Journal of Phonetics 51 (2015) 1–5

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

High is not just the opposite of Low ☆

Jonathan Paul Evans

Academia Sinica, Taiwan

A R T I C L E I N F O

Article history:
Received 29 April 2015
Received in revised form
9 May 2015
Accepted 14 May 2015
Available online 17 June 2015

Keywords:
Fundamental frequency
High
Pitch
Tone
Stress
Larynx
Music

A B S T R A C T

Elevated fundamental frequency (*F*0) has been found to have similar properties across languages. For example, raised pitch often accompanies syllabic stress, emotionally charged speech, infant-directed speech, and questions. In many languages, occurrence of high tone is subject to more constraints than are other tones. Given that these patterns occur frequently in the world's languages, it is natural to ask whether language-independent properties of raised *F*0 could play a role in the existence of typological similarities. The purpose of the present paper is to survey possible language-external factors that appear to play a role in the special status of linguistic H(igh). Moreover, the collection of studies assembled in this Special Issue provides empirical evidence that raised *F*0 attracts listener attention differently from lowered *F*0, that sustained production of high *F*0 may involve unique auditory control mechanisms, and that social context and even semantics may contribute to speaker production of raised *F*0. It is hoped that the articles of the special issue will provide a phonetic basis to explain some of the asymmetries observable in prosodic systems of languages around the world.

1. Introduction

On a hot summer's day, director Mel Larimer began to prepare the Interlochen National Music Camp high school choir to learn the choral portion of Beethoven's Ninth Symphony. He asked the pianist to play high A (880 Hz), the recurring highest note sung more than seventy times throughout the piece. He then asked those first sopranos who were able, to sing the note, which audibly strained their voices. Lastly, he asked the choir, “Which sounded more impressive, the piano or the human voice?” That moment in 1984 planted the seed that has become this article and this Special Issue.

This article sets out to introduce findings that show that raising human voice fundamental frequency is not the mirror image of lowering it. That is, evidence from production and perception suggest that there are physical and psychological bases for the widespread linguistic asymmetries between H(igh) and L(ow).

Because production and perception issues are discussed below and in the articles in this Special Issue, I briefly mention here some common phonological processes that demonstrate a privileged position for elevated pitch. In many languages with H and L tones, there is an active constraint against adjacent H tones (“Meussen's rule” Goldsmith, 1984). Even more restrictive, some phonological grammars do not permit more than one H tone per word (“culminativity”; Evans, 2008). Tonal systems with similar restrictions against L are much less common (Hyman, 2001). In addition, in many cases, raised *F*0 on an individual syllable corresponds with stress (Crystal, 2011).

It is not clear whether acoustic properties of sound play a direct role in the special characteristics of linguistic H. Sound intensity is a property of amplitude, rather than frequency. Thus, a sound with higher *F*0 does not contain more energy than one with lower *F*0. Nevertheless, studies of *F*0 and loudness, the perceptual correlate of intensity, have shown that different fundamental frequencies are perceived at different loudness, even with sound pressure level kept constant (Fletcher & Munson, 1933; Robinson & Dadson, 1956). However, these studies tested responses to pure sinusoidal tones, not complex voice-like tones; they also did not probe distinctions within the normal human spoken *F*0 range. Thus, it is still unknown what shape a vocal pitch/loudness curve would take.

☆ For their comments on earlier versions of this article, I would like to thank Ken DeJong, Yu-Wen Lai and Wen-Chi Yeh. Others are acknowledged within the article for their domain-specific assistance. Funding for this research was provided by Ministry of Science and Technology (Taiwan, R.O.C.) Grant #100-2628-H-001-008-MY4.

E-mail address: jonathan@sinica.edu.tw

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.05.001

Fundamental frequency is conveyed multi-modally in the sound signal. It is the lowest common denominator of the component frequencies, because the harmonic frequencies are multiples of the vibration rate of the vocal folds. Thus, *F*0 is also the distance between each pair of neighboring harmonics. As mentioned in the discussion on perception (Section 3), this multi-modality (wave frequency and distance between harmonics) allows multiple perception mechanisms to identify fundamental frequency.

As the sound wave travels, intensity degrades rapidly, while frequency does not. The idealized speech signal emanates from the speaker as a sphere whose surface area equals $4\pi r^2$. Thus, the sound pressure level on the surface decreases relative to the square of the distance from the source. On the other hand, wave frequency information does not degrade across distance as quickly as intensity does. As a sound wave travels through air, the wave moves air molecules. With each push of a molecule, some wave energy is lost in the form of heat. Lower frequency waves, such as *F*0, lose less energy to heat than do higher frequency waves, such as spectral components (Yun-Feng Hsieh, p.c.). Thus, the intensity of the greatest common denominator frequency decreases more slowly than does that of formants. Due to the physical properties of the sound signal, and the multi-modality of fundamental frequency, *F*0 degrades more slowly across distance than formants and intensity do.

Because *F*0 is more robust across distance than other components of the sound wave are, it is available to speakers and listeners as a relatively invariant signal carrier. Given the relationship between increased intensity and increased *F*0 (Section 2), raised *F*0 could serve as a more faithful indicator of raised intensity than the actual sound pressure level. One example of a correlation between speech intensity and raised *F*0 is that of expression of intense emotions. Feelings such as happiness, anger, fear, and even impatience tend to be expressed by sustained raised *F*0 (Michaud, Vaissière, & Nguyễn, 2015; Pell, Monetta, Paulmann, & Kotz, 2009; Schröder, 2001; Williams & Stevens, 1972). To the extent that this raised *F*0 correlates with higher subglottal pressure, it is a robust indicator of forceful speech.

The following two sections introduce findings related to production and perception of *F*0; they also highlight phenomena that suggest a special role for *F*0 raising.

2. Production and raising of *F*0

Acoustically defined, fundamental frequency (*F*0) is the “lowest frequency component in a complex sound wave” (Crystal, 2011). From a physiological perspective, *F*0 is the rate of vibration of the vocal folds (Gick, Wilson, & Derrick, 2013:86; Ryalls & Behrens, 2000:20). When speakers laryngeally raise the vibration rate, the cricothyroid muscle contracts, which tilts the thyroid cartilage forward, elongating and thinning the vocal folds. Simultaneously, the thyroarytenoid muscles contract, stiffening the vocal folds. Both actions serve to raise the frequency of vibration. In lowering *F*0, parts of the thyroarytenoid muscles contract, shortening the vocal folds. The concomitant increase of mass per unit length slows vocal fold vibration (Gick et al., 2013:86–89; Hirose, 1997:116–136; Reetz & Jongman, 2009:69–71).

Speakers monitor and adjust *F*0. Control of laryngeal structures during speech involves different neural pathways than are invoked during less volitional activities such as cough, swallow, and sniff. (Ludlow, 2005). During swallowing, coughing, etc., the muscles activated are consistent across instances and speakers. However, speakers vary between and within themselves in the mixture of subglottal pressure, airflow, and cricothyroid and thyroarytenoid muscle activation employed to yield a particular combination of intensity and *F*0 (Atkinson, 1976). Speakers rely on both somatosensory feedback and auditory feedback in order to monitor and adjust *F*0 production. Laryngeal muscles move with speed and precision during utterances, which suggests that throughout the language acquisition process, somatosensory feedback aids the speaker in producing the laryngeal gestures that yield the desired vocal output. Recent experiments suggest that mechanoreceptors in the laryngeal mucosa provide the central nervous system with feedback when the larynx is in motion (Ludlow, 2005).

Speakers monitor their own *F*0 auditorily, so that when presented with an *F*0-shifted version of their ongoing speech, they produce a compensatory shift of *F*0 in the opposite direction (Larson, White, Freedland, & Burnett, 1996; Sturgeon, Hubbard, Schmidt, & Loucks, 2015; Ning, Loucks, & Shih, 2015). Differences in compensatory shift have been noted among trained vocalists, speakers of tonal languages, non-tonal language speakers, and L2 speakers of a tonal language (Ning et al., 2015). The existence of these differences suggests that compensatory *F*0 shift is not merely a reflex. Musicians who are not vocalists differ from non-musically trained speakers in their pitch shift responses, suggesting that pitch control experience of a non-laryngeal nature affects the vocal *F*0 control mechanism (Sturgeon et al., 2015).

In addition to raising *F*0 via laryngeal settings, speech uttered with more force also has higher *F*0. For example, *F*0 increases when the rate of airflow across the vocal folds increases and all laryngeal settings are held constant, as confirmed by studies both on humans and on excised canine larynges (Alipour & Scherer, 2007; Baer, 1979; Lieberman, Knudson, & Mead, 1969; Titze, 1989). Higher subglottal pressure leading to greater airflow can occur in various contexts, both linguistic and environmental. *F*0 raising has been noted in at least four contexts in which higher subglottal pressure or more forceful speech could be a cause.

First, higher subglottal pressure could be related to the sudden rises in *F*0 noted at the beginning of new discourse sections (Menn & Boyce, 1982; Pierrehumbert & Hirschberg, 1990 (chap. 14); Mohler & Mayer, 2001; Tseng, Pin, Lee, Wang, & Chen, 2005; Tseng, 2008; Xu, 2006). Second, in the presence of noise, speakers produce speech with both greater intensity and higher *F*0; that is, the “Lombard effect” (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Third, in addition to noise, distance from listener is also a condition in which more speech effort is used. Shih and Lu (2015) find that as the distance between talker and listener increases, there are concomitant increases in intensity, duration, *F*0 maximum and *F*0 range. An increase in *F*0 range indicates that higher sounds are raised more than lower sounds are. Thus, *F*0 raising in this context might be more than just the effect of increased airflow due to higher intensity. Finally, in at least some cases, higher *F*0 on an individual syllable corresponds with stress (Crystal, 2011).

*F*0 raising as an effect of speech forcefulness allows for the possibility of iconic use of *F*0 raising. Jun (2015) suggests that an iconic use of H may occur in Seoul Korean, where [il] has a higher *F*0 realization when the meaning is 'one' than it does when other meanings are encoded. The implication is that the cultural importance of first-ness may have seeped into the articulation of this morpheme.

3. Perception of *F*0 and of *F*0 raising

Pitch perception begins in the auditory periphery, the structures involved in hearing a sound before its neural signal is created (outer ear, cochlea, etc.). The auditory periphery decomposes spectral information, such as the fundamental frequency and the harmonics. The cochlea is more sensitive to the first ten harmonics, with the first five harmonics dominating that region (Plomp, 1967; Ritsma, 1962). Hair cells of the cochlea convert the spectral information into nerve impulses. Detection of pitch is sensitive to information from both fundamental frequency and harmonics. If the fundamental frequency is removed from an acoustic signal, subjects perceive the same pitch. This perception may be due to the fact that the auditory periphery produces a "distortion product" – a signal whose frequency is identical to the difference between neighboring harmonics. For example, a 100 Hz signal would have harmonics at the multiples 200 Hz, 300 Hz, 400 Hz, etc. If the fundamental frequency wave of 100 Hz were removed from the signal, the 100 Hz distortion product is not lost (Yost, 2009). Much more information can be removed from the signal than just *F*0, and yet the pitch remains detectable. In fact, signals that are high pass filtered at 5000 Hz still have detectable pitch (Moore, 1993).

In spite of the sensitivity of the peripheral auditory apparatus, perception of pitch seems to require a combination of the spectral output of the cochlea, as well as temporal analysis performed by neural anatomy (Yost, 2009). Since Licklider (1951), this temporal analysis has been modeled with autocorrelation.

Perception of raised spoken *F*0 activates attention orientation mechanisms in the brain (Hsu, Evans, & Lee, 2015). However, other *F*0 changes, such as lowering of human voice or changes in pure tones, do not show the same effect. *F*0 raising may thus be used by speakers in situations where more listener attention is sought. Roettger and Grice (2015) find that in Tashlhiyt Berber, questions have both higher overall *F*0 and also higher *F*0 peak than statements with identical CV content and morphosyntactic structure. The language also shows a significant tendency for H to occur on the ultimate syllable of questions, but on the penultimate syllable of identically structured statements. Thus, questions, in which a speaker desires that the listener evaluate utterance content and respond to it, are higher in *F*0 than statements, and are also more likely to end on H than are statements. This raising may activate the above mentioned attention orientation mechanisms, increasing the likelihood of a response.

Perception of raised *F*0 may also play a role in language learning. Exaggerated *F*0 peaks in final position are a feature of infant-directed speech, a form of language that babies prefer to listen to more than adult-directed speech (Fernald & Mazzie, 1991). Babies also learn words faster when exposed to such infant-directed speech (Thiessen, Hill, & Saffran, 2005).

Spoken pitch is maintained at a high level as a strategy of keeping one's turn in discourse (Caspers, 1998, 2003; Selting, 1996); likewise, lowering *F*0 can signal a turn change (Caspers, 2003; Koiso, Horiuchi, Tutiya, Ichikawa, & Den, 1998). Perhaps listeners empathetically recognize the increased effort in sustaining high pitch, and allow the one who is making this effort to maintain possession of the conversational turn.

As a complement to *F*0 raising, drops in *F*0 can be used to signal that information does not require as much attention. Crosslinguistically, parenthetical comments are marked by decrease in *F*0, among other prosodic changes (Canals, 2002, 2003; Döring, 2007).

When perception is combined with production, trained vocalists show some similarity in their performance of linguistic tones to native speakers of Mandarin Chinese, suggesting that some of the same brain structures may be involved in this task (Ning, et al., 2015). This study found that in the midst of pitch-shifted audio feedback, control of high pitch is more accurate for trained vocalists and for native speakers of Mandarin Chinese (which has a phonologically H lexical tone), than it was for naïve speakers and for learners of Mandarin. This task, which serves a linguistic function in Mandarin, could have some similarity to controlling sustained pitch while singing alongside others, as in an ensemble. Other categories of vocal control (e.g., other linguistic tones) under these conditions did not show similarities between trained vocalists and native Mandarin speakers. This set of findings suggests that control of H in spoken language could at times invoke mechanisms that are also employed during singing, although production of lower tones and tonal contours might employ non-musical mechanisms. In spite of the fact that trained vocalists demonstrated better production of H, they did not outperform naïve speakers and L2 learners of Mandarin Chinese in the task of Mandarin tone perception.

In another pitch-shift paradigm study, Sturgeon et al. (2015) found differences between the performance of musicians and non-musicians, suggesting that training in musical pitch control affected responses to audio-vocal disturbance, but not necessarily improvement. Surprisingly, musicians displayed increased latency (longer reaction time) and amplitude compared to non-musicians. It may be the case that the monitoring of pitch in the musicians entails a different gain control relationship, leading to this difference. Whether this reflects greater adaptability on the part of the musicians requires further study.

Sturgeon et al. (2015) study also found that as *F*0 was raised, compensatory response amplitude increased and response latency decreased. The decrease in latency as *F*0 was raised across the 9 levels indicates an increasing sensitivity to stimuli corresponding to higher *F*0. However, at *F*0 targets below typical *F*0 for speech, the compensatory response to pitch-shifted feedback did not vary with changes in *F*0. That is, compensatory responses to low sounds was not different from compensatory responses in the range of typical *F*0. Thus, self-monitoring of spoken *F*0 employs different mechanisms in the higher *F*0 range than it does at default or lower levels.

The above discussion mentions differences between musicians and non-musicians in their production of high *F*0. Some instrumental and choral music is written with stepwise upward pitch shifts for the purpose of heightening the attention of listeners. During the last half of marches and choral pieces, such as hymns in some Christian traditions, the key is shifted upward (“modulated”), and maintained until the end of the music. For example, in military marches, like those of John Philip Sousa, during the Trio, the key is typically raised by five semitones, or a perfect fourth (Hilary Evans, p.c.). Choral music is often raised by one to two semitones during the last half of the piece, an effect which heightens the interest of the singers and the audience (Mark Hsia, p.c.). Choral music is raised by a smaller amount than instrumental music, due to *F*0 limitations inherent in the human voice.

4. Conclusion

High vocal pitch corresponds to a bundle of properties of production and perception. Raising *F*0 employs different gestures in the larynx than lowering *F*0 does. It also takes speakers longer to raise than to lower *F*0 (Zhang, 2013). On the perception side, rises in *F*0, especially sudden or stepwise rises, attract listener attention more than *F*0 lowering does. This orientation of attention may be due in part to listener awareness that *F*0 raising often corresponds to more force in speech. In the acoustic dimension, fundamental frequency is less resistant to decay over distance than are intensity and spectral information. Thus, elevated *F*0 is a robust channel for conveying non-pitch information, such as speech intensity. These acoustic, production, and perception properties of raised vocal *F*0 exist universally, independent of language. They may lead to phonological and discourse properties of H(igh) pitch that are found across the world's languages.

As the high school choir members listened to our peers singing at 880 Hz, these girls whose typical conversational average would have been around 200–230 Hz (Draxler, Schiel, & Ellbogen, 2008; Natour & Wingate, 2009), we knew the answer to Mr. Larimer's question. The human voice at elevated pitch makes an impression.

References

Alipour, F., & Scherer, R. C. (2007). On pressure–frequency relations in the excised larynx. *The Journal of the Acoustical Society of America*, *122*(4), 2296–2305.

Atkinson, J. E. (1976). Inter-and intraspeaker variability in fundamental voice frequency. *The Journal of the Acoustical Society of America*, *60*(2), 440–445.

Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. *The Journal of the Acoustical Society of America*, *65*(5), 1271–1275.

Canals, M. P. (2002). Incidental clauses in spoken Catalan: Prosodic characteristics and pragmatic function. *Proceeding Paper of Speech Prosody*, 559–562.

Canals, M. P. (2003). Prosody and pragmatic in parenthetical insertions in Catalan. *Catalan Journal of Linguistics*, *2*, 207–227.

Caspers, J. (1998). Whos' next? The Melodic of Question vs. continuation in Dutch. *Language and Speech*, *41*(3–4), 375–398.

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, *31*, 251–276.

Crystal, D. (2011). *Dictionary of linguistics and phonetics*, Vol. 30. Oxford: Blackwell Publishing Ltd.

Döring, S. (2007). Quieter, faster, lower, and set off by pauses? Reflection on prosodic aspects of parenthetical constructions in modern German. In N. Dehé, & Y. Kavalova (Eds.), *Parentheticals* (pp. 285–307). Amsterdam: John Benjamins Publishing Company.

Draxler, C., Schiel, F., & Ellbogen, T. (2008, May). *F*0 of adolescent speakers-first results for the German Ph@ ttSessionz Database. In *LREC*.

Evans, J. (2008). African'tone in the Sinosphere. *Language and Linguistics*, *9*(3), 463–490.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation*. *Bell System Technical Journal*, *12*(4), 377–430.

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics* (1st ed.). Oxford: Blackwell Publishing Ltd.

Goldsmith, J. (1984). Meeussen's rule. In Aronoff, Mark, & Richard T. Oehrle (Eds.), *Language sound structure*. Cambridge, MA: MIT Press.

Hirose, H. (1997). Investigating the physiology of laryngeal structures. In W. J. Hardcastle, & A. Marchal (Eds.), *The handbook of phonetic sciences* (pp. 130–152). Oxford: Blackwell Publishing Ltd.

Hsu, C. H., Evans, J. P., & Lee, C. Y. (2015). Brain responses to spoken *F*0 changes: Is H special?. *Journal of Phonetics*, *51*, 82–92.

Hyman, L. M. (2001). Privative tone in Bantu. *Cross-Linguistic Studies of Tonal Phenomena*, 237–257.

Jun, S. A. (2015). High-toned [il] in Korean: Phonetics, intonational phonology, and sound change. *Journal of Phonetics*, *51*, 93–108.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map task dialogs. *Language and Speech*, *41*(3–4), 295–321.

Larson, C. R., White, J. P., Freedland, M. B., & Burnett, T. A. (1996). Interactions between voluntary modulation and pitch-shifted feedback signals: Implications for neural control of voice pitch. In *Controlling complexity and chaos: 9th vocal fold physiology symposium*. San Diego, CA: Singular Press (pp. 279–289).

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *The Journal of the Acoustical Society of America*, *23*(1).

Lieberman, P., Knudson, R., & Mead, J. (1969). Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *The Journal of the Acoustical Society of America*, *45*(6), 1537–1543.

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, *147*(2), 205–222.

Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*(4), 341–383.

Michaud, A., Vaissière, J., & Nguyễn, M. C. (2015). Phonetic insights into a simple level-tone system: ‘Careful’ vs ‘impatient’ realizations of Naxi High, Mid and Low tones. In *Proceedings of ICPhS XVIII* (*18th international congress of phonetic sciences*). Glasgow.

Mohler, R., & Mayer,J. (2001). A Discourse model for pitch-range control. In *Proceeding paper of* 4th *ISCA ITRW on Speech Synthesis*.

Moore, B. C. (1993). Frequency analysis and pitch perception. In *Human psychophysics* (pp. 56–115). New York: Springer. Yost, William A., Fay, Richard R. (Eds.).

Natour, Y. S., & Wingate, J. M. (2009). Fundamental frequency characteristics of Jordanian Arabic speakers. *Journal of Voice*, *23*(5), 560–566.

Ning, L. H., Loucks, T. M., & Shih, C. (2015). A unique contribution of high level tone to differentiating vocal sensorimotor control, *Journal of Phonetics*, 51, 50–69. http://dx.doi.org/10.1016/j.wocn.2014.12.003.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107–120.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollac (Eds.), *Intentions in communication*. Cambridge: MIT Press.

Plomp, R. (1967). Pitch of complex tones. *The Journal of the Acoustical Society of America*, *41*(6), 1526–1533.

Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, production, acoustics and perception* (1st ed.). Oxford: Blackwell Publishing Ltd.

Ritsma, R. J. (1962). Existence region of the tonal residue. I. *The Journal of the Acoustical Society of America*, *34*(9A), 1224–1229.

Robinson, D. W., & Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, *7*(5), 166.

Roettger, T. B., & Grice, M. (2015). The role of high pitch in Tashlhiyt Berber: Evidence from production and perception. *Journal of Phonetics*, *51*, 36–49.

Ryalls, J. H., & Behrens, S. (2000). *Introduction to speech science: From basic theories to clinical applications*. Allyn & Bacon.

Schröder, M. (2001). Emotional speech synthesis: A review. In *Proceedings of the 7th European conference on speech communication and techology: Vol. 1. Eurospeech 2001* (pp. 561–564). Aalborg, Denmark: International Speech Communication Association.

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics*, *6*(3), 371–388.

Shih, C., & Lu, H. Y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics*, *51*, 6–35.

Sturgeon, B. A., Hubbard, R. J., Schmidt, S. A., & Loucks, T. M. (2015). High *F*0 and musicianship make a difference: Pitch-shift responses across the vocal range. *Journal of Phonetics*, *51*, 70–81.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, *85*(2), 901–906.

Tseng, C. Y. (2008). Corpus phonetic investigation of discourse prosody and higher level information. *Language and Linguistics*, *9*(3), 659–719.

Tseng, C. Y., Pin, S. H., Lee, Y., Wang, H. M., & Chen, Y. C. (2005). Fluent speech prosody L Framework and modeling. *Speech Communication*, *46*, 284–309.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, *52*(4B), 1238–1250.

Xu, Y. (2006). Tone in connected discourse. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd Ed. Oxford: Elsevier. 12: 742–750.

Yost, W. A. (2009). Pitch perception. *Attention, Perception, & Psychophysics*, *71*(8), 1701–1715.

Zhang, J. (2013). *The effects of duration and sonority on countour tone distribution: A typological survey and formal analysis*. New York: Routledge.