How do you whisper a click?

Acoustic correlates of click voicing in whispered speech

Aaron Braver
Texas Tech University

Slides online at aaronbraver.com/lsa2017

LSA 2017

Introduction

- Whispering has many uses (Cirillo 2004)
- Whispering is quiet
- Clicks are loud (Ladefoged and Traill 1994)

- What happens when you try to whisper a click?

Introduction

- Voicing and whisper require potentially contradictory glottal states

- (How) is the voicing contrast maintained in whispered speech?

Introduction

- Numerous studies on whispered speech and maintenance of voicing contrast in pulmonic sounds

- How do you whisper non-pulmonic sounds?

- (How) do you maintain a voicing contrast in whispered non-pulmonic sounds?

Terminological notes

- Whispered vs. not whispered:
  - "phonation" contrast

- Whispered:
  - "whispered", "whisper phonation"

- Not whispered:
  - "phonated", "normal phonation", "non-whispered"

Outline

- Background
  - Cues to whisper & breathiness
  - Cues to voicing
  - Xhosa (Bantu) and its clicks
- Methods
- Results
  - Whispered vs. phonated clicks
  - Voiced vs. voiceless phonated clicks
  - Voiced vs. voiceless whispered clicks

Background

Cues to whisper

- Intensity (Jovičić and Šarić 2008)
- Periodicity (Monoson and Zemlin 1984)
  - Jitter
  - Harmonic to Noise Ratio (HNR)
- Spectral tilt (Ladefoged and Antonanzas-Barroso 1985)
  - H1-H2
  - H1-A1, H1-A2, H1-A3
- Duration (longer)
  - Overall (Jovičić and Šarić 2008)
  - Closure duration (Osfar 2011)

Cues to voicing

- Voicing in closure
- Closure duration (Kluender et al. 1988)
- Preceding vowel duration (Chen 1970)
- Effects on F0 and F1 of surrounding vowels (Hombert et al. 1979, Kingston and Diehl 1994)
- *Inter alia* (see Lisker 1986)

Cues to voicing in whisper

- C duration (Mills 2003)
- Frication duration (Tartter 1989)
- Aspiration duration (Tartter 1989)
- Preceding V duration (Mills 2003)
- Lip movement velocity in labials (Higashikawa et al. 2003)
- Intensity Δ between [±voice], relative to Δ in normal phonation (Jovičić and Šarić 2008)

Xhosa (Bantu)

- "isiXhosa" [isi‖ʰɔ́sà]
- Spoken mainly in South Africa's Eastern Cape
- 8 million+ speakers (Lewis et al. 2016)

- 15 contrastive clicks

Xhosa clicks

	Voiceless unasp.		Voiceless asp.		Voiced		Nasalized		Nasalized breathy	
Dental	c	[k|]	ch	[k|ʰ]	gc	[g|]	nc	[n|]	ngc	[ng|]
Alveolar	x	[k!]	xh	[k!ʰ]	gx	[g!]	nx	[n!]	ngx	[ng!]
Lateral	q	[k‖]	qh	[k‖ʰ]	gq	[g‖]	nq	[n‖]	ngq	[ng‖]

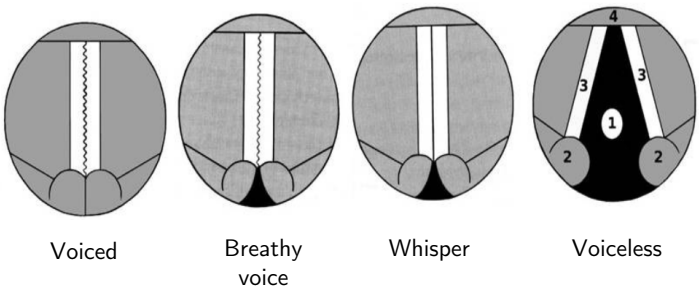
"Voiced" clicks

- Great variability in transcription of voiced and breathy nasal clicks
- Voiced: [g!], [g̬!], [g̊!], [g!ⁿ]

"Voiced" clicks

- Notational variation reflects variation in description:
- "Voiced" (Traditional description, textbooks, Roux 2007)
- "Oral and voiced" (Jordan 1966)
- "Breathy voiced" (Sands 1991, Ladefoged and Maddieson 1996)
- "Slack voice" (Jessen and Roux 2002)
- "Murmured" (Ladefoged and Traill 1994, Ladefoged and Johnson 2010)
- "Depressor consonants" (Sands 1991)

Glottal configuration



Voiced | Breathy voice | Whisper | Voiceless

Diagram modified from Drugman et al. 2014

1: Glottis
2: Arytenoid cartilages
3: Vocal folds
4:Epiglottis

"Voiced" clicks

- No laryngograph/(nas)endoscopy studies of Xhosa click accompaniments
- Claims of no vocal fold vibration come from acoustic studies
  - VOT of 0 (Böhm 2010)
  - 36–56 ms of "breathy voice" following (Böhm 2010)
- Large H1-H2, H1-A3 (Jessen and Roux 2002)
- Depress F0 of following V (Jessen and Roux 2002)

Method

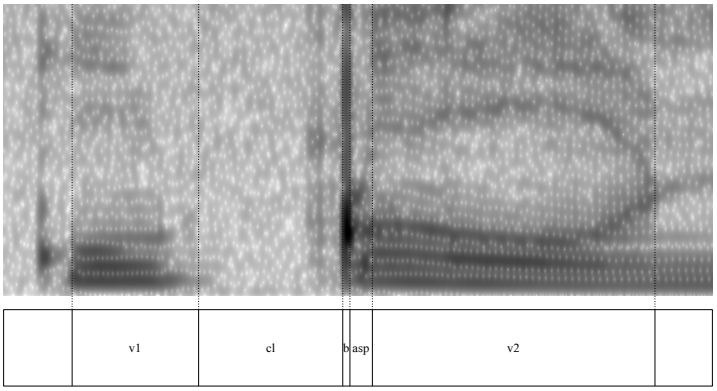
Speaker and stimuli

- Female native speaker of Xhosa from South Africa's Eastern Cape
- 15 real Xhosa words matching the 15 Xhosa click types. All in ...uCo... context
- 2 blocks — one normal, one whispered
- 2 repetitions per word, order randomized

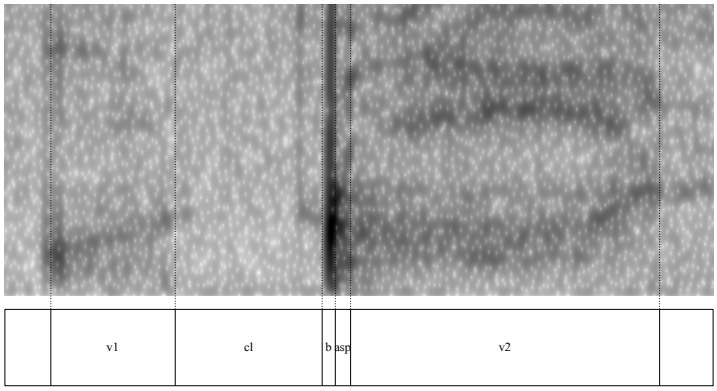
Sample stimuli

Item	Symbol	Description	Gloss
ukuggoba	ɡǃ	Voiced alveolar	to scoop
ukungqoza	ŋgǃ	Breathy nasal alveolar	to admonish
ukunqoma	ŋǃ	Nasal alveolar	to climb up
ukughola	kǃʰ	Voiceless aspirated alveolar	to perfume
ukuqonda	kǃ	Voiceless unaspirated alveolar	understand

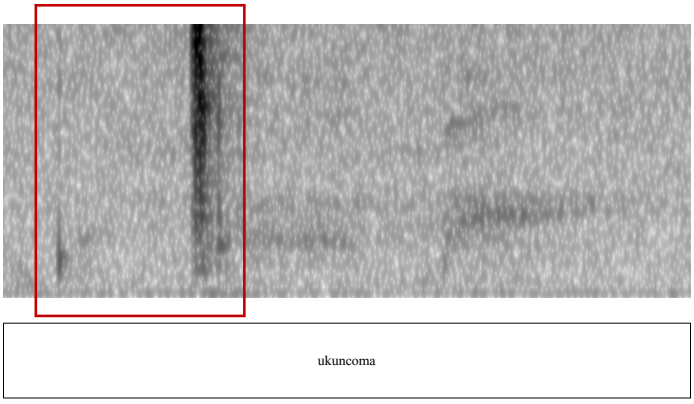
ukuqonda (phonated)



ukuqonda (whispered)



ukuncoma (whispered)



Acoustic variables:

Whisper and breathiness

- Burst amplitude
- Preceding vowel midpoint amplitude
- Spectral tilt
  - H1-H2
  - H1-A3
- Periodicity
  - Harmonic to Noise Ratio (HNR)

Acoustic variables:

Voicing

- Preceding vowel duration
- Closure duration
- Aspiration/breathiness duration
- F0/F1 in following vowel

Statistics

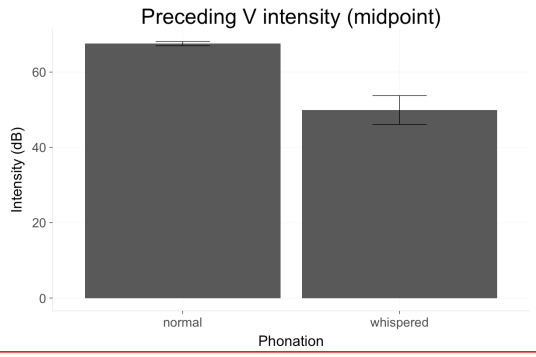
- Per variable:
  - LMM
  - Fixed factors:
    - phonation + voicing
  - Random factor: item

Results

Whisper vs. normal

Intensity

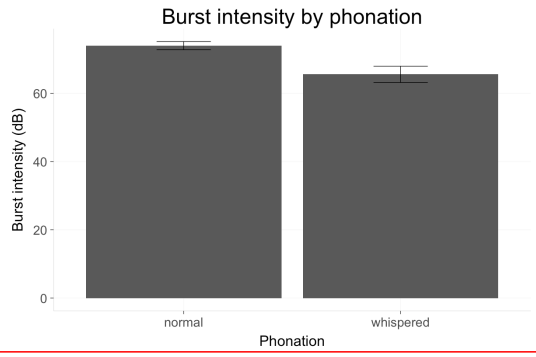
- Whisper vs. normal phonation



Mean diff: 17.63
t(17)=-4.85)
p = 0.0001

Intensity

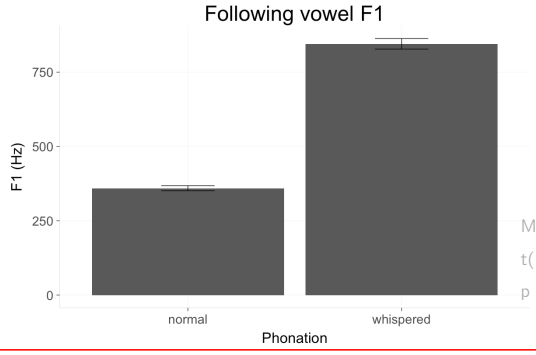
- Whisper vs. normal phonation



Mean dif: 8.417
t(15.2)=-4.352
p = 0.0005

Following vowel F1

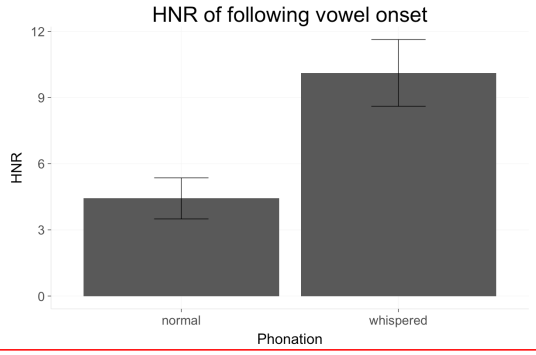
- Whisper vs. normal phonation



Mean dif: 486.04ms
t(17)=-26.31
p < 0.0001

Harmonic to Noise Ratio (HNR)

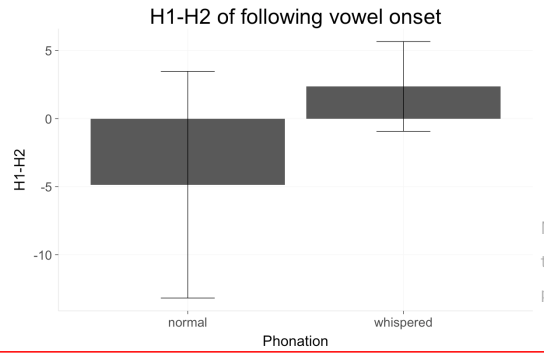
- Whisper vs. normal phonation



Mean dif: 5.691
t(16.2)=3.196
p = 0.006

Spectral tilt (H1-H2)

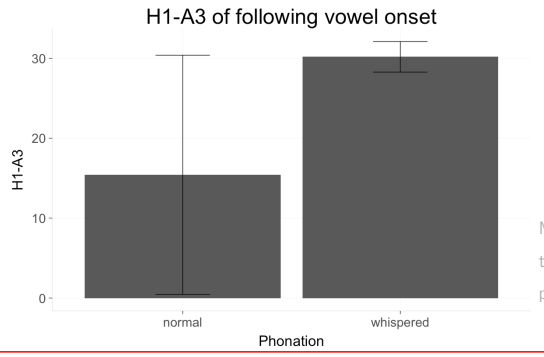
- Whisper vs. normal phonation



Mean dif: 7.22
t(6.2)=0.938
p = 0.388

Spectral tilt (H1-A3)

- Whisper vs. normal phonation



Mean dif: 14.77
t(8.00)=2.428
p = 0.0413

Summary: phonation

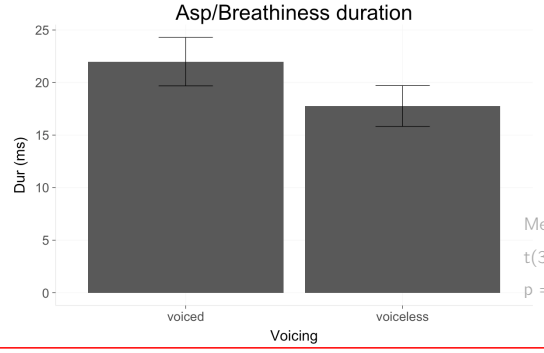
Preceding V intensity	✓
Burst intensity	✓
Following V F1	✓
Following V HNR	(✓)
Following V H1-H2	X
Following V H1-A3	~

Results

Click voicing in normal speech

Aspiration/Breathiness duration

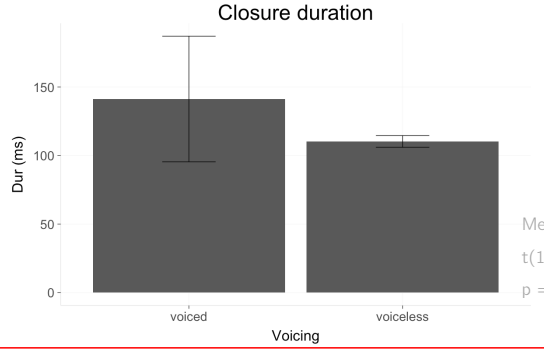
- Voicing contrast in normal speech



Mean dif: 4.224ms
t(3.22)=-0.015
p = 0.990

Closure duration

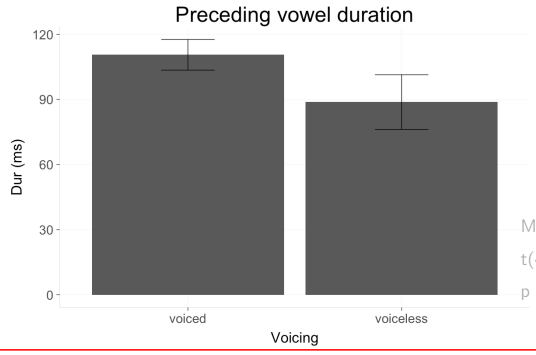
- Voicing contrast in normal speech



Mean dif: 30.94ms
t(15)=-0.507
p = 0.620

Preceding vowel duration

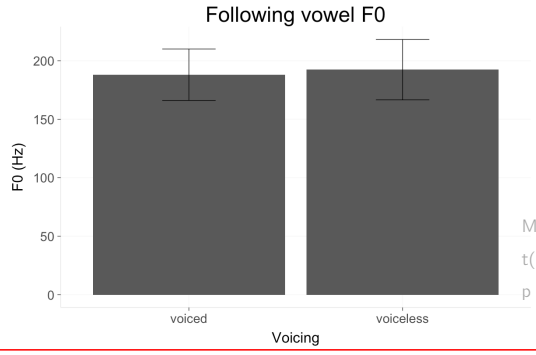
- Voicing contrast in normal speech



Mean dif: 21.78ms
t(4)=-0.791
p = 0.473

Following vowel F0

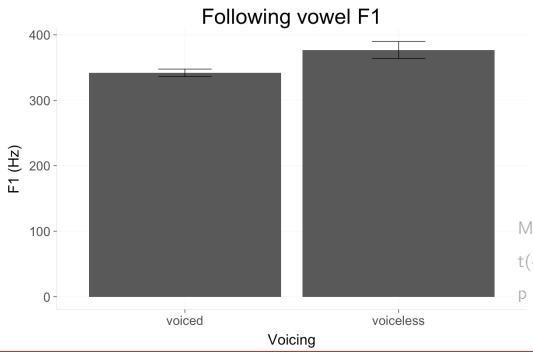
- Voicing contrast in normal speech



Mean dif: 4.33Hz
t(3.5)=-0.452
p = 0.678

Following vowel F1

- Voicing contrast in normal speech



Mean dif: 34.75Hz
t(4)=-0.95
p = 0.396

Normal speech voicing summary

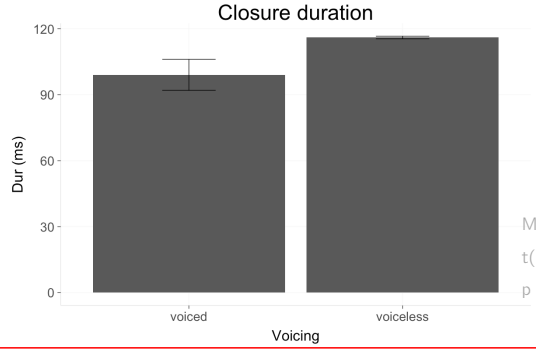
Aspiration/Breathiness Dur	~
Closure duration	X
Preceding V dur	X
Following V F0	X
Following V F1	X

Results

Click voicing in whisper

Closure duration

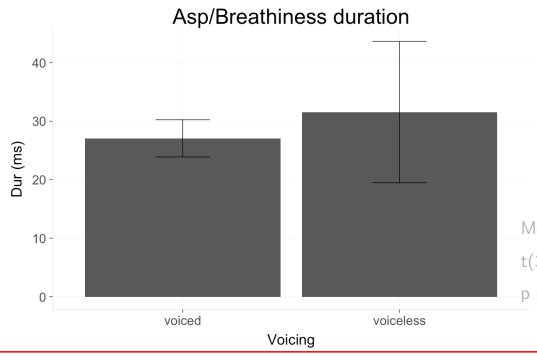
- Voicing in whispered speech



Mean dif: 16.98ms
t(15)=-0.507
p = 0.620

Aspiration/Breathiness duration

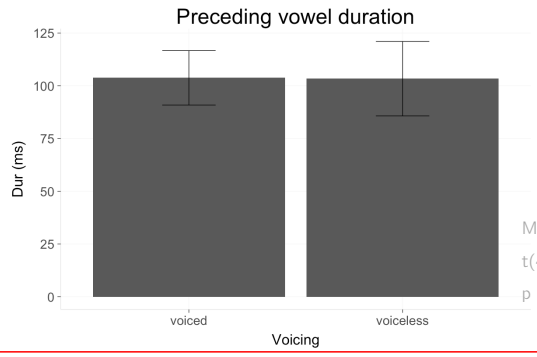
- Voicing in whispered speech



Mean dif: 4.49ms
t(3.22)=0.015
p = 0.989

Preceding vowel duration

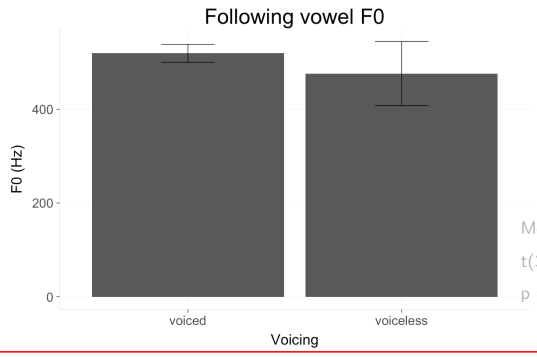
- Voicing in whispered speech



Mean dif: 0.40ms
t(4)=-0.791
p = 0.473

Following vowel F0

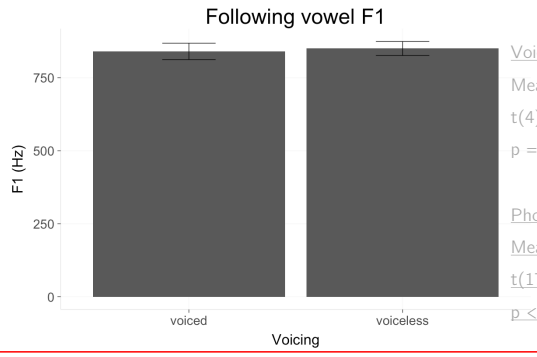
- Voicing in whispered speech



Mean dif: 43.06Hz
t(3.46)=-0.452
p = 0.678

Following vowel F1

- Voicing in whispered speech



Voicing
Mean dif: 9.96
t(4)=0.95
p = 0.396

Phonation
Mean dif: 486.04ms
t(17)=26.31
p < 0.0001

Summary: whispered voicing

Closure duration	~
Aspiration/breathiness dur	~
Preceding V duration	X
Following V F0	X
Following V F1	X

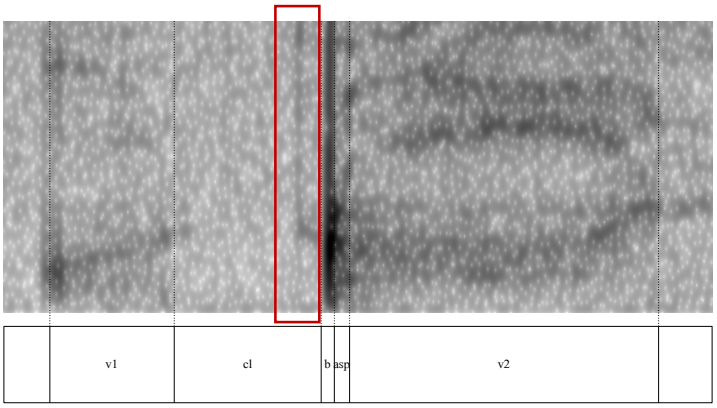
Discussion

So, how do you whisper a click?

- Whisper cued mainly by intensity
  - ~60 dB click burst is not particularly quiet
- HNR and F1 effects

Preceding V intensity	✓
Burst intensity	✓
Following V HNR	✓
Following V H1-H2	X
Following V H1-A3	X
Following V F1	✓

ukuqonda (whispered)



How do you voice a click?

- Main cue is aspiration/breathiness
- No "depressor" effects

Closure duration	X
Preceding V dur	X
Following V F0	X
Following V F1	X
Aspiration/Breathiness Dur	~

How do you voice a whispered click?

- Potentially cued by closure duration and aspiration/breathiness

Closure duration	~
Preceding V duration	X
Following V F0	X
Following V F1	X
Aspiration/breathiness dur	~

Why study whisper?

- Everybody does it
- Perturbations to speech signal provide a natural filter (Mills 2009)
  - Cues that remain are often sufficient for contrast
  - "Filter out" vocal fold vibration
- Speech recognition systems often need to filter out whisper ("whisper island detection") (Zhang and Hansen 2001)

Thank you

- Thanks to Will Bennett and Jeremy Perkins for help with making the recordings, and to Seunghun Lee for helpful discussion

Slides available at aaronbraver.com/lsa2017

References

Böhm, S. (2010). Clicks in Xhosa and Nama: A comparative analysis. Master's thesis, Southern Illinois University at Carbondale.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environ- ment. Phonetica, 22(129-159).

Cirillo, J. (2004). Communication by unvoiced speech: the role of whispering. Anais da Academia Brasileira de Ciências, 76(2):413–423.

Drugman, T., Alku, P., Alwan, A., and Yegnanarayana, B. (2014). Glottal source processing: From analysis to applications. Computer Speech & Language, 28(5):1117–1138.

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

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. Language, 55(1):37–58.

References

Jordan, A. C. (1966). A practical course in Xhosa. Longmans.

Jovičič, S. T. and Šarič, Z. (2008). Acoustic analysis of consonants in whispered speech. Journal of voice, 22(3):263–274.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. Language, 3:419–454.

Kluender, K. R., Diehl, R. L., and Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. Journal of Phonetics, 16:153–169.

Ladefoged, P. and Antoñanzas-Barroso, N. (1985). Computer measures of breathy phonation. UCLA Working Papers in Phonetics, 61:79–86.

Ladefoged, P. and Maddieson, I. (1996). The Sounds of the World's Languages. Blackwell Publishers, Oxford.

Ladefoged, P. and Johnson, K. (2011). A course in phonetics (sixth edition).

Ladefoged, P. and Traill, A. (1994). Clicks and their accompaniments. Journal of Phonetics, 22:33–64.

References

Lewis, M. P., Simons, G. F., and Fennig, C. D., editors (2016). Ethnologue: Languages of the World. SIL International, http://www.ethnologue.com, 19th edition.

Lisker, L. (1986). "Voicing" in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech, 29:3–11.

Mills, T. I. P. (2003). Cues to voicing contrasts in whispered scottish obstruents. whisper, phonation, acoustic cues, Edinburgh University.

Mills, T. I. P. (2009). Speech motor control variables in the production of voicing contrasts and emphatic accent. PhD thesis, The University of Edinburgh.

Monoson, P. and Zemlin, W. R. (1984). Quantitative study of whisper. Folia Phoniatrica et Logopaedica, 36(2):53–65.

Osfar, M. J. (2011). Articulation of whispered alveolar consonants. PhD thesis, University of Illinois at Urbana-Champaign.

References

Roux, J. (2007). Unresolved issues in the representation and phonetic description of click articulation in Xhosa and Zulu. Language Matters, 38(1):8–25.

Sands, B. E. (1991). Evidence for click features: Acoustic characteristics of xhosa clicks. UCLA Working Papers in Phonetics, 80:6–37.

Tartter, V. C. (1989). What's in a whisper? Journal of the Acoustical Society of America, 86(5):1678–1683.

Zhang, C. and Hansen, J. H. (2011). Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing. IEEE Transactions on Audio, Speech, and Language Processing, 19(4):883–894.