Research Article

Effect of Vowel Duration and Noise on the Perception of Stop-Glide Continuum

Saransh Jain, Nuggehalli Puttaveeraiah Nataraja and Sudha P. Nair

JAIISH(2015)
Vol 34, pp. 110-116

Affiliations

JSS Institute of Speech and Hearing, MG road, Mysuru.

Corresponding Author

Saransh Jain
Lecturer, JSS Institute of Speech and Hearing, MG Road, Mysuru.
Email: saranshavi@gmail.com

Key Words

Speech Perception
Categorical Perception
Formant Frequency
Noise
Vowel Length

Abstract

The present study aimed to investigate the effect of vowel length and noise on perception of stop-glide continuum. A nine point stop-glide continuum was created by varying the onset frequency of second formant. The continuum was then manipulated for two lengths of vowel following the consonant and three levels of background noise. The stimuli were presented to 30 normal hearing adult participants and the responses were recorded for identification and reaction time task. The results for the identification task revealed that vowel length had a significant effect on the perceptual process with the categorical boundary shifted towards the perception of stop consonant in the presence of long vowel. However, in the presence of noise, the effects of vowel length cues were limited. In such condition, the probable role of multiple acoustic cues in the perception of speech sounds is suggested. On the contrary, no significant effect of vowel length and/or noise was observed in the time taken for the participants to respond to the stimuli. These results were discussed in terms of the role of perceptual normalcy influencing the effect of noise.

Background

The perception of speech involves breakdown of words into its constituent sounds, each with typical spectral and temporal features. Auditory system disintegrates the spectral information at each time frame to provide a complete representation of the speech signal in the brain (Greenberg & Ainsworth, 2012). The decoding of speech signal involves processing of various acoustic cues to derive at the meaning associated with it (McMurray, Tanenhaus, & Aslin, 2002). Thus, speech perception refers to perceptual mapping of acoustic signal to their linguistic representations (Holt & Lotto, 2010).

The acoustic cues help in identification and discrimination of speech sounds which differ in the phonetic categories. Acoustic cues like fundamental frequency and formant frequencies are important for perception of vowels, whereas slope and duration of formant transition, burst duration, voice onset time, closure duration etc., are essential acoustic cues for perception of consonants. Among consonants, voice onset time is vital for voicing perception, while onset frequency of formants and slope of formant transition is useful in discriminating consonants on the basis of place, manner and laterality parameter. However, it may be noted that no single acoustic cue is responsible for the perception of any specific feature of speech sound and a combination of various cues are required for accurate identification. Nonetheless, the perceptual weighting of some acoustic cues are better than others. In the quest of assessing the perceptual weighting of the acoustic cues, researchers (McMurray et al., 2002) created a synthesized stop-glide continuum by varying the onset frequency of F2 and the steepness of the formant transition slope. The researchers indicated the importance of onset and slope of F2 transition for the perception of stop-glide contrast. In the subsequent experiment, the researchers only varied vowel duration while keeping the other acoustic features as constant, and observed no change in the perception. Lastly, the onset of F2 with the slope of formant transition was varied along the vowel length and a significant change in the categorical boundary was observed. The boundary shifted towards the perception of the stop consonant with long vowel. These results revealed robustness of formant cues in differentiation of speech sounds on the basis of the manner of articulation, whereas vowel duration provides least information regarding the identification of stop-glide continuum.

Contrary to the above findings, in a classical experiment by Miller and Liberman (1979), the researchers strongly advocated that late occurring information significantly affect the stop-glide perception. In a set of experiments, they concluded that the perception of stop consonant in a stop glide continuum improves when followed by a long vowel. Vowel length is also an important cue for perception of voicing feature of the speech sounds (Raphael, 1972). Stop consonants preceded by long vowels are perceived as voiced whereas those with shorter vowels are perceived as unvoiced. Earlier investigation by the present authors also revealed that vowel duration is important in the perception of voicing (Jain, Nataraja, & Nair, 2014) at least in quiet situation. Thus, there is disagreement among the research literature regarding the perceptual importance of vowel duration cues in the perception of stop-glide contrast.

In the presence of noise, perception of voicing has been found to be independent of the vowel duration cues. Researchers have reported different acoustic cues were involved in the perception of vowels (Nábelek, Ovchinnikov, Czyzewski, & Crowley, 1996) and consonants (Alwan, 1992) in the presence of noise. The relative influence of noise on vowel and consonant cues revealed that formant information is important for the perception of vowels, whereas formant transition are important for the perception of consonants, especially stops (Parikh & Loizou, 2005). Thus, the perceptual importance of the acoustic cues is dependent upon the noise. It was earlier observed that the voice-unvoiced distinction of the stop consonants vary in the presence of background noise (Jain et al., 2014). However, such investigations with respect to stop-glide contrasts are limited. It is important to assess the perceptual variation across the classes of speech sounds in noisy situation as speech perception in noise for normal individuals mimics the speech perception of hearing impaired individuals. Such studies help in increasing the knowledge regarding the hearing impaired speech processing, and ultimately help in designing new and improved speech processing algorithms for hearing aids and cochlear implants. Thus, there is a need to investigate the perceptual importance of vowel duration cues in the perception of stop-glide contrast in quiet as well as noisy situation.

Further, the studies reviewed above, considered only identification to assess the categorical boundary. It has been recommended by Pisoni and Tash (1974) that reaction time was another measure to evaluate perceptual analysis judgment. Hence, the present study aimed to assess the effect of vowel length and noise on the perception of stop-glide continuum.

Materials and Methods

Participants

A single group repeated measure research design was used and a total of 30 adults (15 Male, 15 Female) within the age range of 18-25 years (mean age: 21.7 yrs), participated in the present study. All the participants were having normal hearing sensitivity (PTA $\leq$ 25 dB HL; SIS $\geq$ 90%, re: ANSI S3.21, 1997) and no associated neurological, psychological, developmental or related pathology. The participants were native speakers of Kannada and belonging to the Mysore region of southern Karnataka. All participants volunteered for the study. The study was approved from the institutional ethics board for testing on human subjects, and an informed written consent was obtained from each participant before the commencement of the study.

Stimuli/Material

The stimuli consisted of a word continuum for which the second formant frequency (F2) of the initial syllable was manipulated in such a way that it yielded a stop consonant at one end point and a glide at the other end point. Both stop and glide were selected as they were present in the vocabulary of the Kannada language. In the beginning, a token of word —taru— was recorded at sampling frequency of 44,100 Hz. The formant and pitch track of the initial stop consonant —ta— was then extracted using the Praat software (ver. 5.1.41; Boersma & Weenink, 2016). The formant track was extracted based on the 'Split Levinston' algorithm suggested by Willems (1986), implemented in Praat. The formant track of the initial five formants till frequency of 5500 Hz was extracted in the time step of 5 ms. The pitch track was then extracted to increase the naturalness of the synthesized stimuli. An acoustic periodicity detection algorithm was employed, which is based on the autocorrelation method (Boersma, 1993), and the pitch was extracted within the frequency range of 75-600 Hz in the time step of 5 ms. The extracted formant tracks were imported to the Klattwork speech synthesis software (ver. 2.29; McMurray, 2009) where the absolute F2 frequency was determined and the onset of the F2 was raised systematically from 2000 Hz 2800 Hz in nine equal steps of 100 Hz each, by adapting the procedure suggested by McMurray et al. (2002). This resulted in a nine step continuum with stop consonant |t| with lower F2 onset at one end point and a glide |ja| with higher F2 onset at other end point. All other acoustic parameters were kept constant throughout the continuum.

The effect of vowel length on the perception of stop-glide contrast was observed by varying the length of the vowel following the target consonant. Two sets of the continuum were taken and the vowel length in one set was kept constant at 250 ms. and this was considered as long vowel continuum. In another set, the vowel length was reduced to 100 ms. and was considered as short vowel continuum. The duration of both the continua was in agreement with the values suggested by Nataraja (2000) and Raphel (1972). Lastly, to obtain the effect of noise, a multi talker speech babble was recorded and added to both the continua at a signal to noise ratio (SNR) of +2, 0 and -2 dB, using the Colea toolbox (A Matlab based software tool for speech analysis; Loizou, 2000). The reference for these SNR values was obtained from the previous studies (Alwan, 1992; Jain et al., 2014). Thus, a total of 360 stimuli in 8 continua's [9 point along each continuum*2 vowel length*4 levels (3 SNR + 1 quiet)*5 repetition] was presented to each participant.

Procedure

The complete testing was carried out in a sound treated room. The stimuli was presented using the laptop (Dell Inspiron 7750 SE) routed via a calibrated audiometer (Maico MA 53) equipped with standard headphones (TDH 39) at 70 dB HL (average most comfortable loudness level). The participants were instructed to identify the word as fast as possible and the responses were recorded for identification and reaction time task. The identification of the either endpoint of the continuum was noted down by the examiner and the reaction time was measured with the help of DMDX reaction time software (Forster & Forster, 2003) installed in the laptop equipped with standard microphone. The stimulus sequence was presented in random order within each of the continuum. Inter stimulus interval of 3000 ms (3 seconds) was kept and considered as the time for the subject to respond. The entire testing was carried out in a single sitting for 40-45 minutes per participant with a break of 5 minutes in between. Five-Six catch trials (involving different stimulus) were presented randomly to ensure that the participants were attending to the stimuli.

Analysis

In the identification task, the percentage correct response for either stop or glide across each point along the continuum was measured and plotted in a line graph to obtain the categorical boundary. The categorical boundary was defined as the point along the graph where the perception shifts from one endpoint to other with 50% probability function. The categorical boundary was measured using logistic regression model with linear or non linear interpolation function using the Prism software (ver. 5.03; GraphPad Software Inc.). The mean reaction time was also measured at each point along the continuum as well as for the categorical boundary. Reaction time procedure added another level of response analysis. This was calculated as the average reaction time for the region of ambiguity (Pisoni & Tash, 1974), i.e. the points along the continuum where the listener was unsure about the exact categorization of the speech sounds (McMurray & Spivey, 1999). The data was analyzed using univariate analysis of variance where the thresholds obtained for the identification and reaction time task were considered as the dependent variables and vowel length and SNR were considered as the fixed factors. The multiple pair wise comparison as a function of noise was carried out using Bonferroni's post hoc test and as a function of vowel length was carried out using paired sample t-test. The selection of the parametric statistics is based on the Kolmogorov-Smirnov test of normality with Lilliefors significance corrections.

Results

The percentage correct identification scores for either end point along the continuum as a function of vowel length and SNR are plotted in Figure 1. As evident from the figure, in quiet situation, the effect of vowel length was clearly visible, where a shift in the categorical boundary was observed towards the perception of stop consonant with long vowel. However, in the presence of noise, no such obvious shift in the categorical boundary was observed. The identification function was plotted for each of the participant and the mean values at the categorical boundary were considered in terms of the change in the onset frequency of F2. These values were obtained by employing the logistic function and subjected to univariate ANOVA. The results revealed a significant effect of vowel length [F (1, 352) = 43.39; p = 0.001], noise [F (3, 352) = 28.41; p = 0.024] and the interaction effect of vowel length and noise [F (3, 352) = 17.32; p = 0.017] on the perception of stop-glide continuum. Multiple pair wise comparison also revealed significant effect of all the four noise conditions on the perceptual boundary identification. Since, only two length of vowel was considered in the present study, post hoc comparison was not possible to estimate the pair wise difference with respect to each condition of noise. Hence, paired sample t-test was performed to identify the pair wise difference in vowel length. The results revealed significant difference between the identification scores at the perceptual boundary for long and short vowels in quiet situation (p = 0.000) as well as in +2 dB SNR (p = 0.001) and 0 dB SNR condition (p = 0.015), but no such difference was observed for -2 dB SNR condition (p = 0.548).

The mean difference was subjected to univariate ANOVA to obtain the statistical significance of differences. The results revealed no significant effect of vowel length [F (1, 352) = 0.004; p = 0.952], noise [F (3, 352) = 2.203; p = 0.096] and the interaction effect of noise and vowel length [F (3, 352) = 0.076; p = 0.973] on the perception of stop-glide continuum. The mean reaction time at the categorical boundary was measured as the average reaction time across the point where the shift in perception

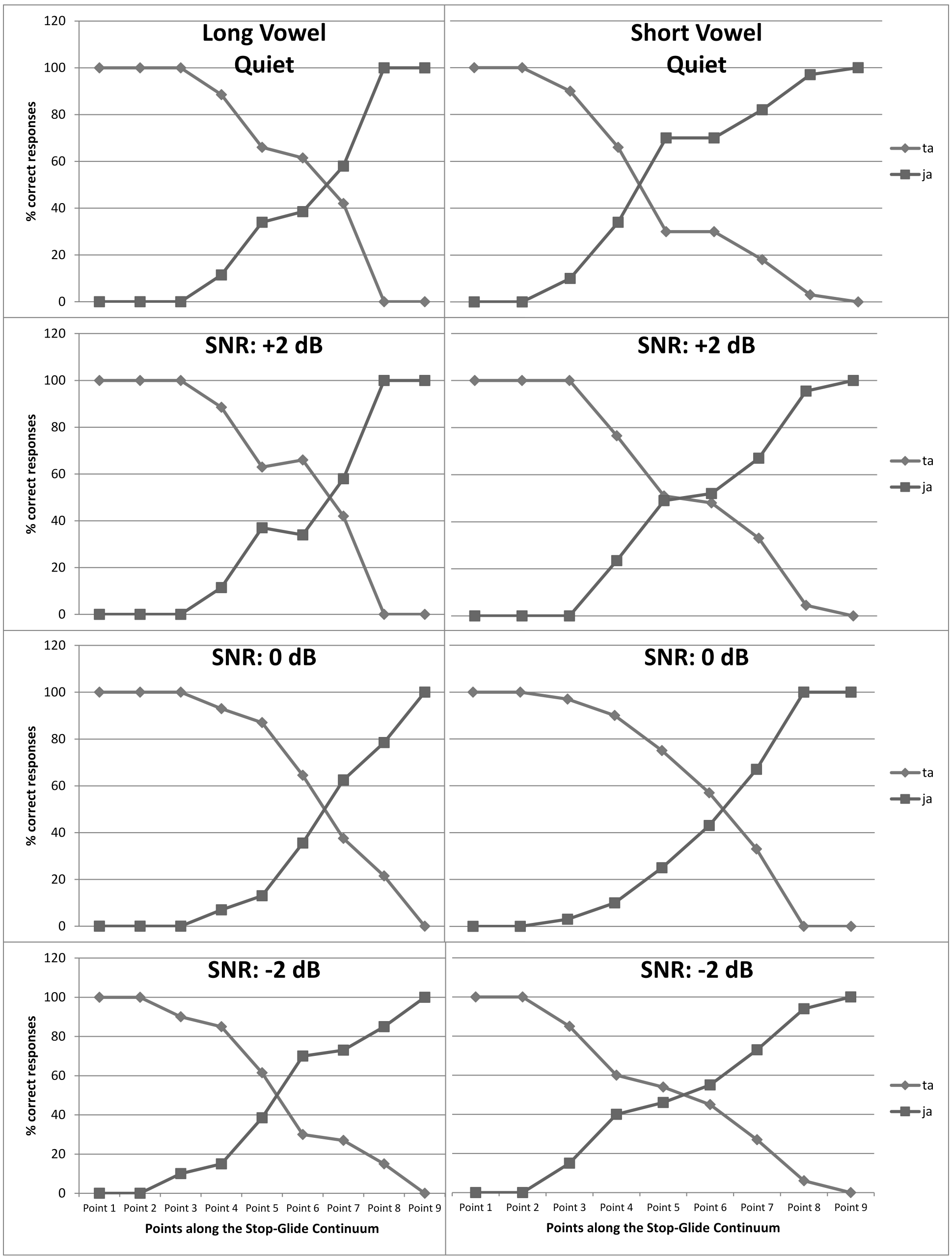

Figure 1: Percentage correct response for the identification of each point along the stop-glide continuum as a function of vowel length and SNR.

was observed. These average values were compared with the average reaction time across all the points along the continuum. The results revealed significantly higher mean reaction time at the perceptual boundary for quiet ($p = 0.037$), +2 dB SNR (0.029) and 0 dB SNR condition ($p = 0.002$), but no such significant difference was observed for -2 dB SNR condition ($p = 0.085$).

Discussion

Previous investigation by the researchers revealed that vowel length has a significant effect on the perception of voicing in quiet situation for Kannada words, however, in the presence of noise; the perceptual weightage of vowel length cue is minimal (Jain et al., 2014). Since, the perception of both voicing and manner of articulation are majorly dependent on the dynamic cues, it was thought to investigate such effect on stop-glide continuum in Kannada. The results of the present study revealed that identification function of stop-glide consonant continuum was influenced by the vowel length.

Long vowel length facilitates the perception of stop consonants in Kannada language. However, in the presence of noise, perceptual weightage of vowel length cues reduces and probable role of other acoustic cues comes into prominence. These results were in agreement with the findings of McMurray,

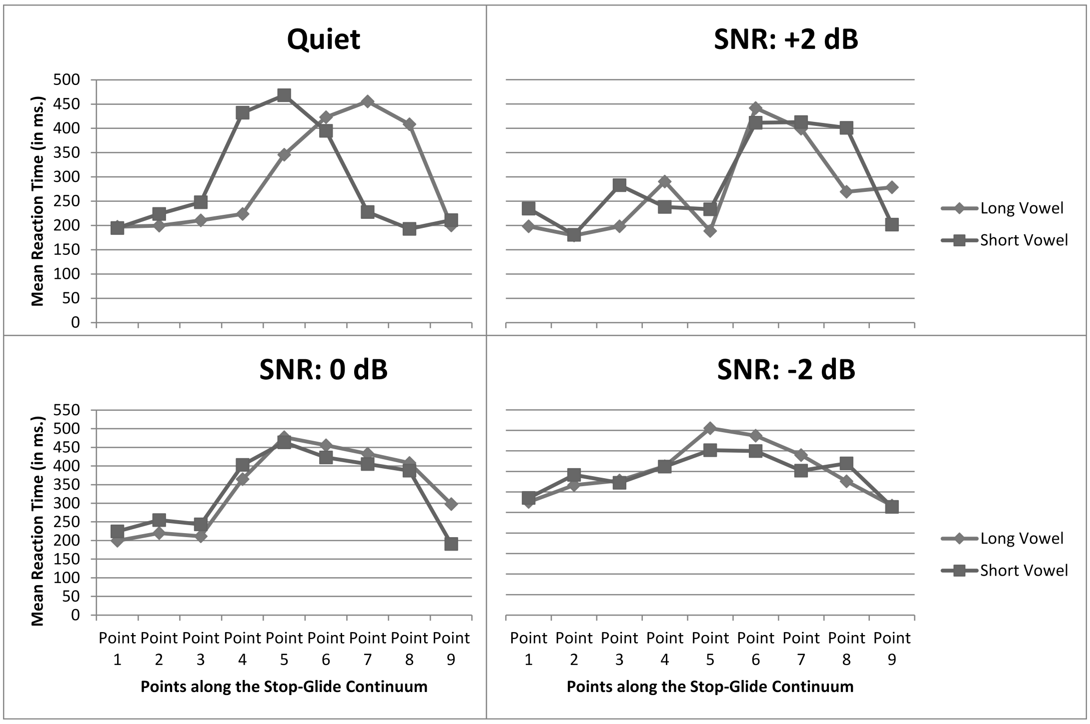

Figure 2: Mean reaction time at each point along the continuum as a function of vowel length and SNR levels for the perception of stop-glide continuum.

Tahensaus and Aslin (2002), in English language. The researcher reported a significant effect of vowel length on manner of articulation with short vowel shifted categorical boundary more towards stop sounds, i.e., increased perception of glide sounds. Wade and Holt (2005) also noted that the perceptual boundary consistently shifted towards the perception of stop consonant (in a |ba|-|wa| distinction) in English i.e. more of |wa| responses when the following sounds were presented in a faster sequence than in slower sequence. Slower sequence facilitate longer vowel duration whereas vowel duration is shorter in faster speech presentation.

Contrary to the present finding, Miller and Liberman (1979) found that the category boundary for the |ba|-|wa| continuum in English could be altered by changing the duration of the steady state portion of the syllable. The researchers found that by increasing the duration of the steady state portion, the categorical boundary shift more towards the |wa| sound. The asymmetry in perception may be due to constellation of cues. Miller and Liberman varied both onset of F2 and transition duration to construct the |ba-wa| continuum. Thus, the effect of multiple dynamic cues may be a reason of differential perception. This reasoning hold true as Shinn (1984) and Shinn, Blumstein and Jongman (1985) reported the effect of syllable length on perception of the |b|-|w| distinction in English gradually disappears as |b| and |w| tokens are synthesized in a less stylized manner and are differentiated by additional, temporal and non-temporal cues such as transition duration, onset formant frequencies, formant-frequency trajectories and noise.

The second important conclusion drawn from the present study is a significant effect of noise on the perception of stop-glide continuum. It was observed that noise has a significant effect on perceptual process, and in its presence, the effect of vowel length cues are minimal. These findings were well reported for other features of speech sounds [voicing: (Jain et al., 2014); manner (stop-fricative contrast): (Soli, Arabie, & Carroll, 1986)] but the explanation with regards to the stop-glide contrast is novel to the present study. Miller and Wayland (1993) showed that in the presence of multi-talker babble noise, rate dependent category boundary shifts hold up even without these additional cues.

The findings suggest that in the presence of speech babble, the listener rely more on the transitional cues than the syllable duration cues, resulting in a less marked distinction between |ba| and |wa|. Zhang and Meng (2011) compared the effects of different acoustic transmission conditions on Mandarin consonants perception. The results showed that the affricates and fricatives were the most intelligible in reverberation, whereas stops were least intelligible. Friction might be the most important cue for consonant perception in reverberation. Those consonants articulated frictionally were more intelligible than others in low frequency noise, while the voiced were more intelligible in high frequency noise. The primate cues for consonant perception in noise were related to the frequency characteristic of noise. For both stops and affricates, the unaspirated sounds were more intelligible than the aspirated sounds in both reverberant and noisy conditions. These findings confirmed the role of multiple acoustic cues in the perception of speech sounds, and the perceptual weightage of each cue vary with respect to the environment in which the speech sounds are perceived.

An interesting finding of the present study is the effect of vowel length or noise on the reaction time task. Unlike the responses for identification time, vowel length and noise had no significant effect on the time taken for the perceptual process to occur. These findings were in opposition to that observed for voicing where introduction of noise had significant effect for the overall mean reaction time and it was found to be higher in noisy conditions (Jain et al., 2014). The results of the present study may be attributed to the perceptual normalization in which the listeners filter out the noise component to arrive at the underlying category (Johnson, 2005). During such situations, the listeners tend to ignore the noise component and pay attention only to the core category. To support the finding, Schwippert and Beinum (1998) also noted no marked differences between within-category and between-category stimulus pairs in a reaction time task for |bak|-|wak| continuum.

Finally, increased reaction time at the region of ambiguity was observed and attributed to the state of confusion which results in the delay in the perceptual process. Such findings are well illustrated in literature (Pisoni & Tash, 1974). Pitt and Samules (1993) changed the onset of F2 and F3 for the |ba| sound and presented in three noise conditions, i.e. binaural, ipsilateral & contralateral noise and found increased reaction time at the phoneme boundary in the 8-point continuum, for all the three conditions.

The findings of the present study may contribute towards understanding the process of speech perception in normal hearing individuals. Under noisy environment, the perceptual mechanism of the normal hearing individuals is similar to that of the hearing impaired listeners in the quiet situation (Zurek & Delhorne, 1987), thus investigation of the normal perceptual processing in noisy environment provide an insight about the speech perception in hearing impaired individuals. Another potential implications of the present study include improving performance of the automatic speech recognition software, developing speech coders for speech recognition and noise reduction algorithms.

Conclusions

The present study aimed at investigating the effect of vowel length and noise on the perception of stop-glide continuum. The results revealed that vowel length has a significant effect on perceptual process in quiet situation. However, in the presence of noise, the perceptual weightage of vowel length cues are not significant. Thus, it may be concluded that noise has a significant effect on the perceptual weightage of acoustic cues and the findings supported a probable role of multiple acoustic cues in the perceptual process.

References

Alwan, A. (1992). *The role of F3 and F4 in identifying the place of articulation for stop consonants.* In Proceedings of the International Conference on Spoken Language Processing (pp. 1063-1066). Banff: Canada.

Boersma, P. (1993). *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.* In Proceedings of the Institute of Phonetic Sciences (Vol. 17, pp. 97-110). Amsterdam: University of Amsterdam.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.10) [Windows]. Retrieved from http://www.praat.org/

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 35*(1), 116-124.

Greenberg, S., & Ainsworth, W. (2012). Listening to Speech: An Auditory Perspective. Retrieved from https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc_100025692216.0x000001

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics, 72*(5), 1218-1227. http://doi.org/10.3758/APP.72.5.1218

Jain, S., Nataraja, N. P., & Nair, S. P. (2014). A study on the role of temporal integration of VOT, vowel length and noise in the perception of bilabial stops. *Journal of All India Institute of Speech and Hearing, 33*(1), 51-58.

Johnson, K. (2005). *Speaker Normalization in Speech Perception.* In D. B. Pisoni & R. E. Remez (Eds.), The Handbook of Speech Perception (pp. 363?389). Oxford, UK: Blackwell Publishing Ltd.

Loizou, P. C. (2000). Colea: A Matlab software tool for speech analysis. Retrieved from http://www.utdallas.edu/~loizou/speech/colea.htm

McMurray, B. (2009). KlattWorks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research. Manuscript in preparation.

McMurray, B., & Spivey, M. J. (1999). *The Categorical Perception of Consonants: The Interaction of Learning and Processing.* In Proceedings of the Chicago Linguistics Society (pp. 205-219). Chicago.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33-42.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(6), 457-465.

Miller, J. L., & Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics, 54*(2), 205-210.

Nábelek, A. K., Ovchinnikov, A., Czyzewski, Z., & Crowley,

H. J. (1996). Cues for perception of synthetic and natural diphthongs in either noise or reverberation. *The Journal of the Acoustical Society of America, 99*(3), 1742-1753.

Nataraja, N. P. (2000). *Transformation of speech of hearing impaired* (Departmental Project No. 1). Mysore, India: All India Institute of Speech and Hearing.

Parikh, G., & Loizou, P. C. (2005). The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America, 118*(6), 3874.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*(2), 285-290.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology. Human Perception and Performance, 19*(4), 699-725.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51*(4), 1296-1303.

Schwippert, C., & Koopmans-van Beinum, F. J. (1998). *Phoneme boundary perception in dyslexic and normal-reading adults.* In Proceedings of the Institute of Phonetic Sciences. Amsterdam.

Shinn, P. (1984). On the role of the amplitude envelope for the perception of [b] and [w]. *The Journal of the Acoustical Society of America, 75*(4), 1243. http://doi.org/10.1121/1.390677.

Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics, 38*(5), 397-407. http://doi.org/10.3758/BF03207170

Soli, S. D., Arabie, P., & Carroll, J. D. (1986). Discrete representation of perceptual structure underlying consonant confusions. *The Journal of the Acoustical Society of America, 79*(3), 826-837.

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 67*(6), 939-950.

Willems, L. (1986). *Robust formant analysis* (IPO Report No. 529) (pp. 1-25). Eindhoven: Institute for Perception Research.

Zhang, S., & Meng, Z. (2011). *A comparative study on perceptual characteristics of mandarin consonants in different acoustic transmission conditions.* In Proceedings of ICPhS, 2011 (pp. 2296-2299). Hong Kong. Retrieved from https://www.internationalphoneticassociation.org

Zurek, P. M., & Delhorne, L. A. (1987). Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. *The Journal of the Acoustical Society of America, 82*(5), 1548-1559.