Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Acoustic correlates of Dutch lexical stress re-examined: Spectral tilt is not always more reliable than intensity

Giulio G.A. Severijnen[1], *Hans Rutger Bosker*[1,2], *James M. McQueen*[1,2]

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, The Netherlands
[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
giulio.severijnen@donders.ru.nl, hansrutger.bosker@mpi.nl, james.mcqueen@donders.ru.nl

Abstract

The present study examined two acoustic cues in the production of lexical stress in Dutch: spectral tilt and overall intensity. Sluijter and Van Heuven (1996) reported that spectral tilt is a more reliable cue to stress than intensity. However, that study included only a small number of talkers (10) and only syllables with the vowels /aː/ and /ɔ/.

The present study re-examined this issue in a larger and more variable dataset. We recorded 38 native speakers of Dutch (20 females) producing 744 tokens of Dutch segmentally overlapping words (e.g., VOORnaam vs. voorNAAM, "first name" vs. "respectable"), targeting 10 different vowels, in variable sentence contexts. For each syllable, we measured overall intensity and spectral tilt following Sluijter and Van Heuven (1996).

Results from Linear Discriminant Analyses showed that, for the vowel /aː/ alone, spectral tilt showed an advantage over intensity, as evidenced by higher stressed/unstressed syllable classification accuracy scores for spectral tilt. However, when all vowels were included in the analysis, the advantage disappeared.

These findings confirm that spectral tilt plays a larger role in signaling stress in Dutch /aː/ but show that, for a larger sample of Dutch vowels, overall intensity and spectral tilt are equally important.

Index Terms: lexical stress, spectral tilt, intensity, speech production, cue weighting

1. Introduction

Lexical stress plays an important role in speech perception in many languages, such as English and Dutch. For instance, it can distinguish segmentally identical word pairs, such as 'OBject' vs. 'obJECT' (capitalization indicates lexical stress) in English, or *VOORnaam* vs. *voorNAAM* ('first name' vs. 'respectable') in Dutch. Previous literature has identified a number of acoustic correlates of lexical stress such as duration, fundamental frequency, intensity, spectral tilt, and vowel quality, but there are between-language differences in the weighting of these cues in perception [1], [2]. The present study investigated the reliability of two acoustic cues in the production of lexical stress: spectral tilt and intensity. Changes in spectral tilt refer to a shift of intensity over the spectrum, such that low frequency components are hardly affected, and the intensity increase is concentrated in higher harmonics [3, p. 2472]. This leads to a less negative tilt in the spectrum of stressed syllables compared to unstressed syllables. In contrast, changes in overall intensity lead to an increase in energy in all frequencies. Previous research suggests that while both cues signal lexical stress patterns, spectral tilt is a more reliable cue in Dutch. However, these results might depend on the specific vowels included in the analyses. We therefore re-examined this issue in a larger and more variable dataset.

The argument in favor of spectral tilt originates from the acoustic effects of vocal effort on the spectral profile of speech. For example, early studies showed that speech sounds that are produced with more effort contain more energy in frequency ranges above 500 Hz [4], [5]. Since stressed syllables are produced with more effort, this has led researchers to investigate whether spectral tilt, as opposed to overall intensity, serves as a better correlate of lexical stress.

Sluijter and Van Heuven [3] (hereafter SvH1996) investigated this in Dutch. They recorded ten Dutch native speakers, who produced lexical (*CAnon* vs. *kaNON*, "canon" vs. "cannon", capitalization indicates lexical stress) and reiterant versions of minimal stress pairs (*NAna* vs. *naNA*) appearing in accented ([F+]; "Will you *CAnon* say (rather than song)") or non-accented ([F-]; "Will you *CAnon* say (rather than write down)") positions. They then measured overall mean intensity and spectral tilt (measured as intensity in four contiguous frequency bands B1-B4: 0-0.5, 0.5-1, 1-2, 2-4 kHz) in the vowel of each syllable. Results from Linear Discriminant Analyses (LDA), which predicted the stress pattern of each produced token based on a set of acoustic predictors, showed that spectral tilt was a better cue to separate stressed from unstressed syllables. More specifically, the analysis in which they included only the higher frequency bands, B2-B4, yielded higher accuracy scores ([F+]: 99% and 94% for lexical and reiterant speech respectively, [F-]: 86% and 81%) compared to the analysis for intensity ([F+]: 88% and 80%, [F-]: 69% and 63%). The authors concluded that frequency regions above 500 Hz are thus more important in signaling lexical stress in Dutch than lower frequency regions and hence that spectral tilt is more important than overall intensity in Dutch.

While SvH1996 provide strong evidence in favor of spectral tilt, they only tested a relatively small speaker sample, producing primarily the vowels /aː/ and /ɔ/. As such, their results may have been confounded by the vowels' formant characteristics (specifically the first formant; F1). That is, /aː/ is one of only very few Dutch vowels that typically display an F1 above 500 Hz (males: 670 Hz, females: 912 Hz [6]). Consequently, the observed power increase in higher frequency bands in stressed syllables may be due to participants increasing the energy around the F1 of /aː/ (which is located in B2; 500-1000 Hz) instead of an overall increase in higher frequencies.

10.21437/SpeechProsody.2022-57

Indeed, the vowel /ɔ/ has an F1 of 419 Hz (females) and 402 Hz (males) [6], and the results showed that, for that vowel, B1 was affected by lexical stress.

Studies on other languages have found contradicting results regarding the role of spectral tilt in cuing stress. For instance, while spectral tilt does seem to be a reliable cue to stress in Polish, Macedonian, and Bulgarian [7], it is not in English [8] and Spanish [9]. Furthermore, in Central Catalan, spectral tilt was only a reliable cue to lexical stress in syllables containing the vowel /a/ (which is reduced to schwa in unstressed syllables). This was not the case for the vowel /i/, which maintains its vowel quality in unstressed syllables [9]. The authors concluded that the changes in spectral tilt are more related to vowel reduction rather than suprasegmental stress.

It thus seems that the strength of spectral tilt as a cue to lexical stress may depend on the language, the vowel, and susceptibility to vowel reduction in unstressed syllables. Research on Dutch has not yet investigated these acoustic correlates in a wider variety of vowels. The present study therefore examined intensity and spectral tilt as acoustic correlates to lexical stress in Dutch in a larger dataset targeting ten different vowels.

2. Methods

We ran a speech production experiment where we recorded Dutch participants producing a large set of segmentally overlapping words (e.g., *VOORnaam* vs. *voorNAAM*, "first name" vs. "respectable"). Following the procedure in SvH1996, we measured overall intensity and spectral tilt for each vowel in stressed and unstressed syllables, and ran LDAs to inform us on the reliability of both cues to signal lexical stress in Dutch.

2.1. Participants

We recruited 38 native speakers of Dutch from the Radboud University participant pool (20 female, 18 male, age range: 17-33, M_{age} = 22.7, SD_{age} = 4.1). All participants gave informed consent and received a monetary reward or course credits for their participation. No participant reported having any speaking and/or reading problems.

2.2. Materials

We selected target word pairs that were fully or partially segmentally identical, but differed in lexical stress. This set included 6 disyllabic minimal stress pairs (e.g., *VOORnaam* vs. *voorNAAM*) and 56 temporarily overlapping pairs (i.e., word pairs with one overlapping syllable; *TAlen /*ˈta:.lən/ vs. *taLENT* /ta:.ˈlɛnt/, 'languages' vs. 'talent'). In these temporarily overlapping pairs, we analyzed only the overlapping syllable (e.g., *ta* in this example; across the set as a whole, half of these overlaps were in first and half in second syllable position). We included these word pairs to increase the external validity of the outcomes, since the number of Dutch minimal stress pairs is limited. The stimuli contained ten different monophthongs: /a:/ (18,4% of the data), /ɑ/ (6,6%), /i/ (20,6%), /ɪ/ (7,4%), /e/ (2,9%), /ɛ/ (6,6%), /u/ (4,4%), /o/ (16,2%), /ɔ/ (6,6%), and /y/ (2,9%). 7% of the stimuli contained diphthongs; these were excluded from the analyses. The complete stimulus list has been made available at:

https://osf.io/drq7g/?view_only=fd52a862000e460597a13a137e066ef6.

We further wanted to separate the acoustic correlates of lexical stress from those of sentence accent, following SvH1996. We thus created a set of carrier sentences in which the target words would appear in different sentence contexts: either in an accented position ([F+]) or a non-accented position ([F-]). For example, the carrier sentences were:

(1) *Eerst had Jan met enthousiasme fiets gezegd* ('First had Jan with enthusiasm bike said'),

(2) *Toen had Jan het woord VOORnaam gezegd*, ('Then had Jan the word first name said'),

(3) *Daarna had Koen het woord VOORnaam gezegd,* ('Afterwards had Koen the word first name said').

The sentences were created in such a way that they would naturally induce correct sentence accent placement, but for clarity, the accented words were always underlined in the recording script provided to the participants. Moreover, the stressed syllables in the target words were given in capitals. The sentence structure of the carriers was identical across trials; only the filler words in the first sentence (e.g., *fiets*), the target words (e.g., *VOORnaam*) and the names (e.g., *Jan, Koen*) were rotated across sentences.

2.3. Procedure

The experiment consisted of a single recording session. Participants were seated in a sound-attenuating booth and wore a head-mounted Omnitronic HS-1100 microphone. A Behringer X-Air XR18 mixer was used and the recordings were digitized at a 44.1 kHz sampling rate.

The target words and sentences were visually presented in Dutch orthography on a computer screen using SpeechRecorder [10], and participants were instructed to read them aloud as naturally as possible. Trials were presented in a pseudo-randomized order, ensuring that the two members of the same word pair were at least 62 trials apart. The stimulus list was presented twice, each time in a different order. All participants received the same order of presentation, reducing any between-talker variability caused by possible order effects. Each trial consisted of two speaking contexts: Participants first produced the word in isolation, followed by the three carrier sentences in the order given in the example above. The analyses focused only on the target words in the sentences.

2.4. Data analysis

2.4.1. Acoustic measurements

The recordings were automatically forced aligned using the WebMAUS Basic tool [10], which segmented the target words and vowels. These were subsequently manually checked by six researchers.

We measured spectral tilt and intensity in the vowel of each stressed and unstressed target syllable. Recall that for the temporarily overlapping pairs, we excluded the irrelevant syllables.

All acoustic measurements were performed in Praat [11]. Intensity was measured as the mean overall intensity in each syllable using the 'Get intensity' function. To allow for comparison to SvH1996, spectral tilt was measured as the mean intensity in four contiguous frequency bands B1-B4: 0-0.5, 0.5-1, 1-2, 2-4 kHz.

2.4.2. Linear Discriminant Analyses

We ran separate LDAs for each cue to examine how well each cue supported classification of stressed and unstressed syllables. An LDA tries to find the optimal linear combination of a set of predictors to separate a dataset into different classes. We trained the model on 80% of the data and then tested it on the remaining 20%. The output contained accuracy scores for how well the model performed the classification.

Using the MASS [12] package in R, we ran a separate model for overall intensity, separate models for each frequency band for spectral tilt, a model for all frequency bands combined, and a final model for only the higher bands (B2-B4). In each analysis, we took stressed vs. unstressed syllable as the dependent variable. Note that this differs from SvH1996, whose models classified stress patterns (initial vs. final stress) by entering two predictors (one value for each syllable). This was possible because they only included minimal stress pairs. In the current dataset, this was not possible due to the temporarily overlapping pairs. To account for this, we ran separate analyses for each syllable position. Finally, we also separated the analyses by sentence accent. This resulted in 5080 ([F+]) and 5163 ([F-]) observations for the first syllable analyses, and 4391 ([F+]) and 4393 ([F-]) observations in the second syllable analyses.

3. Results

3.1. Intensity

Table 1 provides mean intensity values for vowels in stressed and unstressed syllables. In all conditions, we observed a higher intensity in stressed syllables compared to unstressed syllables. Furthermore, this difference was smaller in non-accented position.

Table 1: *Mean overall intensity (dB) in stressed and unstressed vowels*

	[F+]		**[F-]**	
	First syllable	**Second syllable**	**First syllable**	**Second syllable**
Stressed	73	72	68	66
Unstressed	69	67	66	64

Next, we ran LDAs with intensity as predictor. Results for the first syllable yielded 70% and 60% accuracy scores in the [F+] and the [F-] conditions, respectively. For the second syllable, we observed 72% and 62% accuracy scores, respectively.

3.2. Spectral tilt

Mean intensity values in the four contiguous frequency bands (B1-B4) are plotted in Figure 1. For reasons of space, we have plotted this for the cardinal vowels /a:/, /u/, and /i/ only, separating accented and non-accented positions, averaging across first and second syllables.

Figure 1 shows that for /a:/ the power difference between stressed and unstressed syllables seems to be larger in higher frequency bands B2-B4. However, for the other vowels, the power difference seems to be more equally divided across all bands, including B1.

Next, we ran LDAs with the different frequency bands as predictors. These LDAs included data from all vowels, not just the cardinal vowels. Table 2 shows results from these analyses – separately for each syllable position and sentence accent – in each frequency band.

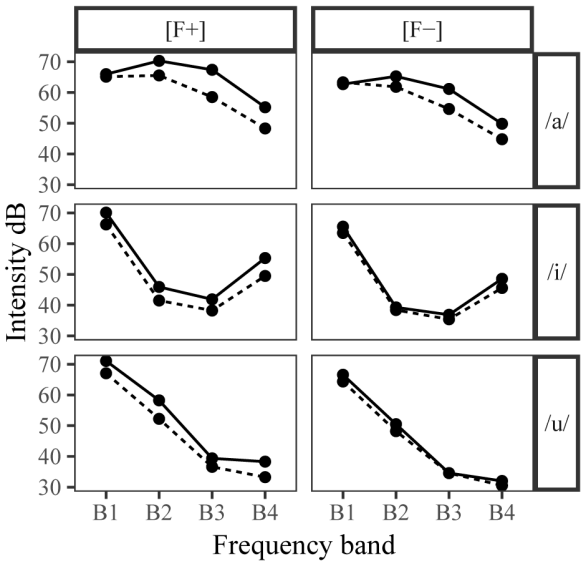

Figure 1: *Mean intensity (dB) of stressed syllables (solid lines) and unstressed syllables (dashed lines) in four contiguous frequency bands. The data have been plotted for three different vowels: /a:/, /i/, and /u/ in two sentence contexts ([F+] and [F-])*

Table 2: *Accuracy scores (%) from linear discriminant analyses split by syllable and sentence context. Results are shown for each frequency band separately, one with all frequency bands combined, and one with only B2-B4.*

	[F+]		**[F-]**	
Frequency band	**First syllable**	**Second syllable**	**First syllable**	**Second syllable**
B1	59	70	51	62
B2	60	57	54	55
B3	59	59	54	55
B4	65	63	59	61
all	67	72	59	63
B2-B4	66	65	58	62

Results showed that, contrary to SvH1996, the higher frequency bands (B2, B3, and B4) did not consistently result in higher accuracy scores compared to B1 alone. This suggests that, in Dutch, spectral tilt is not always the better cue to classify lexical stress compared to intensity. Furthermore, following SvH1996, when comparing the 'all' analysis to the 'B2-B4' analysis, we again see that accuracy scores do not improve. This suggests that in a dataset with a wide variety of vowels, frequencies below 500 Hz are as important as the higher frequencies in productions of lexical stress.

3.3. Comparison to SvH1996: subset analyses

When comparing the accuracy scores for intensity with those for spectral tilt (taking the results from the 'B2-B4' analysis), we see that intensity yielded higher accuracy scores in all conditions. Contrary to SvH1996, this suggests that intensity is a better cue to separate stressed from unstressed

syllables than spectral tilt. This difference from SvH1996 may be explained by the vowels included in the analysis.

To investigate this further, we ran a subset analysis in which we included only syllables that contained the vowel /aː/. In addition, we noticed that we obtained overall lower accuracy scores compared to those reported in SvH1996, possibly due to the larger variability in talkers and vowels in the current dataset. Therefore, we ran the new LDAs on the same number of participants as in SvH1996.

In these subset analyses, we randomly selected 10 participants (balanced gender) and included only the vowel /aː/. We then ran new LDAs for this smaller subset for overall intensity and the 'B2-B4' analysis for spectral tilt.

Results showed that for the first syllable, intensity yielded 71% and 68% accuracy scores in the [F+] and [F-] condition respectively. For the second syllable, we observed 67% and 66% accuracy scores. For spectral balance on the other hand, we observed 81% and 74% for the first syllable, and 79% and 71% for the second syllable. Note that the results for the second syllable varied depending on which trials were randomly sampled by the LDA. This is due to the number of second-syllable observations ([F+]: 120, [F-]: 120) being smaller than the number of first-syllable observations ([F+]: 378, [F-]: 380).

In general, these results confirm that when we reduce variability in talkers and vowels, overall accuracy scores indeed tended to improve, especially for spectral tilt. Moreover, and more interestingly, these results suggest that when we only include the vowel /aː/, spectral tilt does show an advantage over intensity. Crucially, this advantage disappeared when the other vowels were included in the analyses.

4. Discussion and conclusions

The present study examined two acoustic correlates to lexical stress in Dutch: overall intensity and spectral tilt. We recorded native speakers of Dutch, producing segmentally overlapping words. Results from LDAs showed that for the vowel /aː/ accuracy scores were higher in the models for spectral tilt compared to those for intensity, corroborating SvH1996. But when we included other vowels, spectral tilt and intensity were equally important to signal lexical stress in Dutch.

There are two important differences between the present study and SvH1996 that may explain these different results. First, the present study included more participants and more trials for each participant, increasing the external validity. Second, the present study included a wider variety of vowels compared to SvH1996, in which the majority of the syllables contained the vowel /aː/ (with an F1 above 500 Hz). Presumably, the latter is what drives the different results for spectral tilt in the present study compared to SvH1996. Indeed, when running a subset analysis on only the vowel /aː/ in our data, we did replicate the result in SvH1996. When the entire dataset was included, the advantage for spectral tilt disappeared.

We may speculate as to why spectral tilt is a more important cue to lexical stress in the vowel /aː/ compared to other vowels. Note that /aː/ is one of only very few vowels in Dutch with an F1 above 500 Hz [6], resulting in an F1 in frequency band B2. Given this characteristic, two possible acoustic processes may underlie the observed modulation of spectral tilt. First, we propose that there may be more spectral sharpening in stressed syllables. Spectral sharpening enhances the energy difference between the peaks and troughs around the formants in the spectrum, which increases intelligibility [13], [14]. This increase around vowel formants would result in an energy increase in B2 for the vowel /aː/ in stressed (relative to understressed) syllables. Consequently, the spectral energy profile of stressed /aː/ would diverge primarily from unstressed /aː/ from B2 onwards, while for the other vowels the energy difference would be more evenly distributed across the four bands. Second, vowel reduction in unstressed syllables could play a role. In Dutch, the vowel /aː/ is reduced towards /ɑ/ in unstressed syllables, lowering the F1 and possibly moving it from B2 to B1. Consequently, the vowel /aː/ in unstressed syllables (arguably with vowel reduction, pronounced as [ɑ]) would have lower energy in B2, increasing the energy difference in this higher frequency band compared to stressed syllables. This would be in line with previous research on Central Catalan [9].

Our findings thus suggest that, depending on the vowel, spectral tilt becomes a more or less important cue to lexical stress. This varying strength of spectral tilt in turn requires a reconsideration of the importance of intensity, which, depending on the vowel, was equally strong as spectral tilt. We thus propose to extend an explanation offered by SvH1996. They argued that stressed syllables are produced with more vocal effort, which leads to a more pulse-like shape of the glottal source signal and results in more energy in higher frequency regions [4], [5]. We would like to add that speech produced with more effort not only leads to less negative spectral tilt, but also to higher overall intensity, which has also been found in Lombard speech [15] and clear speech [16]. Depending on the vowel, however, spectral tilt may become more important in cueing lexical stress.

Our findings also challenge the conclusion in SvH1996 that intensity is mostly an acoustic correlate of sentence accent, not lexical stress. They based their conclusion on a small difference in overall intensity between stressed and unstressed syllables, and the poor performance of intensity in the LDA. Contrary to their results, we observed a sizable difference in overall intensity in the [F-] condition (syllable 1: 2 dB, syllable 2: 2 dB). Moreover, results from the LDAs in the [F-] condition yielded comparable results for intensity and spectral tilt. We thus propose that intensity, similar to spectral tilt, is a reliable correlate of lexical stress in both accented and unaccented words.

In sum, we found that, for the vowel /aː/, spectral tilt plays a larger role in signaling lexical stress in Dutch than overall intensity. For other vowels, however, overall intensity (including the frequencies below 500 Hz) and spectral tilt are equally important. An interesting avenue for future research would be to investigate whether these cues are also weighed equally in perception.

5. Acknowledgements

This research was funded by the Donders Centre for Cognition. This work is part of a larger project examining individual talker variability in the production of lexical stress in Dutch. We would further like to thank Sanne van Eck, Dennis Joosen, Esther de Kerf, Inge Pasman, Carlijn van Herpt, and Abdellah Elouatiq, who helped to annotate the data.

6. References

[1] N. Cooper, A. Cutler, and R. Wales, "Constraints of Lexical Stress on Lexical Access in English: Evidence from Native and Non-native Listeners," *Lang Speech*, vol. 45, no. 3, pp. 207–228, Sep. 2002, doi: 10.1177/00238309020450030101.

[2] A. Jesse, K. Poellmann, and Y.-Y. Kong, "English Listeners Use Suprasegmental Cues to Lexical Stress Early During Spoken-Word Recognition," *J Speech Lang Hear Res*, vol. 60, no. 1, pp. 190–198, Jan. 2017, doi: 10.1044/2016_JSLHR-H-15-0340.

[3] A. M. C. Sluijter and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *The Journal of the Acoustical Society of America*, vol. 100, no. 4, pp. 2471–2485, Oct. 1996, doi: 10.1121/1.417955.

[4] R. D. Glave and A. C. M. Rietveld, "Is the effort dependence of speech loudness explicable on the basis of acoustical cues?," *The Journal of the Acoustical Society of America*, vol. 58, no. 4, pp. 875–879, Oct. 1975, doi: 10.1121/1.380737.

[5] J. Gauffin and J. Sundberg, "Spectral Correlates of Glottal Voice Source Waveform Characteristics," *J Speech Lang Hear Res*, vol. 32, no. 3, pp. 556–565, Sep. 1989, doi: 10.1044/jshr.3203.556.

[6] P. Adank, R. van Hout, and R. Smits, "An acoustic description of the vowels of Northern and Southern Standard Dutch," *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1729–1738, Sep. 2004, doi: 10.1121/1.1779271.

[7] K. Crosswhite, "Spectral tilt as a cue to stress in Polish, Macedonian and Bulgarian," in *Proceedings of the XVth International Conference of Phonetic Sciences*, Barcelona, 2003, vol. 3, pp. 767–770.

[8] N. Campbell and M. E. Beckman, "Stress, Prominence, and Spectral Tilt," in *Stress, prominence and spectral tilt*, A. Botinis, G. Kouroupetroglou, and G. Carayiannis, Eds. ESCA and University of Athens Department of Informatics, 1997, pp. 67–70.

[9] M. Ortega-Llebaria and P. Prieto, "Acoustic Correlates of Stress in Central Catalan and Castilian Spanish," *Lang Speech*, vol. 54, no. 1, pp. 73–97, 2011, doi: 10.1177/0023830910388014.

[10] C. Draxler and K. Jänsch, *SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software*. Lisbon, 2004.

[11] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2019. [Online]. Available: www.praat.org

[12] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.

[13] A. M. Simpson, B. C. J. Moore, and B. R. Glasberg, "Spectral Enhancement to Improve the Intelligibility of Speech in Noise for Hearing-impaired Listeners," *Acta Oto-Laryngologica*, vol. 109, no. sup469, pp. 101–107, Jan. 1990, doi: 10.1080/00016489.1990.12088415.

[14] T.-C. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression," in *Interspeech 2012*, Sep. 2012, pp. 635–638. doi: 10.21437/Interspeech.2012-197.

[15] M. Cooke and Y. Lu, "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," *J. Acoust. Soc. Am.*, vol. 128, no. 4, pp. 2059–2069, Oct. 2010, doi: 10.1121/1.3478775.

[16] M. A. Picheny, N. I. Durlach, and L. D. Braida, "Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech," *J Speech Lang Hear Res*, vol. 29, no. 4, pp. 434–446, Dec. 1986, doi: 10.1044/jshr.2904.434.