stop pairs b/p and g/k signif-

icantly improved in clear whispered speech. In Speaker 2, perception of all pairs except z/s was significantly better in clear speech; on the other hand, perception of the z/s pair was significantly worse. In Speaker 3, perception of b/p pair significant improved whereas perception of all other pairs was significantly worse. Finally, in Speaker 4, clear speech significantly affected only the z/s and ʤ/ʧ pairs; perception of both was worse in clear speech. These results suggest that clear whispered speech provided consistent perceptual benefit primarily for bilabial stops. At the same time, effect of clear speech on perception of "voicing" in z/s pair was generally detrimental.

3.3 Summary

Summarizing the results of perception of consonant "voicing" in both phonated and whispered speech, it was established that in conversational phonated speech, listeners made very few errors. However, the encountered errors were systematic and were primarily limited to misperception of "voicing" in bilabial stops. In conversational whispered speech, overall perceptual accuracy decreased by about 15%. As with phonated conversational speech, bilabial stops caused most confusion. Overall, clear whispered speech generally provided a limited intelligibility benefit (2%), however the effect of clear whispered speech varied greatly across speakers from -2% ... to 8%. Clear speech benefit afforded by two speakers, Speaker 1 and 2, was in agreement with the predictions made based on the production data. Perceptual benefit, when available, was mostly associated with the b/p pair. Across the speakers, there was substantial variability both in the level of intelligibility in conversational whispered speech and in the amount of benefit provided by clear speech.

Chapter 4

Consonant production-perception relationship

The goal of this section was to investigate how well a combination of acoustic parameters associated with consonant "voicing" can classify phonated and whispered consonants into "voiced" and "voiceless" categories. Prediction success of the logistic regression analysis was then compared with the observed perception accuracy. In addition, correlation analyzes investigated the relationship between prominence of an acoustic contrast and accuracy of consonant "voicing" perception.

4.1 Logistic regression analysis

The acoustic measurements from Study 1 were analyzed using logistic regression (Jiang et al., 2006), where speech tokens were separately classified as either "voiced" "or voiceless" using a combination of acoustic properties. In order to avoid the multicollinearity problem (strongly correlated predictor variables), only C/V duration-ratio (CVDR) was used as a duration-based parameter in the regression. We chose CVDR because it represents a combined measure of vowel duration and consonant duration "voicing" cues and because in the past, it was shown to drive

listeners' judgments of consonant "voicing" judgments (Port & Dalby, 1982). For phonated speech, the analysis was conducted with three continuous predictor variables — CVDR, $F1$-offset dynamics and CVAR — and a categorical variable, consonant-pair; for whispered speech, the analysis was conducted with the same set of variables except CVAR which, as we established, was unavailable in whispered speech as a cue to consonant "voicing". All continuous variables were normalized by speaking-style, speaker, consonant-pair and vowel prior to the analysis.

The analysis was conducted in three steps. First, a separate logistic regression model was developed for the consonants in phonated conversational speech:

$$logit = ln\left(\frac{prob}{1-prob}\right) = \alpha + \beta_1 Acou_1 + \beta_2 Acou_2 + \beta_3 Acou_3 + \beta_4 ConsPair$$

where *logit* is a natural logarithm of the odds ratio, *prob* is the probability of a token being "voiced", α is a constant, β_i is a weighting coefficient, $Acou_i$ is an acoustic feature and $ConsPair$ is consonant-pair. Next, a separate logistic regression model was constructed for the consonants in whispered *conversational* speech which served as the training data set. Finally, a model developed based on the training set (whispered conversational speech data) was applied to whispered *clear* speech data for the purposes of cross-validation.

4.1.1 Results

Table 19 shows the results of logistic regression analysis of "voicing" in phonated consonants. The results of the logistic regression analysis indicated that the linear combination of the regression coefficients classified consonants into "voiced" and "voiceless" categories with 99.2% accuracy. Due to this very high rate of prediction success, further analysis of prediction success for individual speakers and individual consonant-pairs was not conducted.

The relatively large value of the regression coefficient for CVDR compared to $F1$-offset dynamics and CVAR suggested that C/V duration-ratio was a primary contributor to consonant

"voicing" prediction. At the same time, large standard errors suggested that there was a high correlation between the parameters. Further analysis revealed that there was a strong negative correlation between C/V duration-ratio and $F1$-offset dynamics ($r = -0.86$). Also, there was a moderately strong correlation between C/V duration-ratio and C/V amplitude-ratio ($r = 0.67$). To further investigate the contribution of individual acoustic parameters to "voicing" category predictions for consonants in phonated speech, logistic regression analyses were conducted separately with each acoustic parameter as a single predictor. The results indicated that the highest prediction success was achieved with C/V duration-ration (98%), followed by $F1$-offset dynamics (78%) and C/V amplitude-ratio (77%). The contribution of the consonant-pair was not significant in either of the models.

The results of the logistic regression analysis of "voicing" in consonants in whispered conversational speech are given in Table 20. It can be seen from the table that both CVDR and $F1$-offset dynamics significantly contributed to the classification prediction at the $p < 0.05$ level. The linear combination of the two parameters classified consonants into "voiced" and "voiceless" categories with the prediction success rate of 94%. To further explore the contribution of individual acoustic parameters to "voicing" category predictions for consonants in whispered conversational speech, two additional logistic regressions were carried out separately with each acoustic parameter as a single predictor to consonant "voicing" classification. The results indicated that CVDR-based model classified consonants with 93% prediction success rate whereas the model based on the $F1$-offset dynamics classified consonants into "voicing" categories with the prediction success rate of 68%. The contribution of the consonant-pair was not significant in either of the models.

In order to test whether the model constructed based on the whispered conversational speech data was an accurate one, it was fit to data from whispered *clear* speech productions. This new set of consonants was classified by the model with 94% accuracy. These results indicate that C/V duration-ratio could accurately predict "voicing" of a whispered consonant.

Finally, we wanted to investigate whether the results of the statistical classification, i.e. prediction success of the logistic regression model, corresponded to the pattern of the relative speaker and/or consonant-pair intelligibility. Table 21 shows prediction success rate of the logistic regression model for whispered conversational speech (first column) along with A'-transformed consonant intelligibility scores from the perception test (third column). The second and the fourth column show respective ranks of the prediction success and A'-scores, correspondingly. The scores were ranked from 1 through 16 across all speakers and consonant-pairs. It can be seen from the table, that Speakers 1 and 2 were both associated with the highest ranked prediction-success rates *and* with the highest ranked intelligibility scores. Table 22 shows a similar summary for whispered clear speech. On the other hand, there appeared to be less correspondence between the ranks compared to whispered conversational speech. It can be seen from the table, that only Speaker 4 tended to have the lowest ranked prediction-success rates *and* the lowest ranked intelligibility scores; in other speakers, the association between the two series of ranks appeared inconsistent. However, due to the lack of variance in the prediction success rates, this relationship was not further investigated.

To summarize, all parameters considered in the present analysis — C/V duration-ratio, $F1$-offset dynamics and C/V amplitude-ratio in phonated speech, and C/V duration-ratio and $F1$-offset dynamics in whispered speech — significantly contributed to classification of consonants into the "voiced" and the "voiceless" categories. However, in both phonation modes, C/V duration-ratio emerged as the most important contributor to the consonant "voicing" prediction. Results of the cross-validation test with clear whispered speech data indicated high predictive capability of the logistic regression model in whispered speech.

4.2 Correlations between accuracy of consonant "voicing" identification and magnitude of the acoustic contrasts

In the previous section, C/V duration-ratio was identified as the major predictor of consonant "voicing" in whispered speech. The goal of the present analysis was to examine the relationship between the magnitude of the CVDR-contrast and A'-transformed intelligibility scores from the perceptual test. Figures 23 and 24 show correlations between CVDR-contrasts and A'-scores in whispered conversational and whispered clear speech, correspondingly. Each datapoint represents a summary score for a consonant-pair in a particular vowel context for a particular speaker. As it can be seen from the figures, in both speaking-styles, there was a moderately strong correlation between the magnitude of the CVDR-contrast and consonant "voicing" intelligibility in whispered speech ($r = 0.41$ and $r = 0.44$ for conversational and clear speech, correspondingly).

Chapter 5

Discussion

Whispered speech is a naturally distorted speech signal that nevertheless shares certain acoustic similarities with clear speech — a mode of speaking associated with increased speech intelligibility. Thus, intelligibility of whispered speech may be rooted in increased duration of the phonetic segments and favorable consonant/vowel amplitude-ratio. Investigating acoustic properties responsible for intelligibility of whispered speech may help our general understanding of speech perception under suboptimal listening conditions. It may also contribute to our knowledge of acoustically impoverished speech such as speech signals delivered through cochlear implants and hearing aids. The goal of this study was to explore the availability of acoustic cues to final consonant "voicing" in whispered speech. Sentence-medial productions of stops, fricatives and affricates by four speakers who differed naturally in speaking rate were investigated in conversational phonated vs. conversational whispered speech and conversational whispered vs. clear whispered speech. Of particular interest were three duration-based acoustic parameters — vowel duration, consonant duration and consonant-to-vowel duration ratio. The study was divided into two parts. In the first part, detailed acoustic analyses were carried out to document the differences in spectral, intensity and duration-based parameters between phonated and whispered speech. Additional acoustical analysis was conducted on clear whispered speech to compare these parameters in two whispered speaking-styles. In the second part, listeners were

tested on their perception of consonants; Listeners' overall accuracy of consonant identification and their perception of consonant "voicing" in three speech conditions were assessed. Finally, to investigate the production-perception relationship, a logistic regression analysis and correlation analyses between the magnitude of the C/V duration-ratio and consonant "voicing" intelligibility were conducted separately for whispered conversational and whispered clear speech.

5.0.1 Production of consonant "voicing" contrasts in phonated and whispered consonants

Results of the present study indicated that, unlike phonated speech in which spectral, intensity and duration-based acoustic cues to "voicing" were in abundance, in whispered speech, only duration-based acoustic cues were consistently produced. The overall magnitude of these duration-based cues was comparable to that in the phonated speech. Although an increase in absolute durations of the phonetic segments in whispered speech observed in this study corroborated the findings previously reported in the literature (see, for example, Parnell et al., 1977, Schwartz, 1972, Jovicic & Saric, 2006)), only in one speaker (Speaker 2) these changes led to the *enhancement* of the "voicing" contrast in whispered vs. phonated conversational speech on all three duration-based parameters.

In clear whispered speech, consistent with the effect of *phonated* clear speech (Picheny et al., 1986, Smiljanic & Bradlow, 2005), talkers spoke at a slower speaking rate and increased duration of phonetic segments. Vowel duration "voicing" contrasts but not consonant duration "voicing" contrasts were enhanced in whispered clear relative to whispered conversational speech. The enhancement of the vowel duration contrast is attributed to a larger increase of vowel durations in "voiced" (19%) than in "voiceless" (10%) consonant contexts. These results differ from those reported by Smiljanic & Bradlow (2008c) for phonated speech. In their study of 5 speakers' productions of sentence-medial consonant "voicing" contrasts, vowels before "voiced" and "voiceless" codas showed more similar (41% and 46%, correspondingly) proportional lengthening for

clear speech relative to conversational speech. Subsequently, vowel duration "voicing" contrasts in clear phonated speech was of the same magnitude as in conversational phonated speech. Generally smaller proportional lengthening observed in the present study for whispered speech may be explained by the already increased duration of the phonetic segments in whispered conversational speech relative to phonated conversational so that whispered clear speech only allowed for limited additional lengthening. At the same time, the absence of spectral and intensity cues associated with clear phonated speech, e.g. expansion of pitch range and increased consonant-to-vowel amplitude ratio (Picheny et al., 1985, 1986, Smiljanic & Bradlow, 2005), is compensated by the enhancement of the durational contrasts.

Effect of whispered clear speech on the "voiced" and the "voiceless" consonants was equivalent. Subsequently, no enhancement of the consonant "voicing" contrast was observed. In contrast, Maniwa et al. (2009) reported enhanced durational differences between "voiced" and the "voiceless" fricatives in word-medial position in phonated clear speech due to the disproportional increase in frication noise duration in the "voiceless" consonants. Similar results were reported by Smiljanic & Bradlow (2008c) for word-medial stops: closure duration in the "voiceless" stops was lengthened more than closure duration of voiced stops. On the other hand, the reverse pattern — larger increase in the closure duration in the "voiced" stops — was observed for stops in word-initial position. Thus, it appears that potential enhancement of the consonant constriction duration "voicing" cue may depend on the position of the consonant within a word. While the consonants in the Maniwa and Smiljanic & Bradlow studies were in word-medial and/or word-initial consonants, in the present study we investigated the effect of clear speech on word-final consonants. In addition, variability among speakers could be another factor underlying the difference in the results between the studies. Similar to the findings reported in Maniwa et al. (2009), we observed that the effect of clear speech on consonant constriction duration varied among the speakers. Despite an overall effect of clear whispered speech on consonant constriction duration, the actual increase was seen only in two out of four speakers.

The last duration-based acoustic cue to consonant "voicing" that we explored was consonant-to-vowel duration ratio. Although the overall CVDR contrast remained stable in the two speaking styles, changes in CVDR between whispered conversational and whispered clear speech were subject to individual speaker variations. Logistic regression analyses suggested that CVDR was generally a better predictor of consonant "voicing" than consonant constriction duration and/or vowel duration alone. These results generally support the theory put forward by Port and Dalby (1982) that consonant-to-vowel duration ratio may serve as the most perceptually relevant durational acoustic cue to "voicing" of consonants in syllable-final position.

5.0.2 Perception of consonants in phonated and whispered speech

Consonants in phonated speech were identified with near-perfect accuracy although there were a few errors on b/p consonant-pairs. In whispered speech, the accuracy dropped to 85%–87%. Despite a choice of 8 possible consonant responses, perceptual errors were primarily limited to "voicing" errors in all conditions. The absence of place and manner of articulation errors probably reflects the fact that the C/V amplitude-ratio in whispered speech is naturally enhanced. In phonated speech, increased C/V amplitude-ratio has been shown to be associated with improved consonant identification (e.g. Hazan & Simpson, 1998) and an overall increase in speech intelligibility (Picheny et al., 1986, 1985). However, as described in the previous section, the C/V amplitude-ratio contrast as an acoustic cue to consonant "voicing" is not available in whispered speech. Thus, although enhanced C/V amplitude-ratio might have provided a perceptual benefit for identification of place and manner of articulation of consonants, it could not contribute to perception of consonant "voicing". On the other hand, the results reported here suggest better perception of whispered consonants than that observed by Tartter (1989). Very briefly, she investigated perception of 18 initial consonants [b, d, g, p, t, k, m, n, r, l, w, y, v, f, z, s, ʃ, ʒ] in isolated whispered /Cɑ/ syllables produced by two talkers. In her study, the overall accuracy of consonant identification was 64%, accuracy of consonant "voicing" identification was 72%, ac-

curacy of identification of place and manner of articulation was 91% and 86%, correspondingly. It is possible that the difference in the performance can be explained by a larger set of consonants in the previous study and presence of neighboring articulation categories in the response options in the Tartter study. For instance, in her study, listeners confused labial consonants with alveolar consonants more often than with velar consonants. In our study, alveolar pair d/t was not included in the stimulus or response set. At the same time, it is possible that the stimuli used in the present study were, in fact, perceptually more difficult. First, unlike isolated syllables in the Tartter study, we used nonsense words embedded in a sentence. Second, the stimuli were produced by 4 speakers who naturally differed in speaking rate, availability and magnitudes of the "voicing" contrasts. Finally, we introduced additional variability in the stimuli by using four vowels intrinsically different in duration, /ɛ, æ; ɑ:, ʌ/ and used multiple tokens to represent each vowel-consonant combination. Thus, it appears that the difference in performance cannot be simply attributed to a more extensive list of consonants in the Tartter study. It is possible, however, that perception of whispered consonants is affected by the availability of duration-based cues to consonant identity in different positions within a syllable. For instance, in phonated speech, acoustic cues to consonant "voicing" in syllable-initial position include voice onset time (VOT), first formant transition, burst intensity (for stops). In addition, there is frication and aspiration duration for fricatives and stops, correspondingly (Kuhl & Miller, 1975, Stevens & Klatt, 1974, Klatt, 1975). Unlike the consonants in syllable-final position, which benefit from the contribution of vowel duration to their "voicing" identity, time-based cues to "voicing" in syllable-initial consonants are generally limited by the duration of the consonant constriction itself. In this light, a higher proportion of errors in the Tartter study might be associated with a more detrimental effect of whispered speech on syllable-initial consonants than syllable-final consonants at least with regard to their "voicing" identification. However, a more systematic comparison will be required to test this hypothesis.

5.0.3 Clear speech benefit in whispered speech

Overall, the improvement in perception of final consonant "voicing" in clear vs. conversational whispered speech (clear speech benefit) was small, 2% on average. At the same time, the effect of clear whispered speech varied greatly across speakers. Clear speech benefit afforded by two speakers, Speaker 1 and 2, was in agreement with the predictions made on the basis of the production data. Thus, although whispered clear speech provided only limited benefit for consonant "voicing" intelligibility overall, productions of two out of four speakers were, in fact, substantially more intelligible in whispered clear speech. This outcome was not entirely surprising as a number of studies have shown that clear phonated speech does not always provide speech intelligibility benefit and sometimes has a detrimental effect on perception. For instance, results from a recent study by Maniwa et al. (2008) suggest that fricatives were sometimes harder to perceive in clear speech: intelligibility gain scores between clear and conversational speech ranged from -4% to $+11\%$ across different speakers. In agreement with this pattern, perceptual gain for whispered fricatives provided by clear speech in the present study varied from -7% to 6%. On the other hand, perception of stop consonants and, particularly, bilabials showed a more general clear speech benefit in whispered speech. This agrees with the findings of Hazan & Simpson (2000) who showed that artificial cue enhancement, thought to be similar to the acoustic changes in natural clear speech, improved intelligibility of stops the most.

5.0.4 Relationship between production and perception of consonant "voicing" in whispered speech

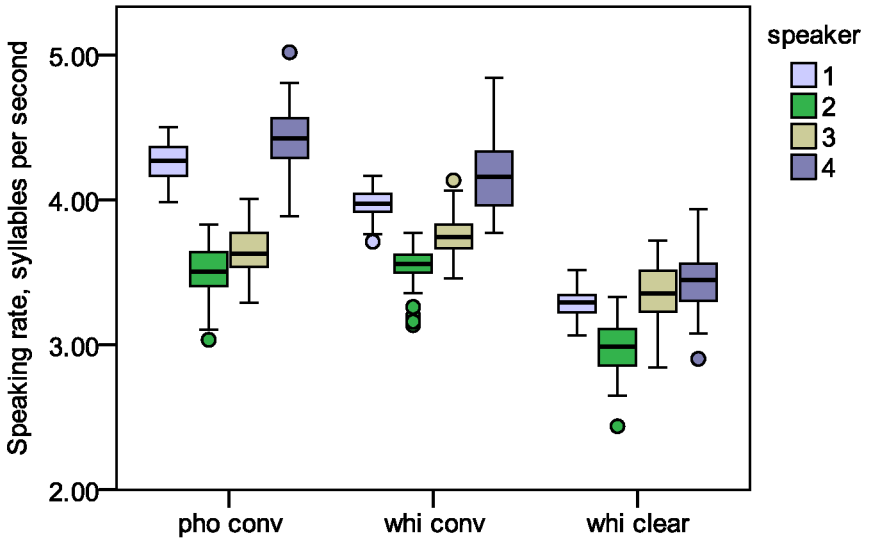
In whispered, as well as in phonated speech mode, C/V duration-ratio emerged as the most important contributor for predicting consonant "voicing". Models for both phonated and whispered speech with a C/V duration-ratio as the only independent factor had high success classifying consonants into the "voiced" and the "voiceless" categories. Furthermore, our analysis showed that

there was a moderately strong correlation between the magnitude of the C/V duration-ratio contrast and consonant "voicing" intelligibility in whispered speech. These results indicated that listeners could, to a certain degree, make use of the C/V duration-ratio cue to consonant "voicing". This agrees with the results Port and Dalby (1982) suggesting that the C/V duration-ratio serves as an important "voicing" cue for consonants in syllable-final position. However, a discrepancy between the prediction success rate of the logistic regression (94%) and the observed accuracy of consonant identification in the perception test (85%–87%) suggests that not all listeners use the C/V duration-ratio in the most effective way. More specifically, the logistic regression models used C/V duration-ratio values standardized by speaker, consonant-pair and vowel. It is possible that such fine tuned normalization is not fully available for listeners to whispered speech. For instance, information about formant-ratios incorporating both $f0$ and $F1$ information — both reduced and/or neutralized in whispered speech — has been previously shown to be important for vowel and speaker normalization (Miller, 1989, Syrdal & Gopal, 1986). However, despite the differences in the outcomes of the statistical classification of whispered consonants and their perception by the human listeners, the results of the logistic regression modeling have an important implication: The "voicedness" of a whispered consonant may be accurately predicted by C/V duration-ratio. This information can be potentially used in the speech processors of cochlear implants to improve consonant identification and overall speech intelligibility.

As one of the human speaking modes, whispered speech is a signal naturally devoid of acoustic redundancy associated with phonated speech. As such it serves as an interesting model for studying time-based acoustic information that remains available. In this study, we set out to investigate production and perception of acoustic cues to final consonant "voicing" in the absence of periodic energy that distinguishes "voiced" and "voiceless" consonants in normal speech. Exploring those acoustic cues to consonant "voicing" that become the primary source of differentiation of the "voiced" vs. "voiceless" consonants in whispered speech may help our general understanding of speech perception under non-optimal listening conditions by the normal hear-

ing. In addition, it may contribute to our knowledge of acoustically impoverished speech such as speech delivered through cochlear implants and hearing aids to hearing impaired listeners. Specifically, this information may be useful in developing algorithms for speech processors in cochlear implants and improving audio processing in hearing aids. Finally, knowledge of acoustic parameters of whispered speech can contribute to improvement of speech recognition algorithms in situations where phonated speech is not appropriate, and for designing speech synthesis systems for voice-impaired patients.

List of Figures



speaker	pho conv			whi conv			whi clear		
	Mean	Range		Mean	Range		Mean	Range	
1	4.26	3.98	4.50	3.97	3.71	4.16	3.29	3.07	3.52
2	3.51	3.03	3.83	3.54	3.13	3.77	2.97	2.44	3.33
3	3.64	3.29	4.01	3.75	3.46	4.13	3.37	2.84	3.72
4	4.42	3.89	5.02	4.15	3.77	4.84	3.43	2.90	3.93
Average	3.97	3.03	5.02	3.85	3.13	4.84	3.26	2.44	3.93

Figure 1: Speaking rate in different speakers in phonated conversational, whispered conversational and whispered clear speech. Circles represent outliers. Two speakers, Speaker 1 and Speaker 4, spoke at a slower rate in conversational whispered speech relative to conversational phonated speech. In clear whispered speech, all speakers spoke at a slower rate compared to both conversational phonated and conversational whispered speech. Relative differences between speakers became less prominent in whispered speech and, particularly, in clear whispered speech.

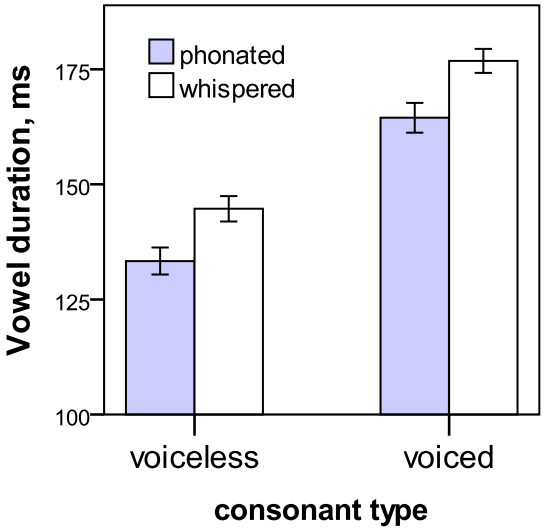


Figure 2: Average vowel durations in “voiced” and “voiceless” consonant contexts in phonated and whispered speech. Bars are mean durations summed across all vowels and speakers in each phonation mode for “voiced” and “voiceless” contexts. Error bars are standard errors of the mean. In whispered speech (light bars), vowel durations were significantly longer than in phonated speech (dark bars) in both “voicing” contexts.

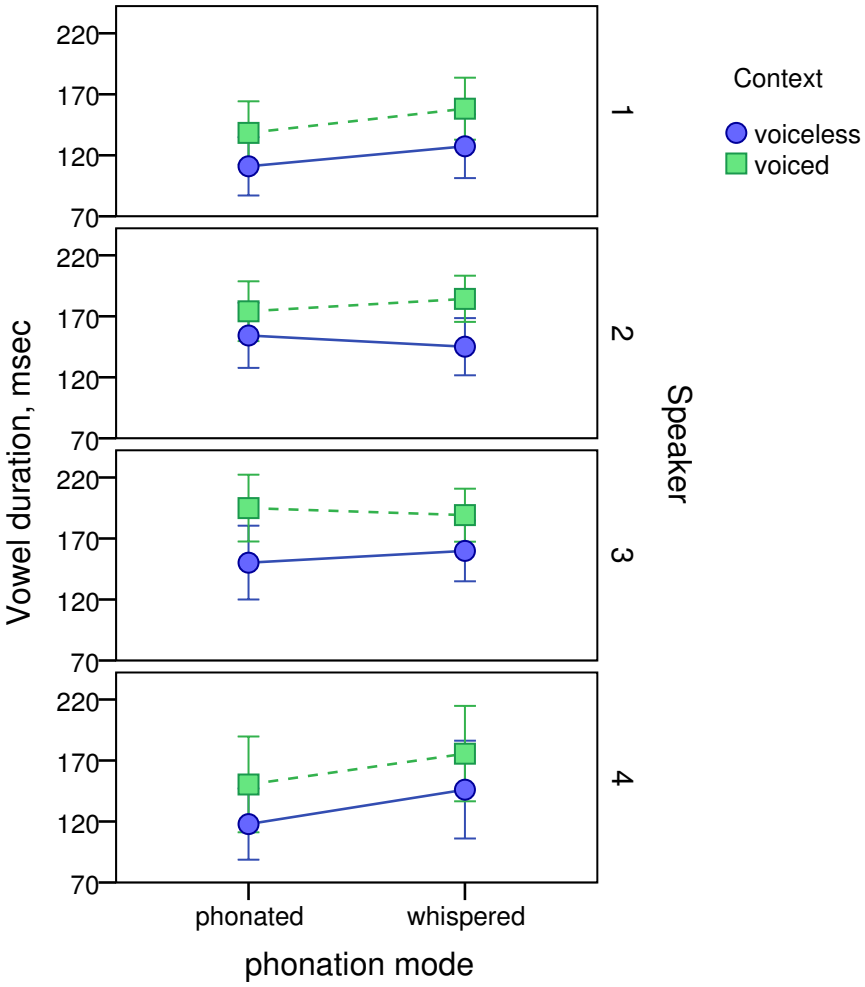


Figure 3: Average vowel durations in "voiced" and "voiceless" contexts in individual speakers in phonated and whispered speech. Symbols represent mean vowel duration values averaged across all four vowels in the "voiced" (squares) and the "voiceless" (circles) consonant contexts in each speaking mode for each of the four speakers. Error bars are standard deviations.

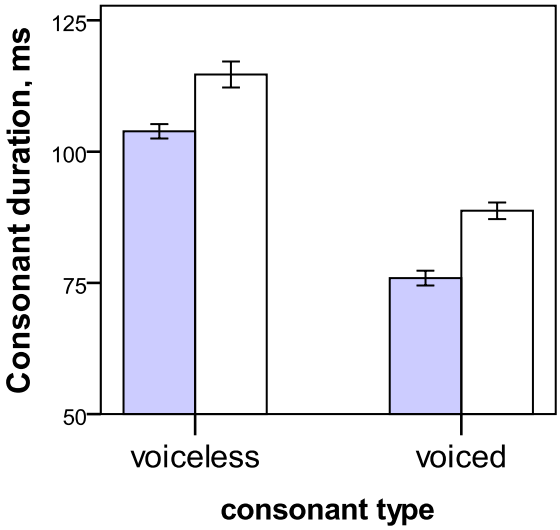


Figure 4: Average consonant durations in "voiced" and "voiceless" consonant contexts in phonated and whispered speech. Bars are mean durations summed across all vowels and speakers in each phonation mode for "voiced" and "voiceless" consonants. Error bars are standard errors of the mean. Both "voiced" and "voiceless" consonant were significantly longer in whispered speech (light bars) compared to phonated speech (dark bars) in both "voicing" contexts.

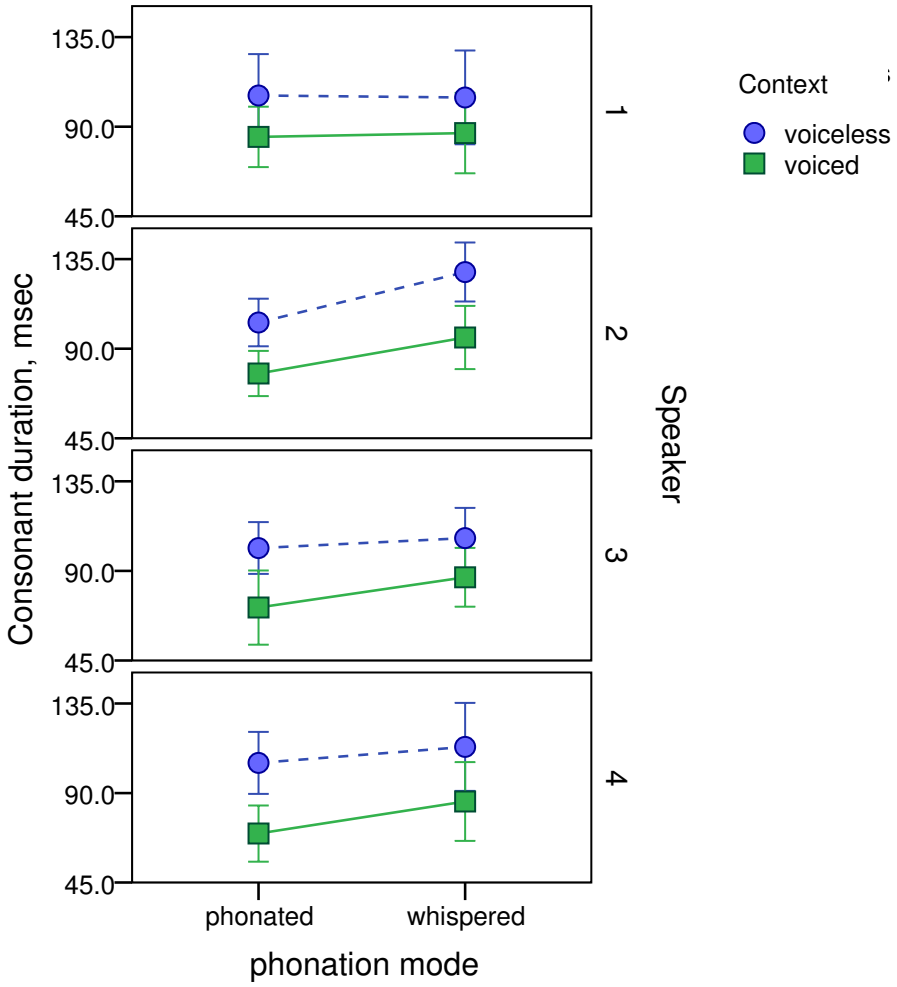


Figure 5: Average consonant durations in "voiced" and "voiceless" contexts in individual speakers in phonated and whispered speech. Symbols represent mean consonant duration values averaged across all four vowels in "voiced" (squares) and "voiceless" (circles) consonant contexts in each speaking mode for each of the four speakers. Error bars are standard deviations.

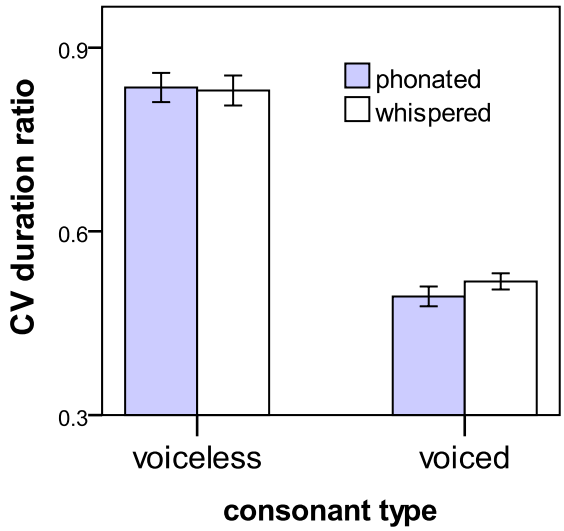


Figure 6: Consonant-to-vowel duration-ratio in phonated and whispered speech. Bars are mean values summed across all vowels and speakers in each phonation mode for “voiced” and “voiceless” consonants. Error bars are standard errors of the mean. Consonant-to-vowel duration-ratios neither in the “voiced” nor in the “voiceless” consonant context did not significantly change between phonated speech (dark bar) and whispered speech (light bar).

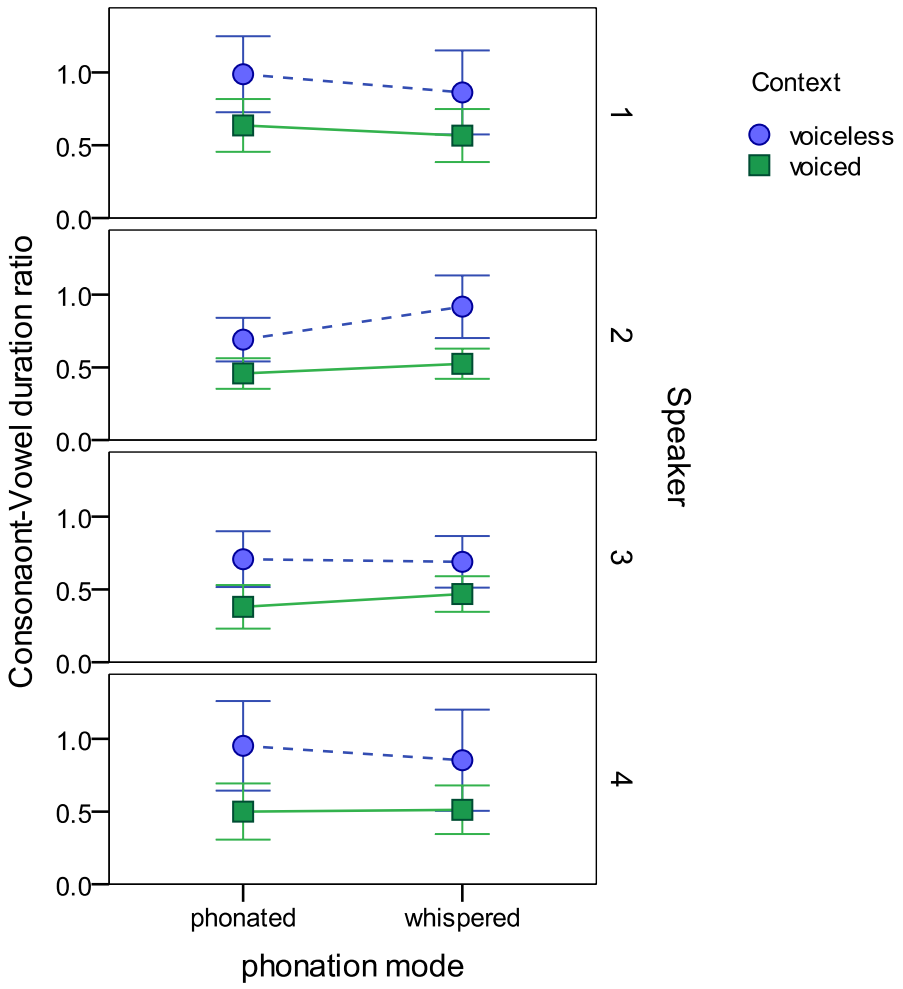


Figure 7: Average C/V duration-ratios in "voiced" and "voiceless" contexts in in phonated and whispered speech in four speakers. Symbols represent mean CVDR values averaged across all four vowels in "voiced" (circles) and "voiceless" (squares) consonant contexts in each speaking mode for each of the four speakers. Error bars are standard deviations. The effect of whispered speech was inconsistent across speakers and "voicing" contexts

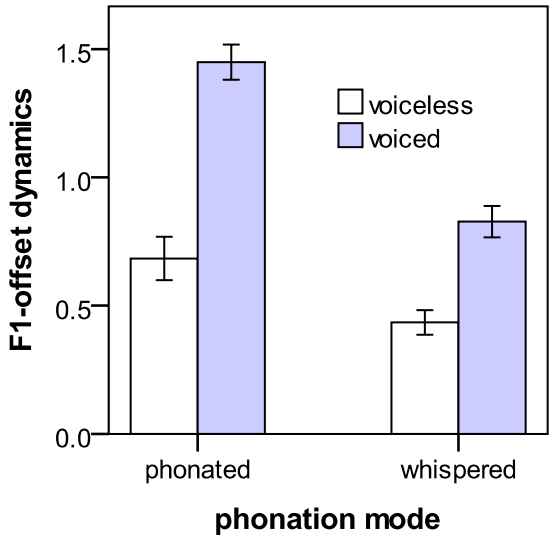


Figure 8: *F1*-offset dynamics in conversational phonated and whispered speech. Bars represent mean values of *F1*-offset dynamics summed over all vowels and speakers for "voiced" and "voiceless" consonant contexts. Error bars are standard errors of the mean.

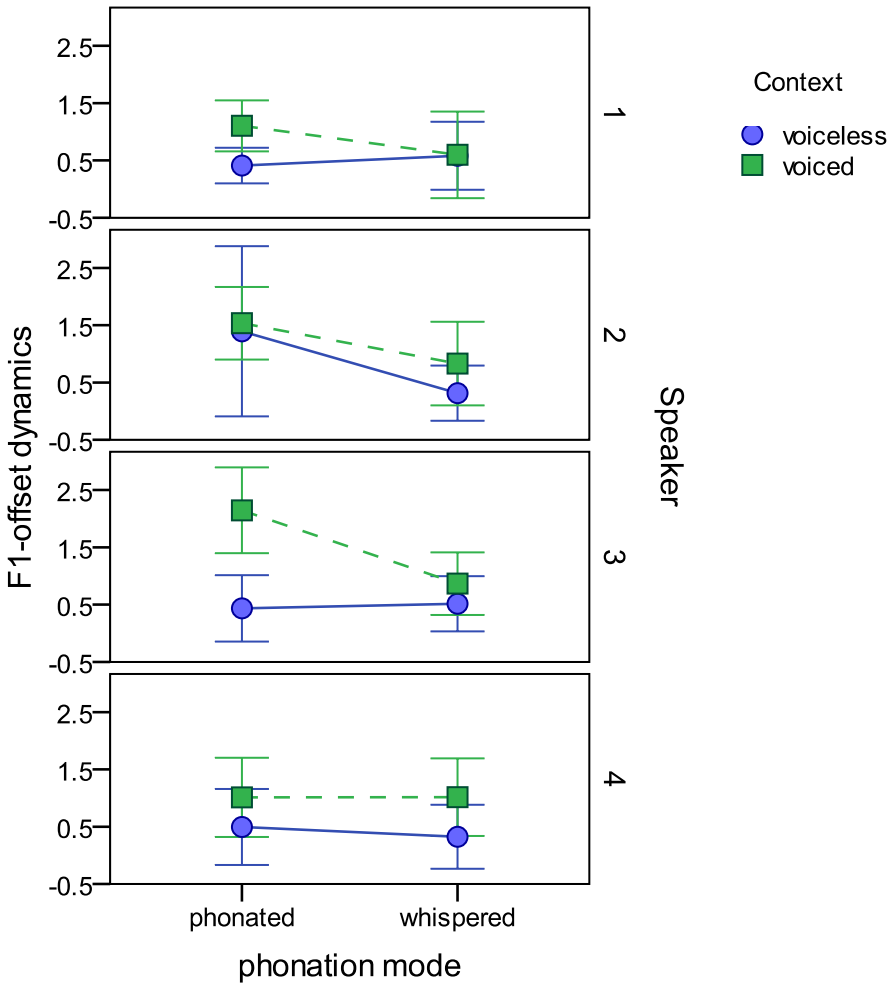


Figure 9: *F1*-offset dynamics in "voiced" and "voiceless" consonants in phonated and whispered conversational speech in four speakers. Symbols represent mean values of *F1*-offset dynamics summed over four vowels for "voiced" (squares) and "voiceless" (circles) consonants in each phonation-mode for each speaker. Error bars are standard errors of the mean. *F1*-offset dynamics is expressed as Bark difference between first formant values at vowel offset and 75% time-point of the vocalic nucleus.

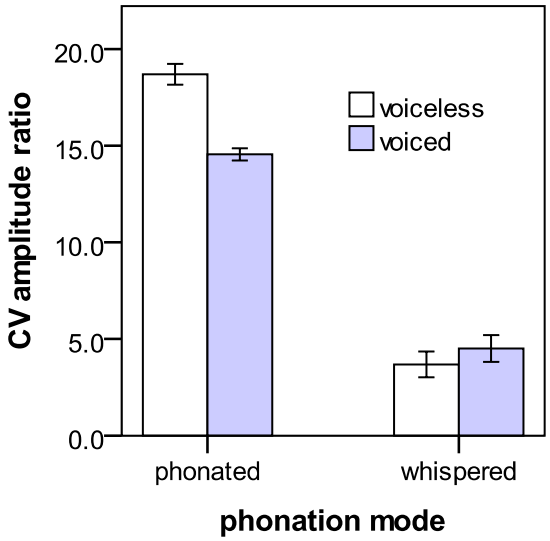


Figure 10: Average magnitude of consonant-to-vowel amplitude-ratio in phonated and whispered conversational speech. Bars represent values of C/V amplitude-ratios summed over four vowels and four speakers in “voiced” and “voiceless” consonant context in each phonation-mode. Error bars are standard errors of the mean

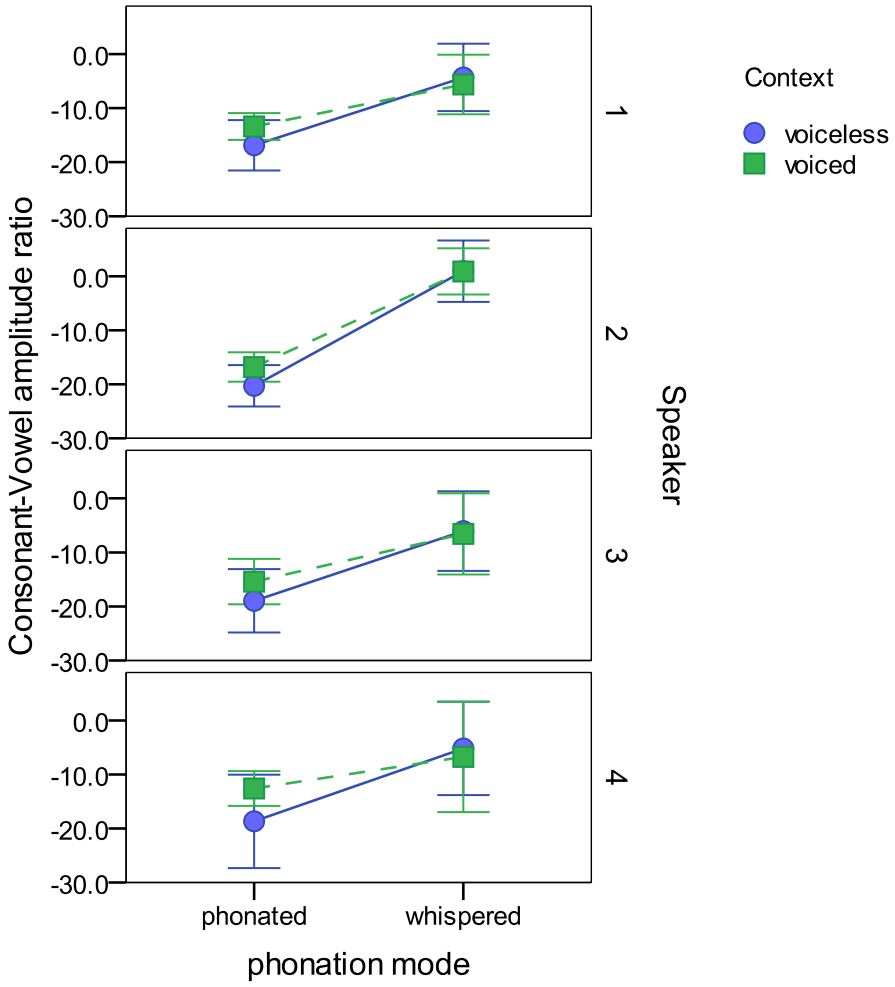


Figure 11: Consonant-to-vowel amplitude-ratio in "voiced" (squares) and "voiceless" (circles) consonants in phonated and whispered conversational speech. Symbols represent *absolute* values of consonant-to-vowel amplitude-ratio summed over four vowels for "voiced" and "voiceless" consonants in each phonation-mode for each speaker. Error bars are standard errors of the mean.

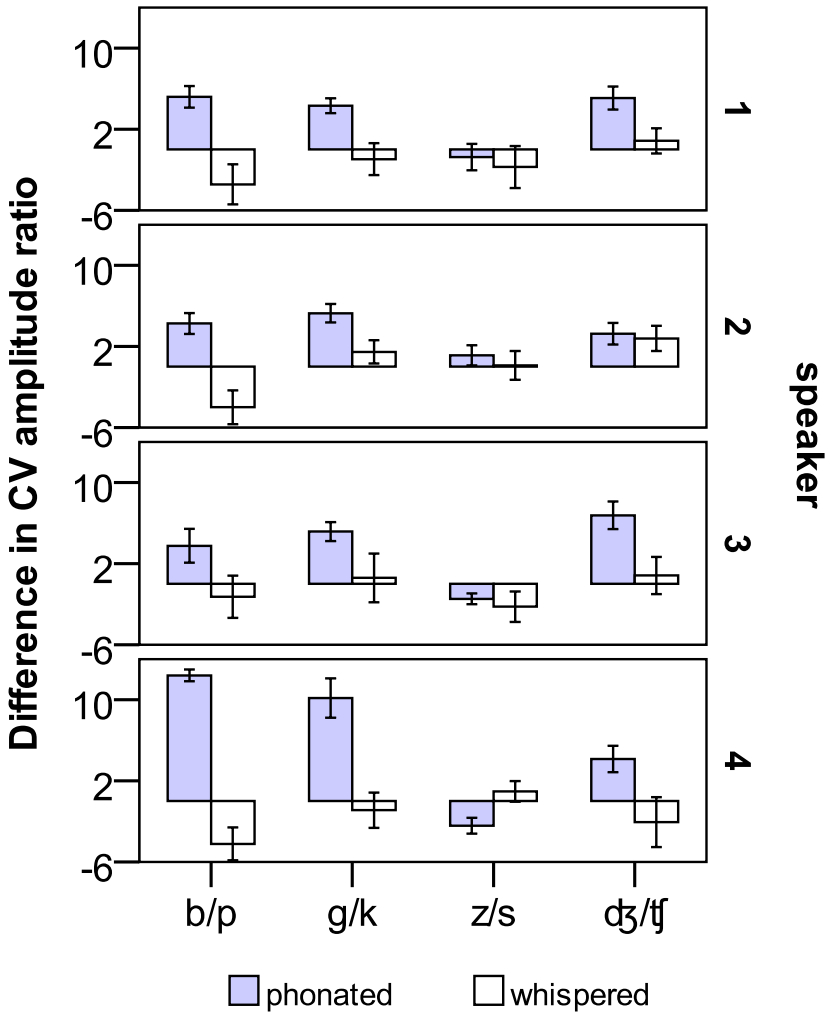


Figure 12: The contrast of C/V amplitude-ratio between "voiced" and "voiceless" consonants in different consonant-pairs in conversational phonated (dark bars) and whispered (light bars) speech. Bars represent *differences* in values of C/V amplitude-ratio between "voiced" and "voiceless" consonants summed over four vowels for each consonant-pair. Error bars are standard errors of the mean.

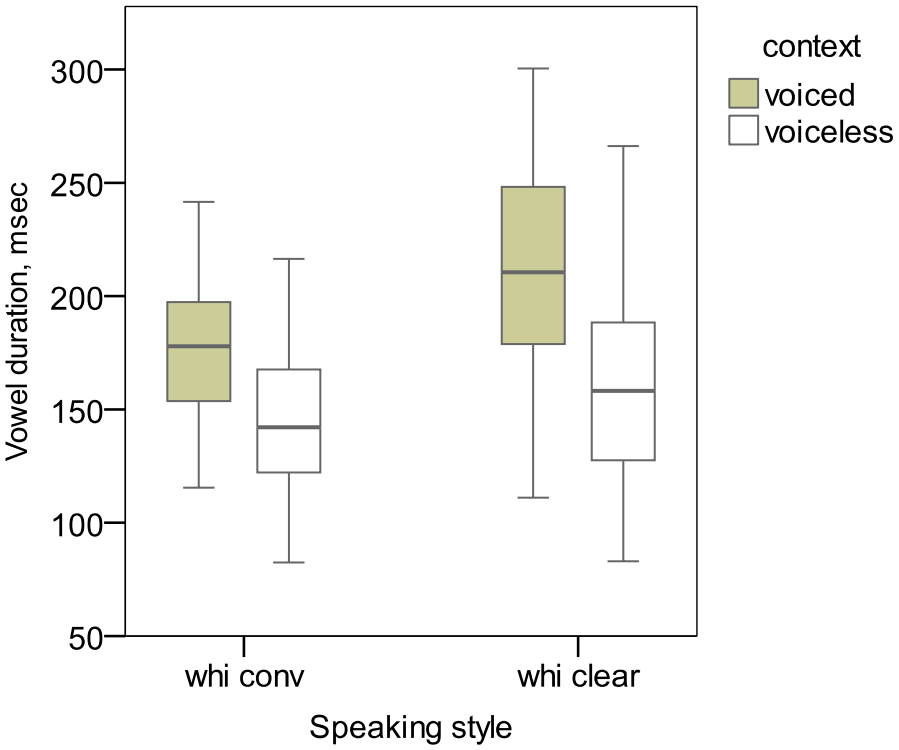


Figure 13: Vowel duration in whispered conversational and clear speech

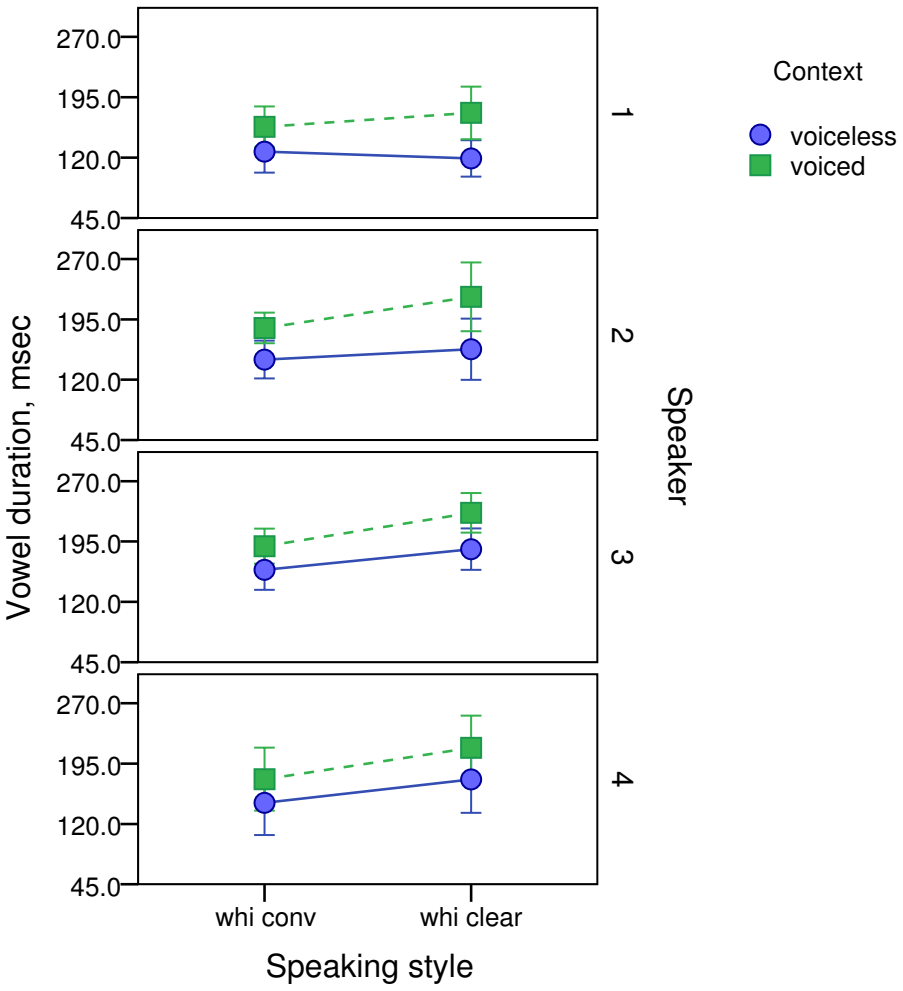


Figure 14: Vowel duration in whispered conversational and clear speech in individual speakers. Points represent vowel durations summed over four vowels in the "voiced"(squares) and the "voiceless" (circles) contexts in whispered conversational and whispered clear speech. Error bars are standard deviations. Vowel durations in clear whispered speech increased relative to conversational whispered speech; the effect was greater for the vowels in the "voiceless" context

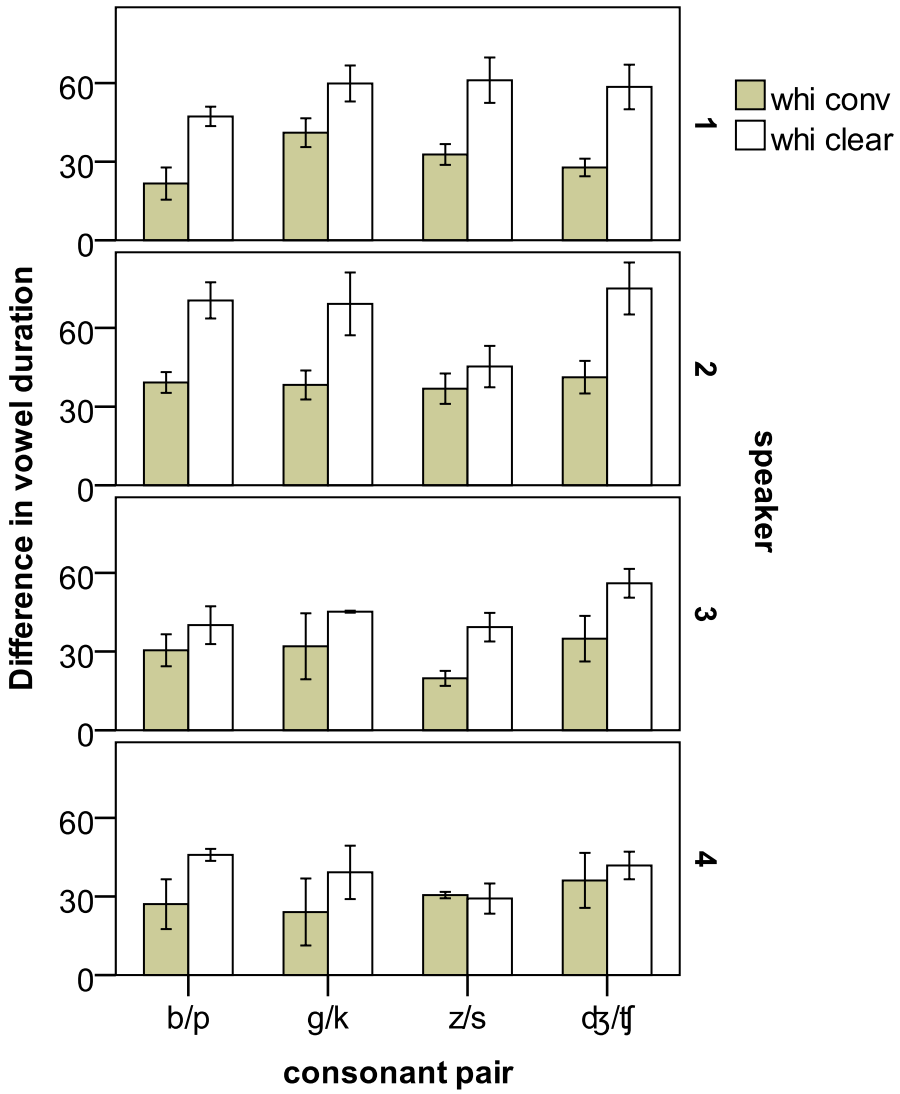


Figure 15: Difference in the vowel duration between “voiced” and “voiceless” consonants in two whispered conditions. Error bars are standard errors of the mean.

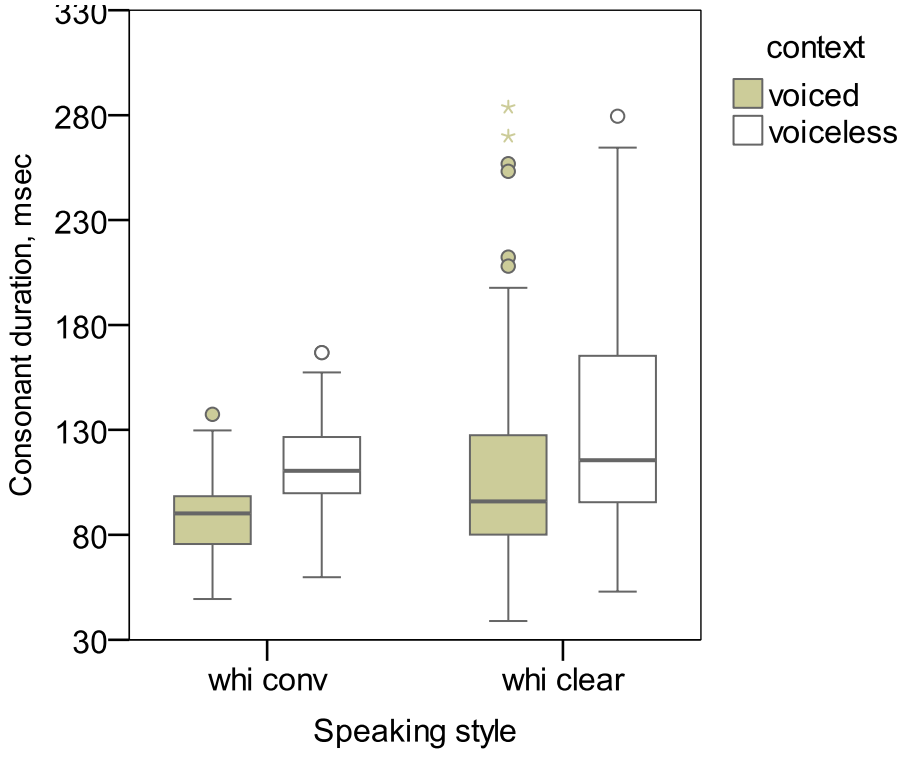


Figure 16: Duration of “voiced” and “voiceless” consonants in whispered conversational and clear speech

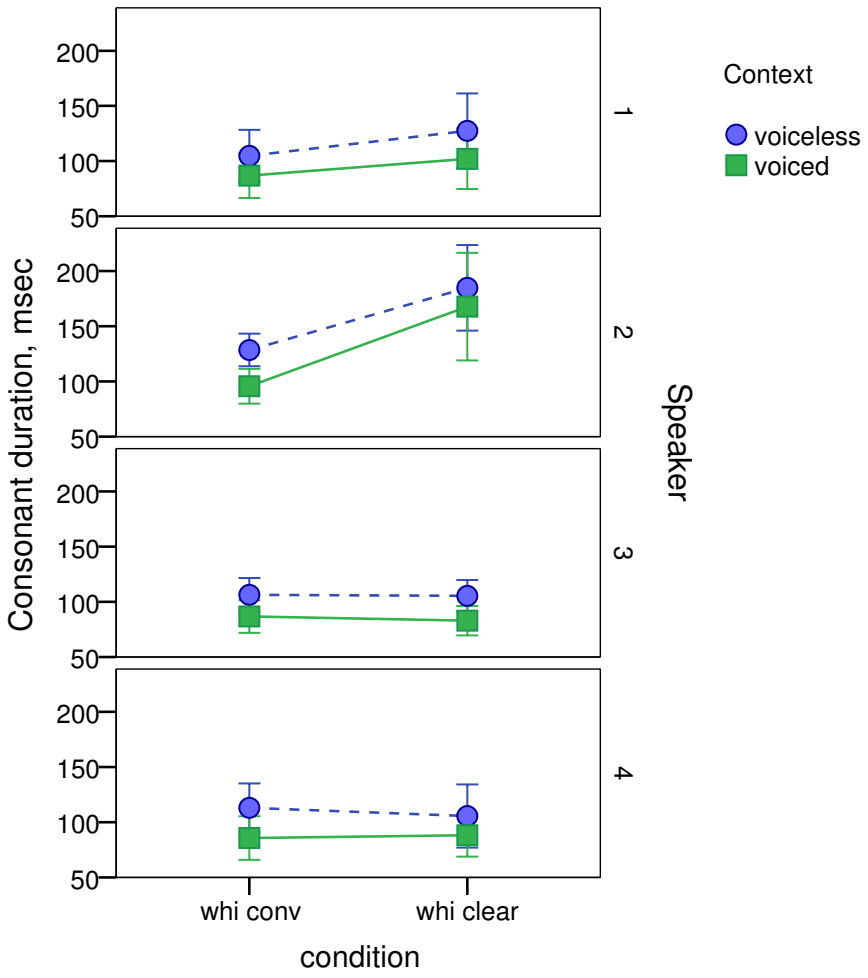


Figure 17: Consonant duration in whispered conversational and clear speech in individual speakers. Points represent consonant durations summed over four preceding vowels for the "voiced" (squares) and the "voiceless" (circles) consonants in whispered conversational and whispered clear speech. Error bars are standard deviations. Duration of both "voiced" and "voiceless" consonants tended to decrease in clear whispered relative to conversational whispered speech.

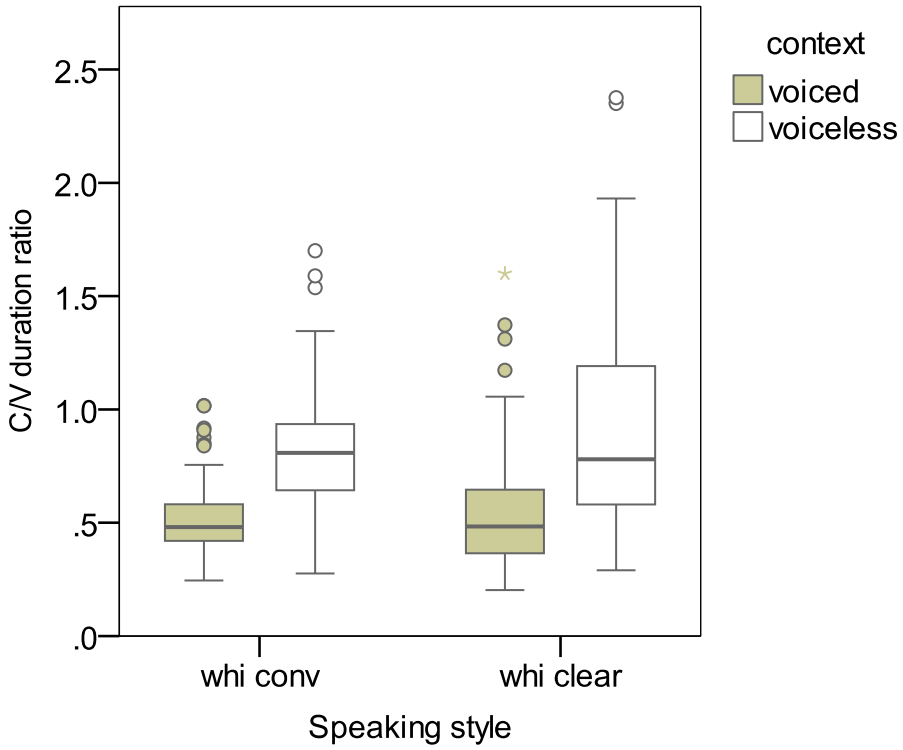


Figure 18: Consonant-to-vowel duration-ratio in whispered conversational and clear speech

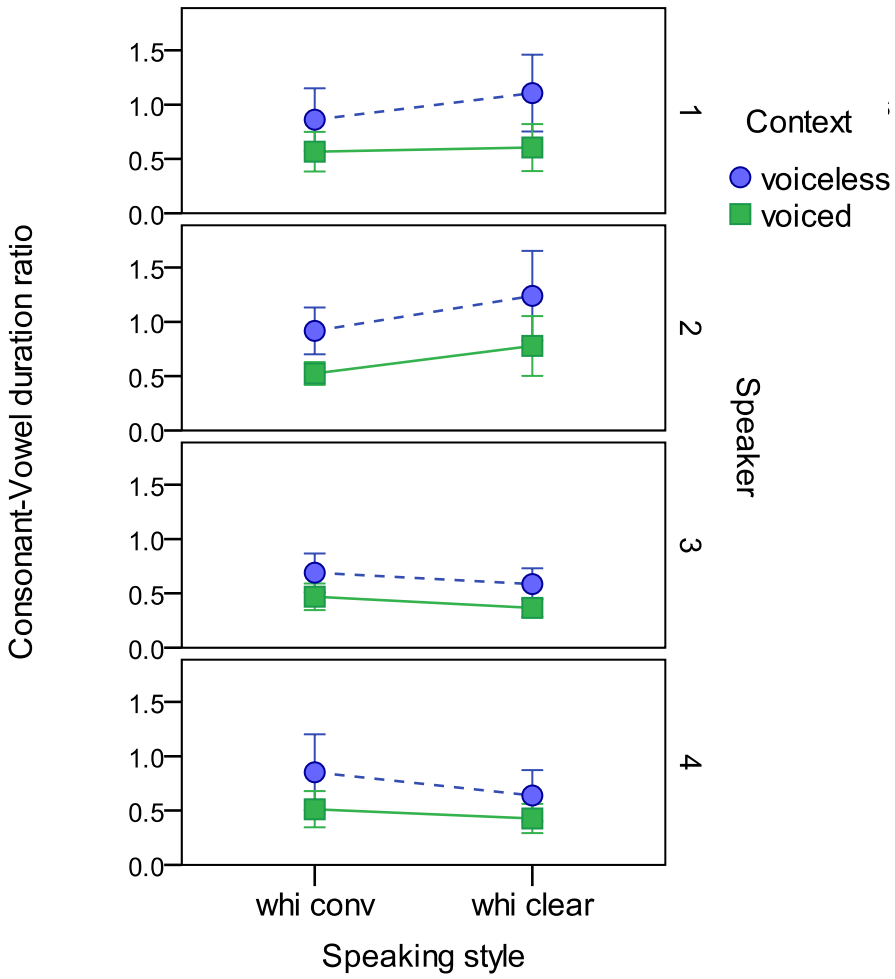


Figure 19: Consonant-to-vowel duration-ratio in whispered conversational and clear speech in individual speakers. Symbols represent CVDRs summed over vowels preceding the "voiced" consonants (squares) and the "voiceless" (circles) consonants in whispered conversational and whispered clear speech. Error bars are standard deviations

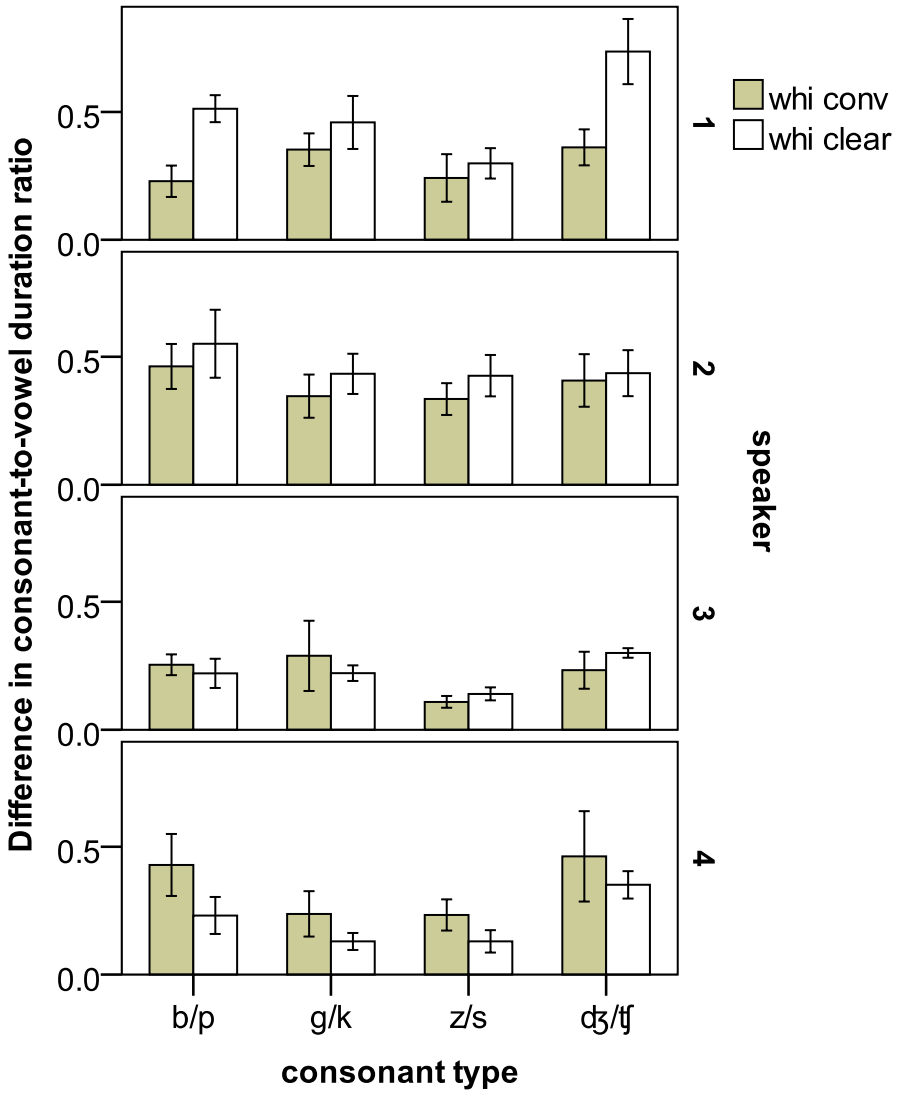
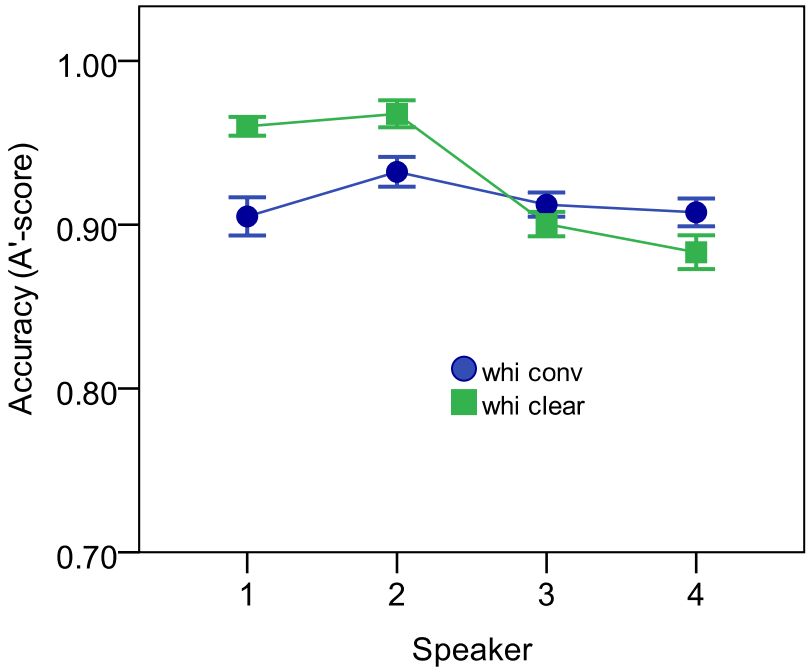
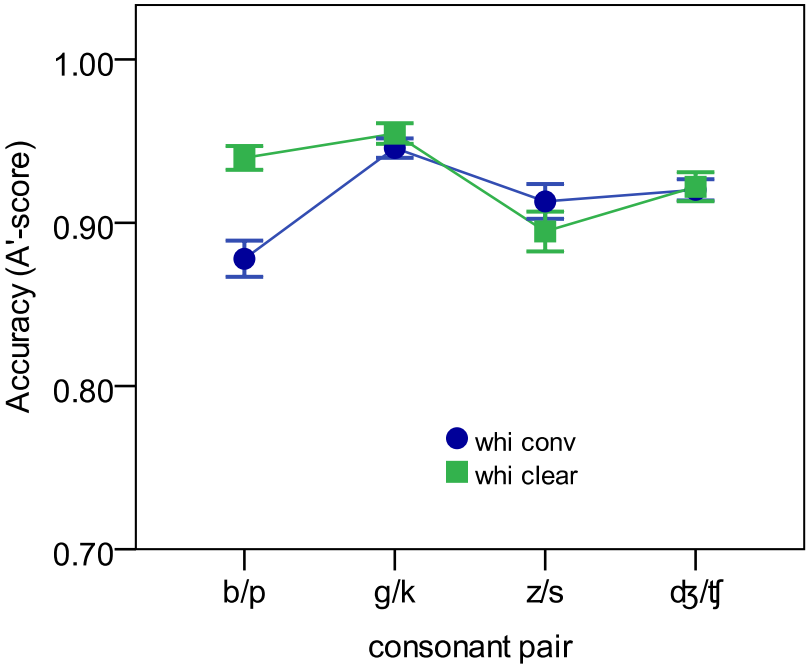


Figure 20: Difference in C/V duration-ratio between "voiced" and "voiceless" consonants in two whispered conditions. Error bars are standard errors of the mean. In whispered clear speech, CV duration-ratio contrast increased in Speaker 1 and 2, decreased in Speaker 4 and minimally alternated in Speaker 3.



(a)



(b)

Figure 21: Accuracy of consonant "voicing" perception in whispered conversational and clear speech in individual speakers (a) and consonant-pairs (b). Error bars are standard errors.

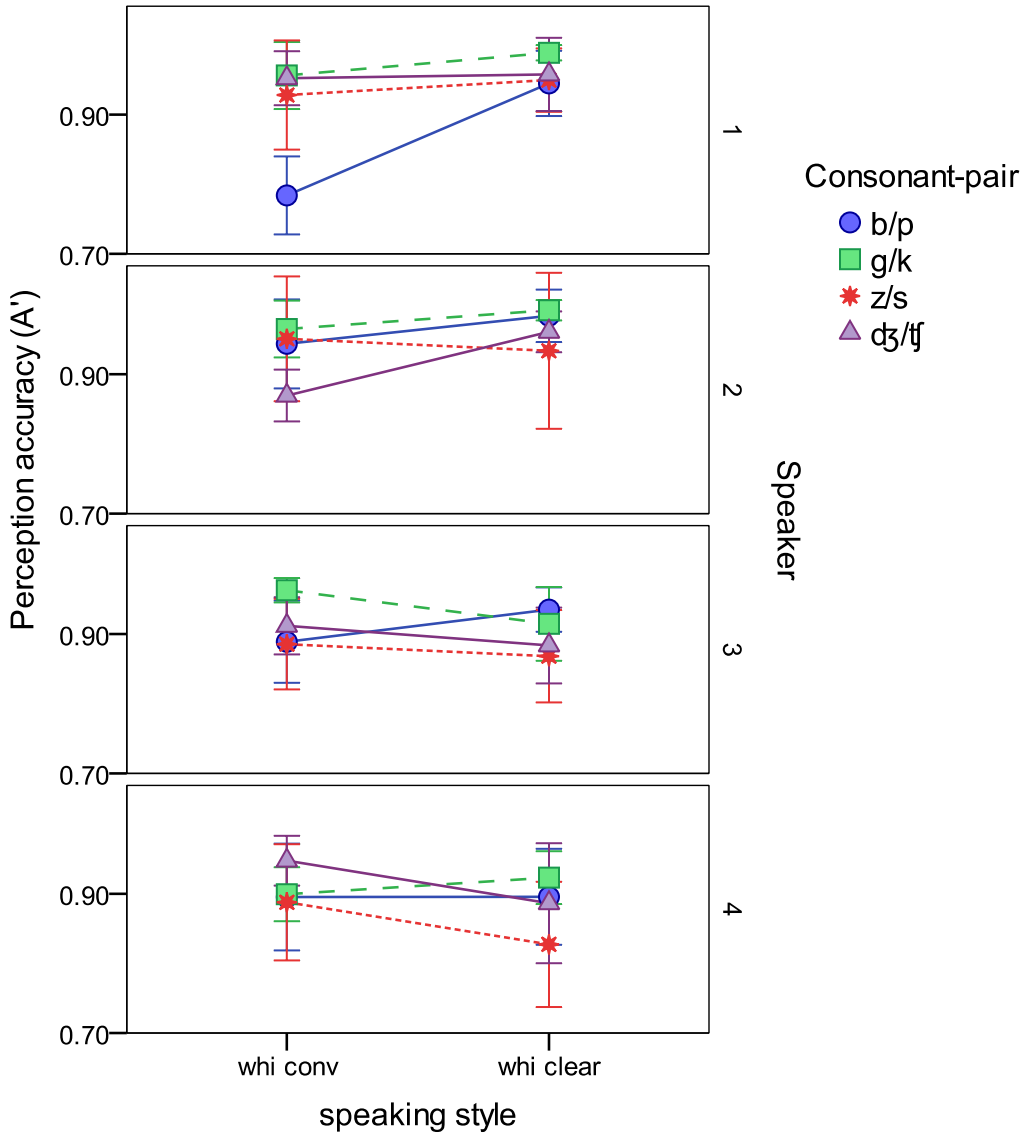


Figure 22: Accuracy of consonant "voicing" perception in whispered conversational and clear speech in different consonant-pairs. Error bars are standard deviations.

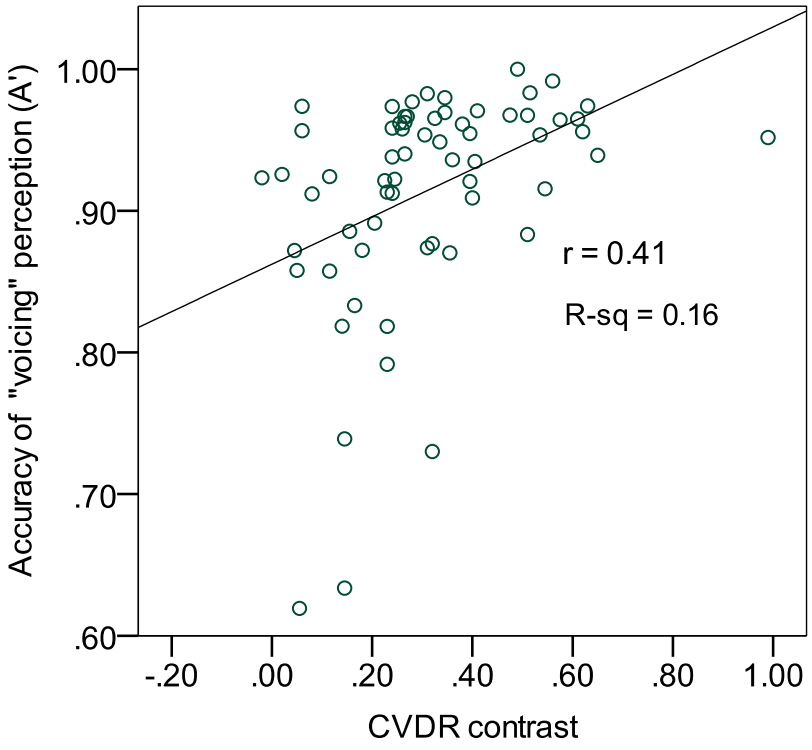


Figure 23: Correlation between consonant “voicing” intelligibility and the magnitude of the C/V duration-ratio in conversational whispered speech.

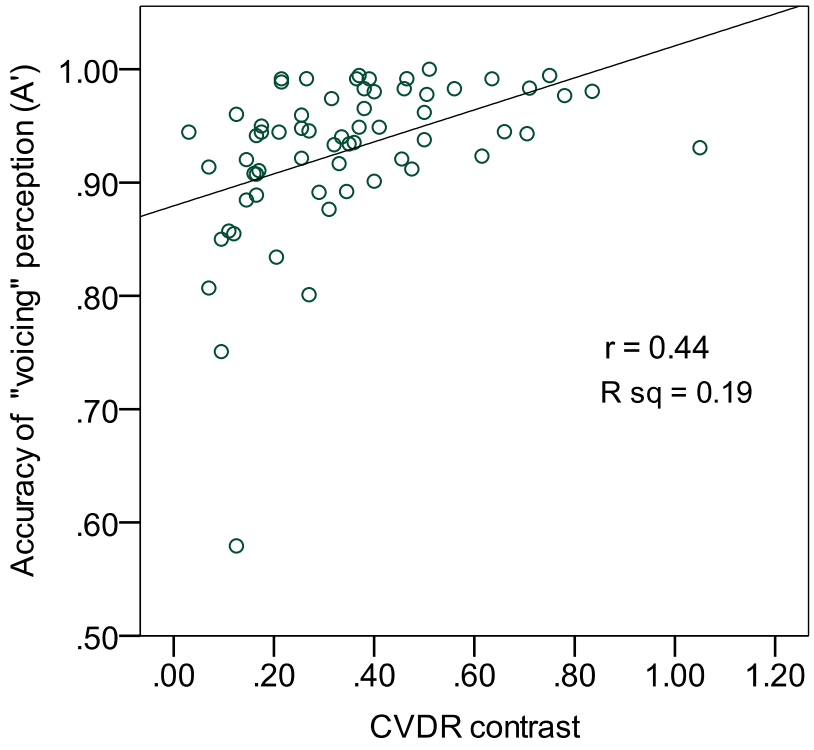


Figure 24: Correlation between consonant “voicing” intelligibility and the magnitude of the C/V duration-ratio in clear whispered speech.

List of Tables

speaker	phonation	vowel duration					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	phonated	138	26	111	24	125	28
	whispered	158	25	127	26	143	30
2	phonated	174	25	154	27	164	27
	whispered	184	19	145	23	165	29
3	phonated	195	27	150	30	173	36
	whispered	189	22	160	25	174	28
4	phonated	150	39	118	29	134	38
	whispered	176	39	146	40	161	42
Average	phonated	164	37	133	33	149	38
	whispered	177	30	145	31	161	34

Table 1: Average vowel duration in “voiced” and “voiceless” contexts in conversational phonated and whispered speech

speaker	phonation	consonant duration					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	phonated	85	15	106	21	95	21
	whispered	87	20	105	24	96	24
2	phonated	78	11	103	12	90	17
	whispered	96	16	129	15	112	22
3	phonated	72	19	101	13	87	22
	whispered	87	15	106	15	97	18
4	phonated	70	14	105	16	87	23
	whispered	86	20	113	22	99	25
Average	phonated	76	16	104	16	90	21
	whispered	89	18	113	21	101	23

Table 2: Average consonant duration in “voiced” and “voiceless” contexts in conversational phonated and whispered speech

speaker	condition	C/V duration ratio					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	phonated	0.64	0.18	0.99	0.26	0.81	0.28
	whispered	0.57	0.18	0.86	0.29	0.71	0.28
2	phonated	0.46	0.11	0.69	0.15	0.57	0.17
	whispered	0.52	0.10	0.92	0.21	0.72	0.26
3	phonated	0.38	0.15	0.71	0.19	0.54	0.24
	whispered	0.47	0.12	0.69	0.18	0.58	0.19
4	phonated	0.50	0.19	0.95	0.31	0.73	0.34
	whispered	0.51	0.17	0.85	0.35	0.68	0.32
Average	phonated	0.49	0.18	0.83	0.27	0.66	0.29
	whispered	0.52	0.15	0.83	0.28	0.67	0.27

Table 3: Average C/V duration-ratio in “voiced” and “voiceless” contexts in conversational phonated and whispered speech

speaker	phonation	*F1*-offset dynamics					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	phonated	1.10	0.44	0.41	0.31	0.76	0.52
	whispered	0.60	0.76	0.58	0.59	0.59	0.67
2	phonated	1.53	0.63	1.39	1.49	1.46	1.13
	whispered	0.83	0.73	0.31	0.48	0.58	0.67
3	phonated	2.15	0.75	0.44	0.58	1.29	1.09
	whispered	0.87	0.55	0.52	0.48	0.69	0.54
4	phonated	1.01	0.69	0.50	0.66	0.75	0.72
	whispered	1.02	0.68	0.32	0.56	0.67	0.71
Average	phonated	1.45	0.78	0.68	0.96	1.07	0.95
	whispered	0.83	0.69	0.43	0.54	0.63	0.65

Table 4: Average values of *F1*-offset dynamics in “voiced” and “voiceless” contexts in conversational phonated and whispered speech

speaker	phonation	C/V amplitude-ratio					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	phonated	-13.40	2.49	-16.86	4.66	-15.13	4.10
	whispered	-5.64	5.51	-4.32	6.24	-4.98	5.88
2	phonated	-16.80	2.73	-20.28	3.84	-18.54	3.74
	whispered	0.90	4.26	0.95	5.68	0.92	4.97
3	phonated	-15.42	4.20	-18.97	5.87	-17.19	5.37
	whispered	-6.57	7.50	-6.05	7.37	-6.31	7.38
4	phonated	-12.59	3.24	-18.66	8.64	-15.62	7.16
	whispered	-6.73	10.22	-5.17	8.66	-5.95	9.43
Average	phonated	-14.55	3.60	-18.69	6.09	-16.62	5.40
	whispered	-4.51	7.82	-3.68	7.51	-4.10	7.66

Table 5: Average C/V amplitude-ratio in “voiced” and “voiceless” contexts in conversational phonated and whispered speech

speaker	condition	vowel duration					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	whi conv	158	25	127	26	143	30
	whi clear	176	33	119	23	147	40
2	whi conv	184	19	145	23	165	29
	whi clear	223	43	158	38	190	52
3	whi conv	189	22	160	25	174	28
	whi clear	231	25	185	26	208	34
4	whi conv	176	39	146	40	161	42
	whi clear	215	40	175	42	195	45
Average	whi conv	177	30	145	31	161	34
	whi clear	211	41	159	41	185	49

Table 6: Average vowel duration in “voiced” and “voiceless” contexts in whispered conversational and clear speech

speaker	condition	consonant duration					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	whi conv	87	20	105	24	96	24
	whi clear	102	27	127	34	115	33
2	whi conv	96	16	129	15	112	22
	whi clear	168	49	185	39	176	45
3	whi conv	87	15	106	15	97	18
	whi clear	83	13	105	14	94	18
4	whi conv	86	20	113	22	99	25
	whi clear	88	19	106	29	97	26
Average	whi conv	89	18	113	21	101	23
	whi clear	110	45	131	44	121	46

Table 7: Average consonant duration in “voiced” and “voiceless” contexts in whispered conversational and clear speech

speaker	condition	C/V duration ratio					
		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD
1	whi conv	0.57	0.18	0.86	0.29	0.71	0.28
	whi clear	0.61	0.22	1.11	0.35	0.86	0.38
2	whi conv	0.52	0.10	0.92	0.22	0.72	0.26
	whi clear	0.78	0.28	1.24	0.41	1.01	0.42
3	whi conv	0.47	0.12	0.69	0.18	0.58	0.19
	whi clear	0.37	0.08	0.59	0.15	0.48	0.16
4	whi conv	0.51	0.17	0.85	0.35	0.68	0.32
	whi clear	0.43	0.13	0.64	0.23	0.53	0.22
Average	whi conv	0.52	0.15	0.83	0.28	0.67	0.27
	whi clear	0.54	0.25	0.89	0.41	0.72	0.38

Table 8: Average C/V duration-ratio in “voiced” and “voiceless” contexts in whispered conversational and clear speech

Speaker	Change in the Acoustic Cue (Clear-Conv), %					
	vowel duration		consonant duration		CVDR	
	voiceless	voiced	voiceless	voiced	voiceless	voiced
1	-7%	11%	22%	18%	28%	7%
2	9%	21%	44%	75%	35%	48%
3	16%	22%	-1%	-4%	-15%	-22%
4	20%	22%	-7%	3%	-25%	-17%
Average	10%	19%	14%	23%	6%	4%

Table 9: Percentage change in the acoustic parameters in clear whispered speech relative to conversational whispered speech in two “voicing” contexts

speaker	consonant pair	vowel duration contrast		
		pho conv	whi conv	% change
1	b/p	21.3725	21.6350	1
	g/k	30.5563	41.0375	34
	z/s	35.2000	32.7013	-7
	dg/ch	22.2825	27.7838	25
	Average	**27.3528**	**30.7894**	**13**
2	b/p	16.6425	39.1788	135
	g/k	16.7025	38.2650	129
	z/s	16.7150	36.8638	121
	dg/ch	29.2900	41.2100	41
	Average	**19.8375**	**38.8794**	**106**
3	b/p	47.5388	30.5175	-36
	g/k	51.2550	32.0013	-38
	z/s	32.3175	19.7888	-39
	dg/ch	47.9000	34.9175	-27
	Average	**44.7528**	**29.3063**	**-35**
4	b/p	24.7396	27.1100	10
	g/k	28.5613	24.0875	-16
	z/s	40.3388	30.5700	-24
	dg/ch	33.7900	36.1525	7
	Average	**31.8574**	**29.4800**	**-6**

(a)

consonant pair	vowel duration contrast		
	pho conv	whi conv	% change
b/p	27.57	29.61	28
g/k	31.77	33.85	28
z/s	31.14	29.98	13
ʤ/ʧ	33.32	35.02	11
Average	**30.95**	**32.11**	**20**

(b)

Table 10: Magnitude of the vowel duration contrast in different speakers (a) and consonant-pairs (b) in conversational phonated speech and whispered speech. Last column shows change in the contrast magnitude in conversational whispered relative to conversational phonated speech expressed as a percentage. Vowel duration contrasts in (b) are average values across speakers

speaker	consonant pair	consonant duration contrast		
		pho conv	whi conv	% change
1	b/p	21.0188	11.3913	-46
	g/k	18.1938	23.1513	27
	z/s	17.7725	13.8113	-22
	dg/ch	26.3000	23.3500	-11
	Average	**20.8213**	**17.9259**	**-13**
2	b/p	20.3013	37.3088	84
	g/k	23.4888	30.0575	28
	z/s	28.8425	35.7950	24
	dg/ch	29.7825	29.0225	-3
	Average	**25.6038**	**33.0459**	**33**
3	b/p	36.0750	21.9600	-39
	g/k	34.4750	26.5063	-23
	z/s	24.0488	10.3063	-57
	dg/ch	25.0800	19.8863	-21
	Average	**29.9197**	**19.6647**	**-35**
4	b/p	26.0021	35.3913	36
	g/k	35.1713	22.7500	-35
	z/s	49.7263	18.7138	-62
	dg/ch	30.3350	32.8913	8
	Average	**35.3086**	**27.4366**	**-13**

(a)

consonant pair	consonant duration contrast		
	pho conv	whi conv	% change
b/p	25.85	26.51	9
g/k	27.83	25.62	-1
z/s	30.10	19.66	-29
ʤ/ʧ	27.87	26.29	-7
Average	**27.91**	**24.52**	**-7**

(b)

Table 11: Magnitude of the consonant duration contrast in different speakers and consonant-pairs in conversational phonated speech and whispered speech. Last column shows change in the contrast magnitude in conversational whispered relative to conversational phonated speech expressed as a percentage

speaker	consonant pair	CV duration ratio contrast		
		pho conv	whi conv	% change
1	b/p	.3713	.2288	-38
	g/k	.3400	.3525	4
	z/s	.3163	.2413	-24
	dg/ch	.3800	.3613	-5
	Average	**.3519**	**.2959**	**-16**
2	b/p	.2163	.4625	114
	g/k	.2050	.3463	69
	z/s	.2175	.3350	54
	dg/ch	.2925	.4075	39
	Average	**.2328**	**.3878**	**69**
3	b/p	.4063	.2538	-38
	g/k	.3550	.2888	-19
	z/s	.2063	.1088	-47
	dg/ch	.3425	.2325	-32
	Average	**.3275**	**.2209**	**-34**
4	b/p	.3508	.4288	22
	g/k	.3963	.2375	-40
	z/s	.5550	.2338	-58
	dg/ch	.4913	.4625	-6
	Average	**.4483**	**.3406**	**-20**

(a)

consonant pair	CV duration contrast		
	pho conv	whi conv	% change
b/p	0.3361	0.3434	15
g/k	0.3241	0.3063	3
z/s	0.3238	0.2297	-19
ʤ/ʧ	0.3766	0.3659	-1
Average	**0.3401**	**0.3113**	**0**

(b)

Table 12: Magnitude of the consonant-to-vowel duration-ratio contrast in different speakers and consonant-pairs in conversational phonated speech and whispered speech. Last column shows change in the contrast magnitude in conversational whispered relative to conversational phonated speech expressed as a percentage

speaker	consonant pair	*F1*-offset dynamics contrast		
		pho conv	whi conv	% change
1	b/p	.3570	.3334	-7
	g/k	.9318	.2350	-75
	z/s	1.0140	-.3007	-130
	dg/ch	.4666	-.2000	-143
	Average	**.6923**	**.0169**	**-88**
2	b/p	.3919	.6286	60
	g/k	-.4386	.6216	242
	z/s	1.0903	.2190	-80
	dg/ch	-.4785	.4695	198
	Average	**.1413**	**.4847**	**105**
3	b/p	1.2857	.1789	-86
	g/k	2.2972	.7984	-65
	z/s	2.3710	.2009	-92
	dg/ch	.8862	.2225	-75
	Average	**1.7100**	**.3501**	**-79**
4	b/p	.3558	.3481	-2
	g/k	.3794	1.0559	178
	z/s	.9388	.5391	-43
	dg/ch	.3951	.8224	108
	Average	**.5173**	**.6914**	**60**

(a)

consonant pair	F1-offset dynamics contrast		
	pho conv	whi conv	% change
b/p	0.5976	0.3722	-9
g/k	0.7924	0.6777	70
z/s	1.3535	0.1646	-86
dg/ch	0.3174	0.3286	-77
Average	**0.7652**	**0.3858**	**-25**

(b)

Table 13: Magnitude of the *F1* offset-dynamics contrast in different speakers and consonant-pairs in conversational phonated and whispered speech. Last column shows change in the contrast magnitude in conversational whispered relative to conversational phonated speech expressed as a percentage

stimulus played \ response selected	b	p	g	k	z	s	ʤ	ʧ
b	351 (91.40%)	33 (8.59%)						
p	3 (0.78%)	381 (99.21%)						
g			381 (99.21%)	3 (0.78%)				
k				384 (100%)				
z					384 (100%)			
s	1 (0.26%)					384 (100%)		
ʤ							384 (100%)	
ʧ							4 (1.04%)	380 (98.96%)

(a)

Speaker	Number of errors	% errors
1	5	0.65
2	16	2.08
3	10	1.30
4	13	1.69

(b)

Table 14: Consonant perception errors in conversational phonated speech by consonant (a) and by speaker (b)

stimulus played \ response selected	b	p	g	k	z	s	ʤ	ʧ
b	224 (58%)	160 (42%)						
p	40 (10%)	344 (90%)						
g	1 (< 1%)		326 (85%)	57 (15%)				
k			11 (3%)	373 (97%)				
z					338 (88%)	45 (12%)	1 (< 1%)	
s					76 (20%)	306 (80%)	2 (< 1%)	
ʤ			1 (< 1%)				339 (88%)	44 (11%)
ʧ							71 (18%)	313 (82%)

(a)

Speaker	Number of errors	% errors
1	133	17.32
2	113	14.71
3	129	16.80
4	130	16.93

(b)

Table 15: Consonant perception errors in conversational whispered speech by consonant (a) and by speaker (b). Confusion matrix shows actual number of responses accompanied by percentages in the parenthesis

speaker	whi conv	whi clear
1	84.65	92.40
2	88.19	94.17
3	84.48	82.53
4	84.03	80.52
Average	85.34	87.40

Table 16: Consonant perception in individual speakers in whispered conversational and clear speech. The values are percentages of correct responses

stimulus played \ response selected	b	p	g	k	z	s	ʤ	ʧ
b	73.89	25.90	0.07	0.14				
p	13.33	86.39		0.14		0.14		
g	0.28	0.07	84.17	15.35			0.14	
k			4.65	95.35				
z					83.13	16.60	0.21	0.07
s					11.25	88.75		
ʤ			0.21				91.88	7.92
ʧ	0.07				0.07		20.69	79.17

(a)

stimulus played \ response selected	b	p	g	k	z	s	ʤ	ʧ
b	90.42	9.03	0.42	0.07		0.07		
p	12.01	87.36	0.21	0.28		0.07		0.07
g	0.21		91.53	8.13	0.07		0.07	
k		0.07	8.68	91.25				
z	0.07				80.00	19.72	0.21	
s			0.07		13.19	86.74		
ʤ			0.35		0.07	0.07	92.15	7.36
ʧ	0.14		0.07				20.00	79.79

(b)

Table 17: Error pattern in consonant perception in conversational whispered (a) and clear whispered (b) speech. Percentage of correct responses are displayed. Errors were observed on "voicing"; errors on place and manner of articulation did not exceed 1%.

speaker	pair	whi conv		whi clear		gain (RAU)	
		RAU	A'	RAU	A'	difference	proportion
1	b/p	71.55	0.78	95.30	0.94	24	0.40
	g/k	99.23	0.96	110.01	0.99	11	0.13
	z/s	91.32	0.93	97.09	0.95	6	0.06
	ʤ/ʧ	95.19	0.95	97.61	0.96	2	0.03
	Average	89.32	0.90	100.00	0.96	10.68	0.15
2	b/p	94.34	0.94	104.07	0.98	10	0.11
	g/k	99.45	0.96	110.04	0.99	11	0.11
	z/s	96.55	0.95	94.18	0.93	-2	-0.03
	ʤ/ʧ	81.53	0.87	99.15	0.96	18	0.24
	Average	92.97	0.93	101.86	0.97	8.89	0.11
3	b/p	82.91	0.89	93.05	0.94	10	0.16
	g/k	100.88	0.96	87.35	0.91	-14	-0.13
	z/s	84.61	0.89	80.07	0.87	-5	-0.05
	ʤ/ʧ	87.85	0.91	82.54	0.88	-5	-0.07
	Average	89.06	0.91	85.75	0.90	-3.31	-0.02
4	b/p	86.52	0.90	86.67	0.90	0	0.00
	g/k	86.30	0.90	87.78	0.92	1	0.03
	z/s	83.14	0.89	75.91	0.83	-7	-0.09
	ʤ/ʧ	93.68	0.95	81.91	0.89	-12	-0.13
	Average	87.41	0.91	83.07	0.88	-4.39	-0.05

(a)

pair	whi conv		whi clear		gain (RAU)	
	RAU	A'	RAU	A'	difference	proportion
b/p	83.83	0.88	94.77	0.94	10.94	0.17
g/k	96.47	0.95	98.80	0.95	2.33	0.03
z/s	88.90	0.91	86.81	0.89	-2.14	-0.03
ʤ/ʧ	89.56	0.92	90.30	0.92	0.74	0.02
Average	89.69	0.91	92.67	0.93	2.97	0.05

(b)

Table 18: Accuracy of consonant perception in whispered conversational and clear speech. Scores expressed in RAU and A′ in four consonant-pairs in individual speakers (a) and across the speakers (b) in two speaking styles. The average intelligibility gain in clear speech is displayed as a difference between clear and conversational RAU scores and as a proportional increase relative to the conversational condition

Predictor	B	SE	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							LB	UB
Intercept	-3.066	2.950	1.080	1	.299			
zCVDR	-17.378	6.523	7.099	1	.008	.000	.000	.010
zF1offset	4.054	1.720	5.558	1	.018	57.627	1.981	1676.188
zCVAR	-2.876	1.675	2.946	1	.086	.056	.002	1.504
consonantPair			5.345	3	.148			
[consonantPair=1.00]	11.895	5.570	4.561	1	.033	146531.786	2.660	8072715739.611
[consonantPair=2.00]	2.145	3.554	.364	1	.546	8.543	.008	9052.414
[consonantPair=3.00]	-1.805	3.216	.315	1	.575	.164	.000	89.863
[consonantPair=4.00]		.	.	0	.	.	.	.

Table 19: Logistic regression analysis of consonant "voicing" in conversational phonated speech with C/V duration-ratio, *F1*-offset dynamics, and C/V amplitude-ratio as predictor variables. C/V duration-ratio emerged as the most important predictor of consonant "voicing"

Predictor	B	SE	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							LB	UB
Intercept	.247	.543	.207	1	.649			
zCVDR	-3.736	.432	74.884	1	.000	.024	.010	.056
zF1offset	.693	.329	4.438	1	.035	2.000	1.049	3.811
consonantPair			1.027	3	.795			
[consonantPair=1.00]	-.751	.791	.902	1	.342	.472	.100	2.223
[consonantPair=2.00]	-.193	.758	.065	1	.799	.825	.187	3.641
[consonantPair=3.00]	-.151	.722	.044	1	.834	.860	.209	3.536
[consonantPair=4.00]				0				

Table 20: Logistic regression analysis of consonant "voicing" in conversational whispered speech with C/V duration-ratio and *F1*-offset dynamics as predictor variables. C/V duration-ratio emerged as the most important predictor of consonant "voicing"

speaker	pair	prediction success %	rank prediction	A'-score	rank A'
1	b/p	93.75	7.0	0.79	1.0
	g/k	100.00	12.5	0.96	13.0
	z/s	87.50	4.0	0.93	9.0
	ʤ/ʧ	100.00	12.5	0.95	12.0
	Average	95.31	9.00	0.91	8.75
2	b/p	100.00	12.5	0.95	11.0
	g/k	100.00	12.5	0.96	16.0
	z/s	100.00	12.5	0.96	14.0
	ʤ/ʧ	93.33	5.0	0.86	2.0
	Average	98.33	10.625	0.93	10.75
3	b/p	100.00	12.5	0.89	4.0
	g/k	81.25	2.5	0.96	15.0
	z/s	81.25	2.5	0.89	5.0
	ʤ/ʧ	93.75	7.0	0.91	8.0
	Average	89.06	6.125	0.91	8.00
4	b/p	93.75	7.0	0.88	3.0
	g/k	75.00	1.0	0.90	7.0
	z/s	100.00	12.5	0.89	6.0
	ʤ/ʧ	100.00	12.5	0.95	10.0
	Average	92.19	8.25	0.90	6.50
Total	b/p	96.88	9.75	0.88	4.75
	g/k	89.06	7.125	0.94	12.75
	z/s	92.19	7.875	0.92	8.50
	ʤ/ʧ	96.77	9.25	0.92	8.00
	Average	93.72	8.50	0.91	8.50

Table 21: Model prediction success and observed perception accuracy of consonant "voicing" in conversational whispered speech. Model prediction success is based on the the logistic regression analysis; perception accuracy is based on the results of the perceptual test and is expressed as A'-scores for individual consonant-pairs

speaker	pair	prediction success %	rank prediction	A'-score	rank A'
1	b/p	100.00	12.5	0.95	9.0
	g/k	100.00	12.5	0.99	15.0
	z/s	75.00	1.0	0.95	11.0
	ʤ/ʧ	100.00	12.5	0.96	12.0
	Average	93.75	9.625	0.96	11.75
2	b/p	81.25	3.0	0.99	14.0
	g/k	93.75	7.0	0.99	16.0
	z/s	100.00	12.5	0.95	10.0
	ʤ/ʧ	81.25	3.0	0.96	13.0
	Average	89.06	6.375	0.97	13.25
3	b/p	100.00	12.5	0.93	8.0
	g/k	93.75	7.0	0.92	6.0
	z/s	100.00	12.5	0.87	2.0
	ʤ/ʧ	100.00	12.5	0.89	4.0
	Average	98.44	11.125	0.90	5.00
4	b/p	93.75	7.0	0.90	5.0
	g/k	87.50	5.0	0.92	7.0
	z/s	81.25	3.0	0.81	1.0
	ʤ/ʧ	100.00	12.5	0.89	3.0
	Average	90.63	6.875	0.88	4.00
Total	b/p	93.75	8.75	0.94	9.00
	g/k	93.75	7.875	0.95	11.00
	z/s	89.06	7.250	0.89	6.00
	ʤ/ʧ	95.31	10.13	0.92	8.00
	Average	92.97	8.50	0.93	8.50

Table 22: Model prediction success and observed perception accuracy of consonant "voicing" in whispered clear speech. Model prediction success is calculated based on the result of fitting the model developed on the whispered conversational speech data into the whispered clear speech data; perception accuracy is based on the results of the perceptual test and is expressed as A'-scores for individual consonant-pairs

Appendix A: Conversational phonated vs. whispered speech

speaker	phonation	context																			
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	phonated	123	20	102	23	139	25	109	25	157	22	122	28	134	28	111	20	138	26	111	24	125	28
	whispered	139	19	118	29	164	20	123	19	170	27	137	27	159	28	132	29	158	25	127	26	143	30
2	phonated	151	21	134	19	174	22	157	27	190	15	173	26	182	23	153	21	174	25	154	27	164	27
	whispered	169	16	130	19	183	10	145	17	202	21	165	24	183	13	139	20	184	19	145	23	165	29
3	phonated	179	30	131	21	198	22	146	26	210	23	178	32	193	28	145	24	195	27	150	30	173	36
	whispered	180	20	149	22	193	17	161	35	194	19	174	20	191	31	156	15	189	22	160	25	174	28
4	phonated	142	39	114	33	148	42	119	26	167	34	127	30	145	43	111	30	150	39	118	29	134	38
	whispered	166	36	139	52	173	49	149	38	190	39	159	38	173	35	137	34	176	39	146	40	161	42
Average	phonated	149	34	120	27	165	36	133	32	181	31	150	38	163	39	130	30	164	37	133	33	149	38
	whispered	164	27	134	34	178	29	144	31	189	29	159	30	176	29	141	26	177	30	145	31	161	34

Table A1: Vowel duration in conversational phonated and whispered speech

Source	SS	df	MS	F	P
Between-Subjects	477810.75	127			
Speaker	120832.42	3	40277.48	13.99	0.000
Error(Speaker)	356978.33	124	2878.86		
Within-Subjects	215398.67	384			
Phonation	18141.08	1	18141.08	117.94	0.000
Speaker x Phonation	15361.26	3	5120.42	33.29	0.000
Error(Speaker x Phonation)	19073.99	124	153.82		
Voicing	127511.24	1	127511.24	860.60	0.000
Speaker x Voicing	1352.16	3	450.72	3.04	0.032
Error(Speaker x Voicing)	18372.45	124	148.17		
Phonation x Voicing	25.27	1	25.27	0.29	0.590
Speaker x Phonation x Voicing	4824.04	3	1608.01	18.57	0.000
Error(Phonation-Voicing)	10737.19	124	86.59		
Total	693209.42	511			

Table A2: ANOVA table for the effects of Voicedness, Speaker, and Phonation-mode on vowel duration.

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig	95% Confidence Interval for Differencea	
						Lower Bound	Upper Bound
1	phonated	whispered	-18.13*	5.053	.000	-28.056	-8.199
2	phonated	whispered	-.427	5.074	.933	-10.395	9.542
3	phonated	whispered	-1.933	5.053	.702	-11.861	7.996
4	phonated	whispered	-26.719*	5.053	.000	-36.647	-16.790

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A3: Pairwise comparisons of vowel durations between conversational phonated and whispered speech. Negative difference indicates increase of vowel duration in whispered speech. Lengthening of vowels was significant only in Speaker 1 and Speaker 4

speaker	phonation	context																			
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	phonated	73	6	94	10	75	8	93	7	91	9	109	28	101	14	128	8	85	15	106	21	95	21
	whispered	73	9	84	10	71	9	94	6	89	8	103	18	114	15	137	13	87	20	105	24	96	24
2	phonated	76	13	96	7	71	7	95	10	74	9	103	6	89	9	118	8	78	11	103	12	90	17
	whispered	86	6	124	10	88	15	118	9	91	7	127	9	118	7	149	12	96	16	129	15	112	22
3	phonated	69	25	105	8	57	7	91	7	70	9	94	9	91	9	116	9	72	19	101	13	87	22
	whispered	87	7	109	4	77	14	103	18	83	16	94	9	100	13	120	14	87	15	106	15	97	18
4	phonated	64	7	91	7	60	7	95	8	65	7	115	9	90	10	120	12	70	14	105	16	87	23
	whispered	64	9	99	17	77	16	100	17	94	6	113	10	108	10	141	13	86	20	113	22	99	25
Average	phonated	70	15	96	9	66	10	93	8	75	13	105	17	93	12	121	10	76	16	104	16	90	21
	whispered	78	12	104	18	78	14	104	16	90	11	109	17	110	13	136	16	89	18	113	21	101	23

Table A4: Consonant duration in conversational phonated and whispered speech

Source	SS	df	MS	F	P
Between-Subjects	106524.08	127			
Speaker	6818.51	3	2272.84	2.83	0.041
Error(Speaker)	99705.57	124	804.08		
Within-Subjects	160481.25	384			
Phonation	15719.86	1	15719.86	130.40	0.000
Speaker x Phonation	7319.26	3	2439.75	20.24	0.000
Error(Speaker x Phonation)	14948.63	124	120.55		
Voicing	88278.77	1	88278.77	738.04	0.000
Speaker x Voicing	2760.98	3	920.33	7.69	0.000
Error(Speaker x Voicing)	14832.01	124	119.61		
Phonation x Voicing	369.58	1	369.58	3.10	0.081
Speaker x Phonation x Voicing	1468.73	3	489.58	4.11	0.008
Error(Phonation-Voicing)	14783.45	124	119.22		
Total	267005.33	511			

Table A5: ANOVA table for the effects of consonant Voicedness, Speaker and Phonation-mode on consonant duration.

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	Std. Error	Sig.a	95% Confidence Interval for Differencea	
						Lower Bound	Upper Bound
1	phonated	whispered	-0.438	3.017	.885	-6.366	5.490
2	phonated	whispered	-21.707*	3.029	.000	-27.659	-15.756
3	phonated	whispered	-10.061*	3.017	.001	-15.989	-4.134
4	phonated	whispered	-12.0706*	3.017	.000	-17.998	-6.143

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A6: Pairwise comparisons of consonant durations between conversational phonated and whispered speech. Negative difference indicates increase of consonant duration in whispered speech. Lengthening of vowels was significant only in Speakers 2, 3 and 4.

speaker	phonation	context																			
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	phonated	0.60	0.08	0.97	0.28	0.55	0.13	0.89	0.21	0.59	0.09	0.90	0.22	0.81	0.26	1.19	0.26	0.64	0.18	0.99	0.26	0.81	0.28
	whispered	0.54	0.14	0.77	0.25	0.44	0.07	0.79	0.19	0.54	0.13	0.79	0.24	0.75	0.22	1.11	0.34	0.57	0.18	0.86	0.29	0.71	0.28
2	phonated	0.52	0.13	0.74	0.14	0.42	0.08	0.62	0.15	0.40	0.06	0.61	0.13	0.50	0.10	0.79	0.13	0.46	0.11	0.69	0.15	0.57	0.17
	whispered	0.52	0.07	0.98	0.21	0.48	0.08	0.83	0.14	0.46	0.07	0.79	0.17	0.65	0.08	1.10	0.23	0.52	0.10	0.92	0.21	0.72	0.26
3	phonated	0.41	0.23	0.82	0.15	0.29	0.06	0.64	0.14	0.34	0.07	0.54	0.12	0.49	0.11	0.83	0.20	0.38	0.15	0.71	0.19	0.54	0.24
	whispered	0.50	0.09	0.75	0.13	0.40	0.05	0.69	0.25	0.44	0.12	0.55	0.10	0.54	0.17	0.78	0.13	0.47	0.12	0.69	0.18	0.58	0.19
4	phonated	0.49	0.17	0.86	0.26	0.43	0.12	0.83	0.18	0.40	0.06	0.95	0.24	0.68	0.25	1.18	0.42	0.50	0.19	0.95	0.31	0.73	0.34
	whispered	0.41	0.14	0.84	0.39	0.48	0.17	0.72	0.24	0.51	0.12	0.75	0.20	0.65	0.16	1.11	0.42	0.51	0.17	0.85	0.35	0.68	0.32
Average	phonated	0.50	0.17	0.84	0.22	0.42	0.14	0.75	0.20	0.43	0.12	0.75	0.25	0.62	0.23	1.00	0.32	0.49	0.18	0.83	0.27	0.66	0.29
	whispered	0.49	0.12	0.83	0.26	0.45	0.11	0.75	0.21	0.49	0.11	0.72	0.20	0.65	0.17	1.02	0.32	0.52	0.15	0.83	0.28	0.67	0.27

Table A7: Consonant-to-vowel duration ratio in conversational phonated and whispered speech

Source	SS	df	MS	F	P
Between-Subjects	19.22	127			
Speaker	2.83	3	0.94	7.14	0.000
Error(Speaker)	16.39	124	0.13		
Within-Subjects	20.49	384			
Phonation	0.01	1	0.01	1.31	0.254
Speaker x Phonation	1.06	3	0.35	39.36	0.000
Error(Speaker x Phonation)	1.12	124	0.01		
Voicing	13.62	1	13.62	635.03	0.000
Speaker x Voicing	0.25	3	0.08	3.91	0.010
Error(Speaker x Voicing)	2.66	124	0.02		
Phonation x Voicing	0.03	1	0.03	2.57	0.111
Speaker x Phonation x Voicing	0.38	3	0.13	11.52	0.000
Error(Phonation-Voicing)	1.36	124	0.01		
Total	39.70	511			

Table A8: ANOVA table for the effects of Voicedness, Speaker and Phonation-mode on C/V duration-ratio.

Dependent Variable:CV duration ratio

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	phonated	whispered	0.098*	.037	.008	.025	.170
2	phonated	whispered	-0.147*	.037	.000	-.220	-.074
3	phonated	whispered	-0.035	.037	.345	-.107	.038
4	phonated	whispered	0.0434	.037	.239	-.029	.116

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A9: Pairwise comparisons of C/V duration-ratio between conversational phonated and whispered speech. Negative difference indicates increase of CVDR in whispered speech. Consonant-to-vowel duration-ratio significantly decreased in Speaker 1 and significantly increased in Speaker 2 in whispered speech

speaker	phonation	context																					
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	phonated	0.70	0.11	0.34	0.28	1.28	0.37	0.35	0.25	1.42	0.31	0.41	0.42	1.01	0.53	0.55	0.27	1.10	0.44	0.41	0.31	0.76	0.52
	whispered	0.63	0.42	0.30	0.54	0.62	0.90	0.39	0.54	0.67	0.99	0.97	0.42	0.46	0.73	0.66	0.70	0.60	0.76	0.58	0.59	0.59	0.67
2	phonated	0.92	0.38	0.53	0.27	1.76	0.54	2.20	1.74	1.81	0.53	0.72	1.35	1.64	0.68	2.12	1.46	1.53	0.63	1.39	1.49	1.46	1.13
	whispered	0.69	0.72	0.07	0.36	1.00	0.88	0.38	0.49	0.67	0.78	0.45	0.65	0.95	0.58	0.36	0.35	0.83	0.73	0.31	0.48	0.58	0.67
3	phonated	1.55	0.30	0.26	0.43	2.64	0.79	0.34	0.40	2.44	0.85	0.07	0.33	1.96	0.47	1.07	0.61	2.15	0.75	0.44	0.58	1.29	1.09
	whispered	0.74	0.41	0.56	0.52	1.08	0.64	0.28	0.50	0.83	0.67	0.63	0.36	0.82	0.48	0.60	0.54	0.87	0.55	0.52	0.48	0.69	0.54
4	phonated	0.63	0.20	0.28	0.70	1.18	0.88	0.81	0.78	1.35	0.92	0.41	0.71	0.88	0.34	0.49	0.41	1.01	0.69	0.50	0.66	0.75	0.72
	whispered	0.90	0.46	0.56	0.56	1.17	0.58	0.11	0.63	0.68	0.86	0.14	0.58	1.31	0.68	0.49	0.39	1.02	0.68	0.32	0.56	0.67	0.71
Average	phonated	0.95	0.45	0.35	0.45	1.72	0.87	0.92	1.21	1.76	0.80	0.40	0.81	1.37	0.67	1.06	1.03	1.45	0.78	0.68	0.96	1.07	0.95
	whispered	0.74	0.50	0.37	0.52	0.97	0.76	0.29	0.53	0.71	0.79	0.55	0.58	0.89	0.67	0.53	0.51	0.83	0.69	0.43	0.54	0.63	0.65

Table A10: *F1*-offset dynamics in conversational phonated and whispered speech

Source	SS	df	MS	F	P
Between-Subjects	90.66	127			
Speaker	12.86	3	4.29	6.83	0.000
Error(Speaker)	77.80	124	0.63		
Within-Subjects	271.19	384			
Phonation	23.99	1	23.99	56.50	0.000
Speaker x Phonation	13.45	3	4.48	10.56	0.000
Error(Speaker x Phonation)	52.66	124	0.43		
Voicing	42.56	1	42.56	86.00	0.000
Speaker x Voicing	10.33	3	3.44	6.96	0.000
Error(Speaker x Voicing)	61.37	124	0.50		
Phonation x Voicing	4.55	1	4.55	11.96	0.001
Speaker x Phonation x Voicing	15.14	3	5.05	13.28	0.000
Error(Phonation-Voicing)	47.14	124	0.38		
Total	361.85	511			

Table A11: ANOVA table for the effects of “Voicing”, Speaker, and Phonation-mode on *F1*-offset dynamics.

Dependent Variable: *F1*-offset dynamics

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	phonated	whispered	0.168	.123	.173	-.073	.409
2	phonated	whispered	0.892*	.123	.000	.650	1.134
3	phonated	whispered	0.600*	.123	.000	.359	.841
4	phonated	whispered	0.0840	.123	.494	-.157	.325

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A12: Pairwise comparisons of *F1*-offset dynamics between conversational phonated and whispered speech. *F1*-offset was significantly less dynamic in whispered speech for Speakers 2 and 3

Dependent Variable: *F1*-offset dynamics

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
phonated	voiced	voiceless	0.765*	.087	.000	.595	.936
whispered	voiced	voiceless	0.394*	.087	.000	.223	.565

Table A13: Pairwise comparisons of *F1*-offset dynamics between "voiced" and "voiceless" contexts in conversational phonated and whispered speech. The differences between two "voicing" contexts was significant in both phonation modes

difference in contrast (whi – pho)	Mean	SD	SE	95% CI of the Difference Lower	Upper	t	df	Sig.
vowel duration	1.164	17.673	2.209	-3.251	5.578	.527	63	.300
consonant duration	-3.395	17.246	2.156	-7.703	0.913	-1.575	63	.060
CVDR	-0.029	0.184	0.023	-0.075	0.017	-1.250	63	.108
F1-offset dynamics	-0.379	1.207	0.151	-0.681	-0.078	-2.514	63	.007
CVAR	-4.965	5.771	0.721	-6.407	-3.523	-6.882	63	.000

Table A14: Difference in "voicing" contrasts between conversational phonated and whispered speech

					95% CI of the Difference				
Speaker	Paired difference	Mean	SD	SEM	LB	UB	t	df	Sig. [1]
1	pho - whi	4.77	4.61	1.15	2.31	7.23	4.139	15	0.00
2	pho - whi	3.38	4.67	1.17	0.89	5.87	2.896	15	0.01
3	pho - whi	4.07	4.70	1.18	1.57	6.58	3.467	15	0.00
4	pho - whi	7.63	7.97	1.99	3.38	11.88	3.830	15	0.00

[1]one-tailed

Table A15: Pairwise comparisons of the CVAR contrast between conversational phonated and whispered speech. The CVAR contrast in whispered speech was significantly smaller for all four speakers.

speaker	phonation	context																					
		b		p		g		k		z		s		ʤ		ʧ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	phonated	-12.14	1.64	-17.32	2.82	-16.25	1.28	-20.56	2.23	-11.06	1.71	-10.31	2.08	-14.16	1.46	-19.23	2.77	-13.40	2.49	-16.86	4.66	-15.13	4.10
	whispered	-6.65	3.28	-3.22	3.61	-10.65	2.23	-9.70	2.15	2.42	2.83	4.14	3.79	-7.67	1.47	-8.52	1.96	-5.64	5.51	-4.32	6.24	-4.98	5.88
2	phonated	-20.34	2.30	-24.61	1.83	-15.21	0.93	-20.49	3.13	-15.51	1.88	-16.63	2.80	-16.14	1.94	-19.40	2.62	-16.80	2.73	-20.28	3.84	-18.54	3.74
	whispered	3.26	3.96	7.26	3.25	-2.99	3.41	-4.45	2.99	3.94	1.95	3.82	2.60	-0.63	3.54	-3.37	2.47	0.90	4.26	0.95	5.68	0.92	4.97
3	phonated	-17.47	3.97	-21.23	1.12	-19.75	2.78	-24.91	2.03	-11.54	1.46	-10.07	1.42	-12.91	1.37	-19.67	2.83	-15.42	4.20	-18.97	5.87	-17.19	5.37
	whispered	-10.04	4.16	-8.77	1.94	-13.24	4.98	-13.83	3.65	2.20	6.39	4.45	3.31	-5.21	3.42	-6.04	2.93	-6.57	7.50	-6.05	7.37	-6.31	7.38
4	phonated	-13.59	2.25	-26.01	3.02	-15.82	2.30	-25.98	2.13	-8.56	1.58	-6.13	1.13	-12.37	1.38	-16.52	2.86	-12.59	3.24	-18.66	8.64	-15.62	7.16
	whispered	-17.12	3.66	-12.90	3.73	-12.78	4.01	-11.88	3.34	7.48	3.75	6.53	2.64	-4.51	4.24	-2.44	3.90	-6.73	10.22	-5.17	8.66	-5.95	9.43
Average	phonated	-15.88	4.15	-22.29	4.06	-16.76	2.59	-22.98	3.42	-11.67	2.99	-10.78	4.25	-13.89	2.09	-18.70	2.93	-14.55	3.60	-18.69	6.09	-16.62	5.40
	whispered	-7.64	8.28	-4.41	8.26	-9.91	5.52	-9.96	4.61	4.01	4.43	4.74	3.16	-4.50	4.07	-5.15	3.69	-4.51	7.82	-3.68	7.51	-4.10	7.66

Table A16: Consonant-to-vowel amplitude-ratio in conversational phonated and whispered speech

Source	SS	df	MS	F	P
Between-Subjects	13930.31	127			
Speaker	557.70	3	185.90	1.72	0.166
Error(Speaker)	13372.61	124	107.84		
Within-Subjects	28371.00	384			
Phonation	19900.13	1	19900.13	939.14	0.000
Speaker x Phonation	2044.36	3	681.45	32.16	0.000
Error(Speaker x Phonation)	2627.52	124	21.19		
Voicing	334.76	1	334.76	31.81	0.000
Speaker x Voicing	22.26	3	7.42	0.71	0.551
Error(Speaker x Voicing)	1304.98	124	10.52		
Phonation x Voicing	770.28	1	770.28	74.54	0.000
Speaker x Phonation x Voicing	85.39	3	28.46	2.76	0.045
Error(Phonation-Voicing)	1281.33	124	10.33		
Total	42301.31	511			

Table A17: ANOVA table for the effects of Consonant-pair, Speaker, and Phonation-mode on C/V amplitude ratio between “voiced” and “voiceless” consonants.

Source	SS	df	MS	F	P
Between-Subjects	128.492	15			
Speaker	22.148	3	7.383	0.833	0.501
Error(Speaker)	106.344	12	8.862		
Within-Subjects	2651.38	112			
Phonation	785.07	1	785.07	99.069	0
Speaker x Phonation	84.461	3	28.154	3.553	0.048
Error(Speaker x Phonation)	95.094	12	7.924		
ConsonantPair	304.023	3	101.341	10.14	0
Speaker x ConsonantPair	151.07	9	16.786	1.68	0.13
Error(Speaker x ConsonantPair)	359.781	36	9.994		
Phonation x ConsonantPair	391.211	3	130.404	16.824	0
Speaker x Phonation x ConsonantPair	201.633	9	22.404	2.89	0.011
Error(Phonation-ConsonantPair)	279.031	36	7.751		
Total	2779.87	127			

Table A18: ANOVA table for the effects of Consonant-pair, Speaker, and Phonation-mode on consonant-to-vowel amplitude-ratio contrast.

Dependent Variable:C/V amplitude ratio

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
phonated	voiced	voiceless	4.139*	.809	.000	2.550	5.728
whispered	voiced	voiceless	-0.827	.810	.308	-2.419	.764

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A19: Pairwise comparisons of C/V amplitude-ratio between "voiced" and "voiceless" contexts in conversational phonated and whispered speech. The difference in C/V amplitude-ratio in "voiced" vs. "voiceless" context was significantly different only in phonated speech

Dependent Variable:C/V amplitude ratio

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	whispered	phonated	10.149*	1.080	.000	8.027	12.271
2	whispered	phonated	19.463*	1.084	.000	17.333	21.594
3	whispered	phonated	10.884*	1.080	.000	8.762	13.006
4	whispered	phonated	9.672*	1.080	.000	7.550	11.794

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table A20: Pairwise comparisons of C/V amplitude-ratio between conversational phonated and whispered speech. C/V amplitude-ratio was significantly smaller in whispered speech in all speakers

Appendix B: Whispered conversational vs. clear speech

speaker	condition	context																			
		b		p		g		k		z		s		dʒ		ʧ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	whi conv	139	19	118	29	164	20	123	19	170	27	137	27	159	28	132	29	158	25	127	26	143	30
	whi clear	154	30	107	22	178	30	118	23	193	36	131	26	178	27	119	16	176	33	119	23	147	40
2	whi conv	169	16	130	19	183	10	145	17	202	21	165	24	183	13	139	20	184	19	145	23	165	29
	whi clear	200	43	129	31	224	47	154	28	239	31	194	38	229	46	154	28	223	43	158	38	190	52
3	whi conv	180	20	149	22	193	17	161	35	194	19	174	20	191	31	156	15	189	22	160	25	174	28
	whi clear	219	21	179	25	237	22	192	25	238	25	199	27	228	29	172	21	231	25	185	26	208	34
4	whi conv	166	36	139	52	173	49	149	38	190	39	159	38	173	35	137	34	176	39	146	40	161	42
	whi clear	213	39	167	44	205	42	166	31	227	39	198	49	213	45	171	40	215	40	175	42	195	45
Average	whi conv	164	27	134	34	178	29	144	31	189	29	159	30	176	29	141	26	177	30	145	31	161	34
	whi clear	197	42	146	42	211	41	157	37	224	37	180	45	212	42	154	34	211	41	159	41	185	49

Table B1: Vowel duration in whispered conversational and clear speech

Source	SS	df	MS	F	P
Between-Subjects	561184.99	127			
Speaker	147469.15	3	49156.38	14.73	0.000
Error(Speaker)	413715.84	124	3336.42		
Within-Subjects	448292.81	384			
Style	80159.08	1	80159.08	240.20	0.000
Speaker x Style	18766.46	3	6255.49	18.75	0.000
Error(Speaker x Style)	41381.23	124	333.72		
Voicing	227167.49	1	227167.49	871.47	0.000
Speaker x Voicing	7582.39	3	2527.46	9.70	0.000
Error(Speaker x Voicing)	32323.12	124	260.67		
Style x Voicing	11178.68	1	11178.68	49.08	0.000
Speaker x Style x Voicing	1490.95	3	496.98	2.18	0.094
Error(Style-Voicing)	28243.42	124	227.77		
Total	1009477.80	511			

Table B2: ANOVA table for the effects of consonant Voicedness, Speaker, and Speaking-style on the vowel duration in whispered speech.

speaker	(I) condition	(J) condition	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	whi conv	whi clear	-4.456	5.580	.425	-15.419	6.507
2	whi conv	whi clear	-25.766*	5.602	.000	-36.773	-14.759
3	whi conv	whi clear	-33.546*	5.580	.000	-44.509	-22.583
4	whi conv	whi clear	-34.064*	5.580	.000	-45.028	-23.102

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table B3: Pairwise comparisons of vowel durations between whispered clear and conversational speech. Negative difference indicates increase of vowel duration in clear speech. Lengthening of vowels was significant in all speakers except Speaker 1

difference in contrast (conv – clear)	Mean	SD	SE	95% CI of the Difference Lower	Upper	t	df	Sig.
vowel duration	-19.359	20.017	2.502	-24.359	-14.359	-7.737	63	.000
consonant duration	3.987	20.634	2.579	-1.168	9.141	1.546	63	.064
CVDR	-0.037	0.207	0.026	-0.089	0.014	-1.443	63	.077
F1-offset dynamics	0.021	0.906	0.113	-0.206	0.247	.184	63	.427
CVAR	0.098	4.778	0.597	-1.095	1.292	.165	63	.435

Table B4: Difference in “voicing” contrasts between whispered conversational and clear speech

Source	SS	df	MS	F	P
Between-Subjects	7957.667	15			
Speaker	5897.585	3	1965.862	11.451	0.0008
Error(Speaker)	2060.082	12	171.674		
Within-Subjects	37587.172	112			
Style	11993.52	1	11993.52	41.226	0
Speaker x Style	1570.021	3	523.34	1.799	0.201
Error(Speaker x Style)	3491.054	12	290.921		
ConsonantPair	1651.596	3	550.532	2.201	0.1048
Speaker x ConsonantPair	2314.212	9	257.135	1.028	0.4373
Error(Speaker x ConsonantPair)	9006.543	36	250.182		
Style x ConsonantPair	381.034	3	127.011	0.778	0.514
Speaker x Style x ConsonantPair	1301.734	9	144.637	0.886	0.5468
Error(Style-ConsonantPair)	5877.458	36	163.263		
Total	45544.839	127			

Table B5: ANOVA table for the effects of Speaker, Consonant-pair and Speaking-style on the vowel duration contrast in whispered speech

speaker	condition	context																					
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	whi conv	73	9	84	10	71	9	94	6	89	8	103	18	114	15	137	13	87	20	105	24	96	24
	whi clear	79	11	109	10	80	9	103	17	117	18	120	28	133	19	176	11	102	27	127	34	115	33
2	whi conv	86	6	124	10	88	15	118	9	91	7	127	9	118	7	149	12	96	16	129	15	112	22
	whi clear	143	41	156	35	150	29	166	26	152	27	200	23	227	44	217	37	168	49	185	39	176	45
3	whi conv	87	7	109	4	77	14	103	18	83	16	94	9	100	13	120	14	87	15	106	15	97	18
	whi clear	80	12	103	8	77	14	103	9	78	7	92	9	96	11	124	9	83	13	105	14	94	18
4	whi conv	64	9	99	17	77	16	100	17	94	6	113	10	108	10	141	13	86	20	113	22	99	25
	whi clear	69	15	87	6	81	12	87	18	94	11	104	13	109	13	145	23	88	19	106	29	97	26
Average	whi conv	78	12	104	18	78	14	104	16	90	11	109	17	110	13	136	16	89	18	113	21	101	23
	whi clear	93	37	114	32	97	35	115	36	110	32	129	47	141	57	166	42	110	45	131	44	121	46

Table B6: Consonant duration in whispered conversational and clear speech

Source	SS	df	MS	F	P
Between-Subjects	390027.47	127			
Speaker	189471.66	3	63157.22	39.05	0.000
Error(Speaker)	200555.81	124	1617.39		
Within-Subjects	347150.50	384			
Style	51681.13	1	51681.13	109.92	0.000
Speaker x Style	100976.13	3	33658.71	71.59	0.000
Error(Speaker x Style)	58302.25	124	470.18		
Voicing	64216.32	1	64216.32	184.42	0.000
Speaker x Voicing	73.77	3	24.59	0.07	0.976
Error(Speaker x Voicing)	43178.41	124	348.21		
Style x Voicing	431.45	1	431.45	2.02	0.158
Speaker x Style x Voicing	1750.46	3	583.49	2.73	0.047
Error(Style-Voicing)	26540.59	124	214.04		
Total	737177.97	511			

Table B7: ANOVA table for the effects of consonant Voicedness, Speaker, and Speaking-style on consonant duration in whispered speech

Dependent Variable:consonant duration

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	whi conv	whi clear	-18.994*	4.451	.000	-27.740	-10.248
2	whi conv	whi clear	-64.242*	4.469	.000	-73.023	-55.461
3	whi conv	whi clear	2.440	4.451	.584	-6.306	11.186
4	whi conv	whi clear	2.517	4.451	.572	-6.229	11.263

Table B8: Pairwise comparisons of consonant durations between whispered clear and conversational speech. Negative difference indicates increase of consonant duration in clear speech. Lengthening of consonants was significant in Speakers 1 and 2

speaker	condition	context																					
		b		p		g		k		z		s		dʒ		tʃ		voiced		voiceless		Average	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	whi conv	0.54	0.14	0.77	0.24	0.44	0.07	0.79	0.19	0.55	0.13	0.79	0.24	0.75	0.22	1.11	0.34	0.57	0.18	0.86	0.29	0.71	0.28
	whi clear	0.54	0.18	1.06	0.23	0.46	0.10	0.92	0.30	0.64	0.22	0.94	0.27	0.78	0.23	1.51	0.27	0.61	0.22	1.11	0.35	0.86	0.38
2	whi conv	0.52	0.07	0.98	0.21	0.48	0.08	0.82	0.14	0.46	0.06	0.79	0.17	0.65	0.08	1.09	0.23	0.52	0.10	0.92	0.22	0.72	0.26
	whi clear	0.74	0.27	1.29	0.54	0.69	0.16	1.12	0.29	0.66	0.20	1.08	0.29	1.03	0.31	1.46	0.44	0.78	0.28	1.24	0.41	1.01	0.42
3	whi conv	0.49	0.09	0.75	0.13	0.40	0.05	0.69	0.25	0.44	0.12	0.55	0.10	0.54	0.16	0.78	0.13	0.47	0.12	0.69	0.18	0.58	0.19
	whi clear	0.37	0.06	0.59	0.12	0.33	0.08	0.55	0.11	0.33	0.06	0.47	0.09	0.43	0.10	0.73	0.13	0.37	0.08	0.59	0.15	0.48	0.16
4	whi conv	0.41	0.14	0.83	0.39	0.48	0.17	0.72	0.24	0.51	0.12	0.75	0.20	0.65	0.16	1.11	0.42	0.51	0.17	0.85	0.35	0.68	0.32
	whi clear	0.33	0.09	0.56	0.17	0.41	0.10	0.54	0.14	0.43	0.10	0.56	0.19	0.54	0.17	0.89	0.25	0.43	0.13	0.64	0.23	0.53	0.22
Average	whi conv	0.49	0.12	0.83	0.26	0.45	0.11	0.75	0.21	0.49	0.11	0.72	0.20	0.65	0.17	1.02	0.32	0.52	0.15	0.83	0.28	0.67	0.27
	whi clear	0.50	0.23	0.87	0.43	0.47	0.17	0.78	0.33	0.51	0.21	0.76	0.33	0.69	0.31	1.15	0.44	0.54	0.25	0.89	0.41	0.72	0.38

Table B9: C/V duration-ratio in whispered conversational and clear speech

Source	SS	df	MS	F	P
Between-Subjects	27.81	127			
Speaker	8.99	3	3.00	19.73	0.000
Error(Speaker)	18.82	124	0.15		
Within-Subjects	29.22	384			
Style	0.29	1	0.29	10.76	0.001
Speaker x Style	4.34	3	1.45	54.22	0.000
Error(Speaker x Style)	3.31	124	0.03		
Voicing	13.75	1	13.75	393.25	0.000
Speaker x Voicing	0.86	3	0.29	8.21	0.000
Error(Speaker x Voicing)	4.33	124	0.04		
Style x Voicing	0.06	1	0.06	3.83	0.053
Speaker x Style x Voicing	0.48	3	0.16	10.85	0.000
Error(Style-Voicing)	1.81	124	0.02		
Total	57.03	511			

Table B10: ANOVA table for the effects of consonant Voicedness, Speaker, and Speaking style on consonant-to-vowel duration ratio in whispered speech.

Dependent Variable:CV duration ratio

speaker	(I) phonation mode	(J) phonation mode	Mean Difference (I-J)	SE	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
1	whi conv	whi clear	-0.141*	.042	.001	-.223	-.060
2	whi conv	whi clear	-0.288*	.042	.000	-.370	-.206
3	whi conv	whi clear	0.104*	.042	.013	.022	.186
4	whi conv	whi clear	0.150*	.042	.000	.068	.232

The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

Table B11: Pairwise comparisons of C/V duration-ratio between whispered clear and conversational speech. Negative difference indicates increase in CVDR in whispered clear speech. C/V duration-ratio in clear speech significantly increased in Speakers 1 and 2 and significantly decreased in Speakers 3 and 4

Paired Differences								
				95% Confidence Interval Of the Difference				
comparison	Mean	Std. Deviation	SE	Lower	Upper	t	df	Sig.
speaker1 - speaker2	0.040	0.279	0.070	-0.109	0.189	.568	15	.578
speaker1 - speaker3	0.281	0.173	0.043	0.189	0.374	6.486	15	.000
speaker1 - speaker4	0.291	0.209	0.052	0.180	0.402	5.567	15	.000
speaker2 - speaker3	0.242	0.168	0.042	0.152	0.331	5.760	15	.000
speaker2 - speaker4	0.251	0.163	0.041	0.165	0.338	6.183	15	.000

Table B12: Pairwise comparisons of C/V duration-ratio difference between speakers in whispered clear speech.

Speaker	Paired difference	Mean	SD	SEM	95% CI of the Difference LB	UB	t	df	Sig. [1]
1	convo - clear	-0.21	0.22	0.06	-0.32	-0.09	-3.700	15	0.00
2	convo - clear	-0.07	0.15	0.04	-0.16	0.01	-1.931	15	0.04
3	convo - clear	0.00	0.13	0.03	-0.07	0.07	.019	15	0.49
4	convo - clear	0.13	0.16	0.04	0.04	0.22	3.257	15	0.00

[1]one-tailed

Table B13: Pairwise comparisons of the CVDR contrast between conversational whispered and clear whispered speech.

Source	SS	df	MS	F	P
Between-Subjects	2.069	15			
Speaker	0.926	3	0.309	3.24	0.0604
Error(Speaker)	1.143	12	0.095		
Within-Subjects	2.998	112			
Style	0.045	1	0.045	1.876	0.1958
Speaker x Style	0.464	3	0.155	6.508	0.0073
Error(Speaker x Style)	0.285	12	0.024		
ConsonantPair	0.518	3	0.173	7.583	0.0005
Speaker x ConsonantPair	0.283	9	0.031	1.383	0.2321
Error(Speaker x ConsonantPair)	0.82	36	0.023		
Style x ConsonantPair	0.033	3	0.011	0.96	0.4222
Speaker x Style x ConsonantPair	0.141	9	0.016	1.386	0.2307
Error(Style-ConsonantPair)	0.408	36	0.011		
Total	5.067	127			

Table B14: ANOVA table for the effects of Speaker, Consonant-pair and Speaking-style on the C/V duration-ratio contrast in whispered speech

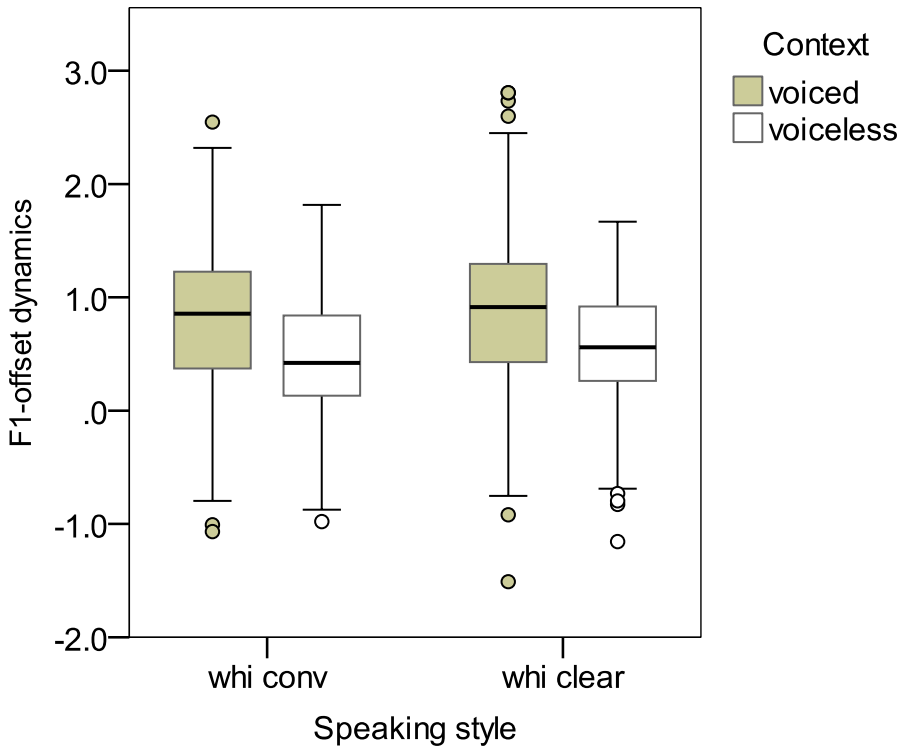


Figure B1: *F1*-offset dynamics in whispered conversational and clear speech. In ANOVA on *F1*-offset dynamics with Speaker, Voicedness, and Speaking-style, there was no significant effect of Speaking-style indicating that clear speaking-style did not influence *F1*-offset dynamics in whispered speech.

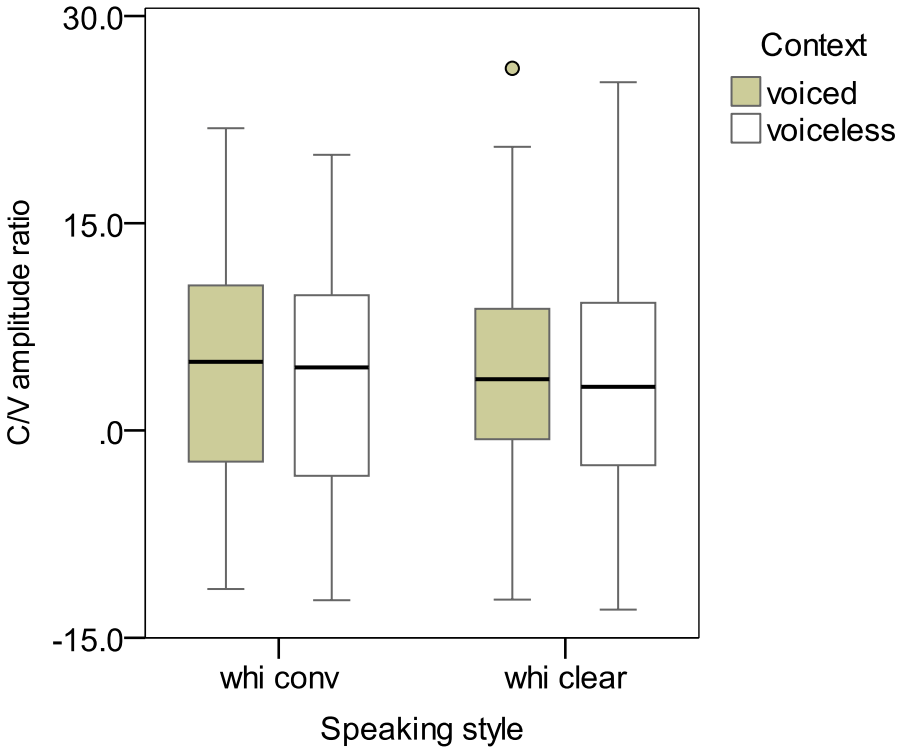


Figure B2: In ANOVA on C/V amplitude-ratio, the effect of Speaking-style was not significant indicating that clear speaking-style did not influence C/V amplitude-ratio in whispered speech

Appendix C: Perception of consonant "voicing" in phonated and whispered speech

Speaker 1								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	50.56	49.17		0.28				
p	14.44	85.56						
g			84.17	15.83				
k			0.83	99.17				
z					79.44	20.00	0.28	0.28
s					4.17	95.83		
ʤ							92.78	7.22
ʧ							10.28	89.72

Speaker 2								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	84.72	15.28						
p	5.00	94.72				0.28		
g	0.56		93.33	6.11				
k			6.94	93.06				
z					91.39	8.33	0.28	
s					6.94	93.06		
ʤ							88.33	11.67
ʧ							33.06	66.94

Speaker 3								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	68.61	30.83	0.28	0.28				
p	7.50	92.22		0.28				
g	0.56		89.17	10.28				
k			4.17	95.83				
z					80.83	19.17		
s					17.50	82.50		
ʤ							96.11	3.61
ʧ					0.28		29.17	70.56

Speaker 4								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	91.67	8.33						
p	26.39	73.06		0.28		0.28		
g		0.28	70.00	29.17			0.56	
k			6.67	93.33				
z					80.83	18.89	0.28	
s					16.39	83.61		
ʤ			0.56				90.28	9.17
ʧ	0.28						10.28	89.44

Table C1: Error patterns in conversational whispered speech. Scores are percentages of correct responses

Speaker 1								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	81.39	18.33				0.28		
p	1.94	97.78		0.28				
g			96.11	3.89				
k			0.56	99.44				
z					81.94	17.78	0.28	
s					1.11	98.89		
ʤ			0.56		0.28		88.61	10.56
ʧ	0.28						4.72	95.00

Speaker 2								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	95.83	2.78	1.39					
p	2.78	96.11		0.56		0.28		0.28
g	0.28		98.06	1.39			0.28	
k			1.67	98.33				
z	0.28				85.28	14.44		
s					5.56	94.44		
ʤ						0.28	94.72	5.00
ʧ							9.44	90.56

Speaker 3								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	92.50	7.22		0.28				
p	16.11	83.89						
g	0.56		93.89	5.56				
k			25.56	74.44				
z					71.94	27.78	0.28	
s			0.28		14.17	85.56		
ʤ			0.56				93.89	5.56
ʧ			0.28				35.56	64.17

Speaker 4								
	response selected							
stimulus played	b	p	g	k	z	s	ʤ	ʧ
b	91.94	7.78	0.28					
p	27.22	71.67	0.83	0.28				
g			78.06	21.67	0.28			
k		0.28	6.94	92.78				
z					80.83	18.89	0.28	
s					31.94	68.06		
ʤ			0.28				91.39	8.33
ʧ	0.28						30.28	69.44

Table C2: Error patterns in clear whispered speech. Scores are percentages of correct responses

Dependent Variable: Accuracy in A'-scores

Source	Sum of Squares	df	Mean Square	F	Sig.
Between-Subjects	0.737	14			
Within-Subjects	1.894	465			
condition	0.022	1	0.022	17.19	0.001
Error(condition)	0.018	14	0.001		
speaker	0.222	3	0.074	35.675	0.000
Error(speaker)	0.087	42	0.002		
consPair	0.155	3	0.052	8.417	0.000
Error(consPair)	0.258	42	0.006		
condition * speaker	0.128	3	0.043	51.444	0.000
Error(condition * speaker)	0.035	42	0.001		
condition * consPair	0.105	3	0.035	19.768	0.000
Error(condition * consPair)	0.075	42	0.002		
speaker * consPair	0.241	9	0.027	18.649	0.000
Error(speaker-consPair)	0.181	126	0.001		
condition * speaker * consPair	0.134	9	0.015	14.699	0.000
Error(condition-speaker-consPair)	0.127	126	0.001		
Total	2.526	479			

Table C3: ANOVA table for the effects of Speaker, Consonant-pair, and Speaking-style on consonant “voicing” identification accuracy in whispered speech

Bibliography

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech-i: global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun.*, *20*(3-4), 255–272.

Dannenbring, G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Percept Mot Skills*, *51*(3 Pt 1), 979–85.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. *J Acoust Soc Am*, *116*(4 Pt 1), 2365–73.

Hazan, V., & Simpson, A. (1998). The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise1. *Speech Communication*, *24*(3), 211–226.

Hazan, V., & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Lang Speech*, *43*(Pt 3), 273–94.

Higashikawa, M., Green, J. R., Moore, C. A., & Minifie, F. D. (2003). Lip kinematics for /p/ and /b/ production during whispered and voiced speech. *Folia Phoniatr Logop*, *55*(1), 17–27.

Hirahara, T., Otani, M., Shimizu, S., Toda, T., Nakamura, K., Nakajima, Y., & Shikano, K. (2009). Silent-speech enhancement using body-conducted vocal-tract resonance signals. *Speech Communication*.

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, *45*(2), 139–152.

Jiang, J., Chen, M., & Alwan, A. (2006). On the perception of voicing in syllable-initial plosives in noise. *J Acoust Soc Am*, *119*(2), 1092–1105.

Jovicic, S. (1998). Formant feature differences between whispered and voiced sustained vowels. *Acta Acustica united with Acustica*, *84*, 739–743.

Jovicic, S., & Saric, Z. (2006). Acoustic analysis of consonants in whispered speech. *J Voice*, *22*(3), 263–74.

Kallail, K. J., & Emanuel, F. W. (1984a). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, *12*(2), 175–186.

Kallail, K. J., & Emanuel, F. W. (1984b). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *J Speech Hear Res*, *27*(2), 245–51.

Kallail, K. J., & Emanuel, F. W. (1985). The identifiability of isolated whispered and phonated vowel samples. *Journal of Phonetics*, *13*(1), 11–17.

Kent, R., & Read, C. (2002). *The acoustic analysis of speech*. Singular/Thomson Learning, 2 ed.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *J Speech Hear Res*, *18*(4), 686–706.

Kostek, B. (2005). *Perception-based data processing in acoustics: applications to music information retrieval and psychophysiology of hearing*. Springer, 1st edition ed.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am*, *115*(1), 362–78.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69–72.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Li, X. L., & Xu, B. L. (2005). Formant comparison between whispered and voiced vowels in mandarin. *Acta Acustica united with Acustica*, *91*(6), 1079–1085.

Loizou, P. (1999). Colea: A matlab software tool for speech analysis. URL `http://www.utdallas.edu/~loizou/speech/colea.htm`

Maniwa, K., Jongman, A., & Wade, T. (2008). Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners. *Journal of the Acoustical Society of America*, *123*(2), 1114–1125.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken english fricatives. *Journal of the Acoustical Society of America*, *125*(6), 3962–3973.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *J Acoust Soc Am*, *85*(5), 2114–34.

Morris, & Clements (2002). Reconstruction of speech from whispers. *Medical engineering & physics*, *24*(7), 515–520.

Munro, M. (1990). Perception of "voicing" in whispered stops. *Phonetica*, *47*(3-4), 173–181.

Parnell, M., Amerman, J. D., & Wells, G. B. (1977). Closure and constriction duration for alveolar consonants during voiced and whispered speaking conditions. *J Acoust Soc Am*, *61*(2), 612–3.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing i: Intelligibility differences between clear and conversational speech. *J Speech Hear Res*, *28*(1), 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. ii: Acoustic characteristics of clear and conversational speech. *J Speech Hear Res*, *29*(4), 434–46.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in english. *Percept Psychophys*, *32*(2), 141–52.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in american english. *J Acoust Soc Am*, *51*(4), 1296–303.

Repp, B. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*(2), 173.

Revoile, S., Pickett, J. M., Holden, L. D., & Talkin, D. (1982). Acoustic cues to final stop voicing for impaired-and normal hearing listeners. *J Acoust Soc Am*, *72*(4), 1145–54.

Schwartz, M. F. (1972). Bilabial closure durations for p, b, and m in voiced and whispered vowel environments. *J Acoust Soc Am*, *51*(6), 2025–9.

Smiljanic, R., & Bradlow, A. (2005). Production and perception of clear speech in croatian and english. *The Journal of the Acoustical Society of America*, *118*, 1677.

Smiljanic, R., & Bradlow, A. (2008a). Stability of temporal contrasts across speaking styles in english and croatian. *Journal of Phonetics*, *36*(1), 91–113.

Smiljanic, R., & Bradlow, A. (2008b). Temporal organization of english clear and conversational speech. *Journal of the Acoustical Society of America*, *124*(5), 3171–3182.

Smiljanic, R., & Bradlow, A. R. (2008c). Stability of temporal contrasts across speaking styles in english and croatian. *J Phon*, *36*(1), 91–113.

Stent, A., Huffman, M., & Brennan, S. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, *50*(3), 163–178.

Stevens, K., & Klatt, D. (1974). Role of formant transitions in the voiced voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*, 653.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of american english vowels. *J Acoust Soc Am*, *79*(4), 1086–1100.

Tagliaferri, B. (2005). Perception research systems.

URL http://www.paradigmexperiments.com

Tartter, V. C. (1989). What's in a whisper? *J Acoust Soc Am*, *86*(5), 1678–83.

Tartter, V. C. (1991). Identifiability of vowels and speakers from whispered syllables. *Percept Psychophys*, *49*(4), 365–72.

Tran, V. A., Bailly, G., Loevenbruck, H., & Toda, T. (2010). Improvement to a nam-captured whisper-to-speech system. *Speech Communication*, *52*(4), 314–326.

Tsunoda, K., Niimi, S., & Hirose, H. (1994). The roles of the posterior cricoarytenoid and thyropharyngeus muscles in whispered speech. *Folia Phoniatr Logop*, *46*(3), 139–51.

Weismer, G., & Longstreth, D. (1980). Segmental gestures at the laryngeal level in whispered speech: evidence from an aerodynamic study. *J Speech Hear Res*, *23*(2), 383–92.