Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Voice transforms for affect control in Irish speech synthesis

Anna Maria Giovannini[1]*, Zihan Wang*[1]*, Maria O'Reilly*[1]*, Ailbhe Ní Chasaide*[1]*, Christer Gobl*[1]

[1]School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland
giovanna@tcd.ie, zwang7@tcd.ie, moreil12@tcd.ie, anichsid@tcd.ie, cegobl@tcd.ie

Abstract

This paper reports on an experiment using voice transforms to alter the perceived affect in synthetic utterances of Irish, with a view to controlling affect in the spoken output of an Irish AAC device. The transforms were guided by prior experience and by voice source analyses of utterances by a male speaker with an *angry, happy, sad, bored, relaxed* and *neutral* voice. The neutral utterance was modified to incorporate stylised voice transforms targeting these affects. Modifications included global shifts affecting the entire utterance, local shifts affecting only accented syllables, and a combination of global and local changes. Stimuli targeting *sad* and *happy* included tempo changes and formant shifts were included for *happy*. Listeners' evaluations most positively identified the high activation affects *happy* and *angry*. Stimuli targeting *sad* were also effective, while those targeting *bored* and *relaxed* were not, although *bored* was positively associated with some of the sad-targeting stimuli. Results for low activations states are confounded by the fact that the *neutral* stimulus was to some degree biased towards *bored, sad* and *relaxed* affects. Of the three types of transforms, global, local and combined, the most effective appears to vary with the targeted affect.

Index Terms: voice quality, voice source, affect, emotion, speech synthesis, emotion, Irish

1. Introduction

In this paper, we explore possible affect-targeting voice transforms, i.e., formula-based changes that could potentially be applied to a neutral utterance to alter its perceived affective colouring. These transforms are envisaged as shifts primarily in the voice source and f_0 characteristics of the utterance, and for some affects include shifts in tempo and formant values.

Many studies of the vocal correlates of affect have focused on features like f_0, intensity and tempo [1-5]. As often pointed out in these studies, a particular gap has been the coverage of voice quality features, and this has been suggested as a reason why perception experiments replicating the production features mentioned above often yield disappointing results.

Studies that have focused on the mapping of voice quality to affect have shown that changes in the voice quality of an utterance can greatly affect its perceived affective colouring [6-8]. These studies found no one-to-one cueing of voice quality for affect: rather, a particular voice quality would appear to be associated with a range of affects. For example, a lax-creaky voice was found to be associated with sadness, intimacy and boredom, whereas a tense voice quality was found to be associated with a range of high-activation states. The present study examines, therefore, not only how effectively a particular transform signals an affect, but also how affect-specific the cueing is.

Most research on the vocal correlates of affect has emerged from the field of psychology, and results tend to be presented in terms of generalised trends, i.e., global, utterance-wide shifts in the mean and standard deviation (or level and dynamic range) of measures like f_0, intensity and tempo. From a more linguistically oriented perspective, the question arises as to whether the observed changes are indeed global, or whether they might also involve specific utterance-internal changes, linked to the prosodic constituents of the utterance. Of relevance here are the considerable utterance-internal voice source modulations found in utterances, particularly in the realisation of accented, focally accented and unaccented syllables, as well as sentence declination [9-15]. It seems likely that affective prosody, i.e. the affect-related changes in voice parameters, involves not just global, uniform shifts across an utterance, but possibly local changes, affecting its constituents differently. The transforms explored in this study consider such possibilities.

The focus here is on practical goals, and on a limited palette of expressive transforms. The aim is to formulate transforms that could potentially be implemented in synthetic utterances generated by our TTS system, so that the output could be modulated towards different affects. This goal stems from a very real need. An Irish language AAC system [16], currently being tested with non-speaking autistic children, uses the ABAIR project's Irish synthetic voices [17]. For AAC users, the synthetic voice is their own voice: it would greatly enhance their ability to express themselves if they could control affect in the spoken output.

In an earlier study [18], voice source parameters (including f_0) were modified in intuitive ways with the aim of conjuring anger and sadness. In the present study, the transforms are guided in large part by detailed voice source analyses of utterances recorded to portray the affects *happy, angry, sad, bored, relaxed* as well as a *neutral* rendition, produced by a male speaker, whose voice features as one of the Irish synthetic voices. As in the earlier study, the transforms carried out involve global, local and a combination of global and local manipulations to voice source parameters. Tempo and formant manipulations were included in two of the transforms. To summarise, this experiment addresses three questions:

(i) How effective are these transforms in altering the perceived affect towards the intended affective targets?

(ii) Which of the three transform types is the most effective? We hypothesise that global and local effects are additive and that the combined transforms are the most effective.

(iii) How uniquely are individual affects signalled, i.e., differentiated among low-activation (or high-activation) affects?

While the present focus is on affective prosody, and on a very specific application, the methods adopted here should also enable a wider exploration of both linguistic and affective prosody and of how these two aspects interact.

10.21437/SpeechProsody.2024-61

2. Methodology

Recordings were made of a male native speaker of Kerry (Dingle) Irish producing a number of semantically neutral Irish sentences. Each sentence was read to portray the affects *angry*, *happy*, *sad*, *bored*, and *relaxed* as well as *neutral*, with multiple repetitions of each affective rendering. The recordings were carried out in a semi-anechoic studio using high-quality equipment (a Brüel & Kjær 4191 microphone with 2669-C pre-amplifier and Nexus 2690 conditioning amplifier) to ensure phase linearity, as required for the voice source analysis. The sampling frequency was 44.1 kHz, but the data selected for analysis were downsampled to 10 kHz.

Based on an informal listening test, the sentence set which best conveyed the target affects overall was chosen and, among the repetitions, one that was deemed the most successful in conveying the targeted affect was chosen for detailed voice source analysis. The sentence was:

'Geobhaimid bád ar an lá sin'

'Get-future tense-we boat on the day that' (morpheme gloss)

'We will get the boat on that day' (translation)

The accented syllables are '*bád*' and '*lá*'. In the chosen sentences, the intonation contour was consistent with the past description of high falling H*+L tones for this dialect [19, 20], which was also found to be the dominant contour across the recorded sentences.

2.1. Analysis Methods

The selected sentences were analysed through a manual interactive inverse filtering system described in [21, 22]. While time-consuming, the manual approach was used due to its superior accuracy compared to automatic analysis methods. The estimated voice source signal was then parameterised using the Liljencrants-Fant (LF) model [23]. The LF model is a mathematical model of differentiated glottal flow, allowing for extraction and manipulation of voice source parameters. The parameters explored here are f_0 (fundamental frequency), E_e (excitation strength) and R_d, a global waveshape parameter proposed by Fant [24]. E_e is defined as the negative amplitude of the differentiated glottal flow at the main excitation, which occurs at the instant of glottal closure. This amplitude is a major determinant of the overall intensity of the source signal. R_d is defined as $0.11^{-1} f_0 U_p / E_e$, where U_p is the amplitude of the glottal flow pulse. U_p / E_e is the so-called *declination time* of the glottal pulse, T_d. Hence, R_d is T_d normalised to the fundamental period and scaled by the factor 0.11^{-1} so that the numerical value of R_d corresponds to T_d in milliseconds for an f_0 of 110 Hz [24]. R_d is a global parameter in the sense that it captures some of the natural covariation in the LF model parameters. It has been shown to relate to vocal fold tension: a low R_d value generally corresponds to a tense voice, while a high value corresponds to a lax voice [24, 25].

PRAAT [26] was used to segment the sentences, and the segment durations of the non-neutral utterances were then normalised to the segment durations of the neutral utterance using a PRAAT script [27]. Visualisations of parameter tracks helped guide decisions regarding the stylisation of the parameter contours and points at which scaling factors were calculated.

2.2. Parameter Modification and Resynthesis Methods

Parameter modifications and resynthesis were done using the Voice Source Generator (VSG) [28, 29], a system developed for the generation of realistic voice source signals using the MATLAB App Designer [30]. The system takes the LF model parameter data of the analysed neutral (baseline) utterance as the input and, given specified scaling factors of the parameter transforms, generates a new source signal using the aliasing-free version of the LF model [31].

Three different kinds of manipulations were carried out to create the affect-targeted stimuli: global, local, and combined (a combination of global and local changes) and are schematically illustrated in Figure 1. Additionally, all versions of the *sad* and *happy* stimuli included tempo shifts: a 10% increase in duration of the *sad* stimuli, and a 10% decrease in the duration of the *happy* stimuli – prompted by the literature (see review in [2]). Furthermore, the *happy* stimuli included a moderate increase in the formant frequencies of 4% across the utterance, prompted by [32]. These modifications were carried out using the PSOLA algorithm [33] in PRAAT.

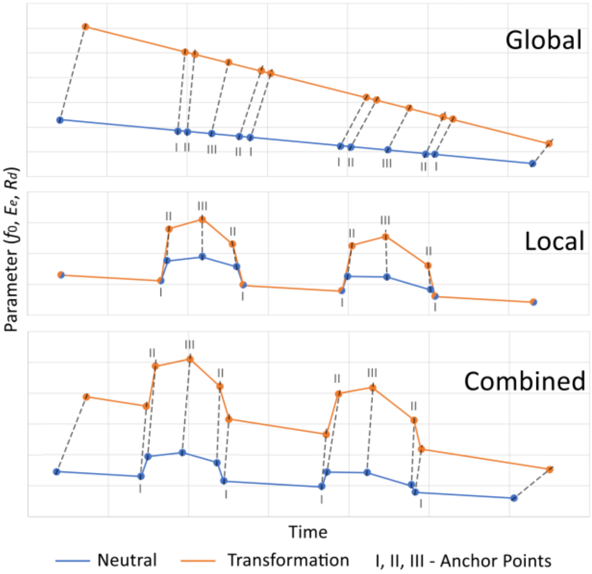

Figure 1: *Schematic representation of the voice transforms. Anchor points I, II and III for the local stimuli were used for the time-normalisation of the global shifts.*

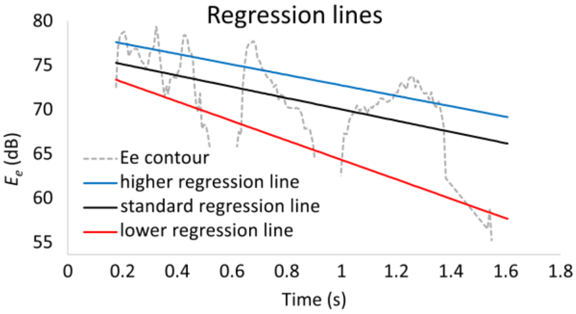

Figure 2: *Example of the different regression lines illustrated for E_e.*

Global transforms: Linear regression analysis was conducted to capture the global characteristics of the source parameter data, as illustrated in Figure 2 for the E_e parameter. Initially, an overall, utterance-level regression line was obtained for each parameter. Similar to the approach used in [34], a secondary linear regression analysis was then conducted using only the data points either below (for f_0 and E_e) or above (for

R_d) the overall regression line, as this analysis would be less affected by the local contours associated with the accented syllables, and therefore potentially provide better representations of the global voice source characteristics.

For f_0 and E_e the data points below the overall regression line were used, as one would expect values to increase during the accented syllables. The opposite is the case for R_d: one would expect a lower R_d value during the accented syllables, as R_d tends to be negatively correlated with vocal tension. Thus, for R_d, the data points above the overall regression line were used for the secondary regression analysis.

For the global transformations, the VSG calculates the scaling factors by dividing the values derived from the secondary regression lines for the targeted affect by the corresponding values of the secondary regression lines of the neutral utterance. These scaling factors are then applied to the neutral utterance to yield the global transform. Because the neutral and target utterances have different durations, anchor timepoints are taken at the start and end of the utterance to ensure time-alignment of the regression lines.

Local transforms: For the local shifts, an anchor-point system was used to extract measurements and implement modifications. Based on phonetic-level segmentations, these anchor points were: (i) the first and last voiced pulses of the vowel in the accented syllables (‘*bá̲d*’ and ‘*lá̲*’); (ii) pulses located 20% into the vowel from the anchor points in (i); and (iii) the middle point of each accented vowel, i.e., the pulse at the midpoint between the anchor points of (i). Local contours were stylised by linear interpolation of the parameter values between the anchor points (see Figure 1).

The local scaling factors were calculated for each parameter by dividing the interpolated parameter values by the corresponding values of the secondary regression line. The local transforms were then achieved by multiplying the scaling factors by the corresponding value from the secondary regression line of the neutral utterance.

Combined transforms: The combined stimuli bring together the global and local shifts, with the scaling factors of the local changes applied directly to the secondary regression line of the target affect after the global shift is applied.

Table 1: *For each affect, the mean change in the parameters for the voice transforms.*

Affect	**f_0 diff (semitones)**			**E_e diff (dB)**			**R_d diff (%)**		
	global	local	combined	global	local	combined	global	local	combined
angry	2.4	1.4	3.8	6.0	1.5	7.4	-14%	-5%	-18%
happy	4.1	4.6	8.7	4.1	2.6	6.7	-1%	4%	3%
bored	-2.0	0.0	-1.9	-3.4	-2.7	-6.1	0%	19%	20%
sad	-1.0	0.2	-0.8	-1.3	-1.3	-2.5	8%	13%	21%
relaxed	1.0	0.1	1.1	0.4	-0.3	0.1	16%	13%	30%

Table 1 shows the change in the mean values of the voice source parameters for the accented syllable *bád* for local, global and combined manipulations.

2.3. Perception Tests

Each of the target affects (*angry, happy, sad, bored, relaxed*) were presented in three versions: global, local and combined. The neutral baseline stimulus, resynthesised through the VSG system, was also included, yielding a total of 16 unique stimuli. The test involved 23 participants, all speakers of Irish. They first heard the full set of stimuli, after which 6 tests were carried out where all 16 stimuli were presented in random order. Listeners were asked to select the expressed emotion from a list that included all affects and a ‘no affect’ option. If an affect was chosen, listeners were then asked to rate how strong the emotion was deemed to be on a 5-point scale from weak (1) to strong (5). Instructions were given in Irish, and responses to the first test set were discarded as it was treated as a practice run to familiarise listeners with the task.

3. Results and Discussion

Each panel in Figure 3 presents results for the triad of stimuli (local, global and combined) targeting a given affect. For each of these, the percentage of responses obtained for the targeted affect are shown alongside responses for the other possible affect choices. The black dotted lines show the percentage of responses to the neutral stimulus for each possible choice of affect. Regarding the perceived strength of affect judged by the listeners, results were little differentiated: averages ranged from 1.8 to 2.75 (from a potential range of 1 to 5). Therefore, in the following, we focus on the identification rate with which specific affects were chosen for different stimuli.

To address our research questions, one-way repeated measures ANOVA tests examined: (i) which of the affect-targeting stimuli yielded ratings for that affect that were significantly higher than those for the neutral stimulus (significant cases are marked with an asterisk in Figure 3); (ii) whether ratings for the three stimulus types for a targeted affect are significantly different from each other, and (iii) how different were ratings among pairs of high-activation (*happy, angry*) or low-activation (*sad, bored*) affects. The significance level was set to 0.05.

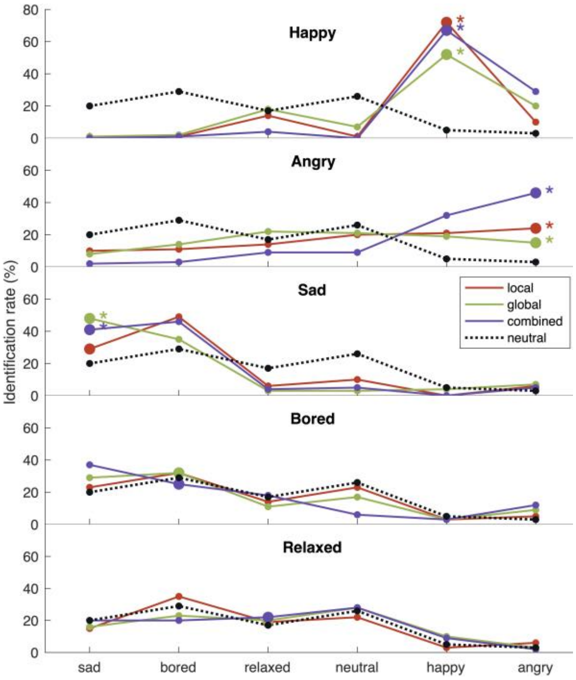

Figure 3: *Identification rates for each triad of affect-targeting stimuli: global (green), local (orange) and combined (purple) for both targeted and other affects. Asterisks mark values significantly higher than values obtained for the neutral stimulus (black dotted lines).*

The high-activation states (*happy, angry*) were more effectively cued than low-activation states (*sad, bored, relaxed*). The *happy* stimuli yielded the highest ratings: 72% for the local, 67% for the combined, and 54% for the global stimulus. All ratings were significantly higher than those for the neutral stimulus. The high ratings for the *happy* stimuli were somewhat surprising, as happiness has been notoriously elusive to cue in perception. As these happy-targeting stimuli involve voice source features as well as tempo and formant shifts, one cannot draw inferences concerning which manipulations were the most effective in conveying happiness to listeners.

The cueing of *happy* also appears to be rather unambiguous, as a comparison with *angry* reveals. Particularly for the most highly rated local stimulus, ratings for *angry*, at 10%, are 62 percentage-points lower than for *happy*. The difference in *happy* vs. *angry* responses is less striking for the combined and global stimuli, but in all cases these differences are statistically significant, and we conclude that the signalling of *happy* is quite specific to this affect. Of the three stimulus types, the local stimulus is clearly the most effective, both in terms of the frequency and uniqueness with which *happy* is identified. As the addition of global modifications adds nothing to the local, the initial hypothesis that the combined transforms would be the most effective is not supported here.

The *angry* affect appears to be effectively signalled, even if ratings are much lower than those obtained for *happy*. The combined stimulus is the most effective, with a rating of 46%, while the other two angry-targeting stimuli achieve relatively poorer ratings (24% for local, 15% for global). Nonetheless, ratings for all three stimuli are significantly higher than for the *neutral* stimulus. Here, it does appear that the perceptual effects of local and global manipulations are additive: while individually relatively ineffective, combining them yields a relatively strong effect. The findings for *angry* replicate the findings of the earlier experiment for *angry* [18]. Unlike the *happy*-targeted stimuli, the *angry*-targeted stimuli are not affect-specific: in no case are *angry* and *happy* responses significantly different.

Overall, ratings for the low-activation affects were poorer and more ambiguous than for the high-activation states. The *sad* affect achieves relatively strong signalling. Ratings of 48% for the global stimulus and 41% for the combined stimulus are significantly higher than those for the *neutral* stimulus. Here, we find no support for the initial hypothesis that the combined stimuli would be the most effective: the addition of local changes reduces rather than enhances the perception of *sad*. Furthermore, the affect cueing of the *sad*-targeted stimuli is not specific to the *sad* affect. For none of the three *sad*-targeted stimuli are responses for *sad* significantly higher than those for *bored*. In fact, *bored* emerges as a more frequent response to both the local (49%) and the combined (46%) stimuli.

Ratings for the *bored*-targeted stimuli are low, and none are significantly higher than the ratings for the neutral stimulus. But, as just noted, a *bored* affect was in fact strongly associated with the local and combined stimuli targeting *sad*. And in contrast to the case of the *sad* affect, where the global stimulus was the most highly rated and the addition of local modifications reduced the *sad* effect, the local *sad*-targeting stimulus appears to shift perceptions towards *bored*, and the addition of global changes reduced rather than enhanced the *bored* effect. Furthermore, for the local *sad*-targeting stimulus, the difference in *bored* and *sad* responses is statistically significant, suggesting this stimulus achieves relatively unambiguous cueing of *bored*.

The relaxed-targeted stimuli were ineffective. Regardless of stimulus type, ratings are never significantly different from ratings for the neutral stimulus.

The results for the low-activation states are confounded by the fact that the *neutral* baseline emerges as not affectively neutral, but rather skewed towards the low-activation states. Thus, even where, in absolute terms, rather high *sad* (or *bored*) ratings are achieved with *sad*-targeting stimuli, the relative shift from *neutral* is considerably smaller. The present transforms are thus likely to underestimate what is needed to change a truly affect-neutral utterance towards a *sad* or *bored* rendition.

4. Conclusions

The results of this experiment are somewhat mixed. The transforms for the high activation affects *happy* and *angry* are promising. The cueing of *happy* is much more effective than *angry* – both in terms of the frequency of identification and the relative uniqueness with which this affect is cued – particularly with the local stimulus. This is of particular interest, as the *happy* affect has traditionally been found difficult to capture. Given that the *happy* transforms involve not only voice source parameters but also formant and tempo shifts, further experiments are needed to assess the relative contribution of the different factors.

The cueing of the low-activation states was less effective, even if *sad* and *bored* did emerged reasonably strongly, and even if the cueing of *bored* was not in response to the intended transforms. However, for the low-activation affects, our results are somewhat inconclusive, given that the intended neutral stimulus, as perceived, was biased towards the low activation states. A priority for future work will be the establishment of a neutral baseline. Our plan at this juncture is to take a sentence produced by the TTS system for this speaker's voice as the neutral template, and to carry out preliminary tests to establish its affect-status. Informal listening tests suggest it will provide a better starting point for this type of experiment.

A hypothesis that the combined stimuli would be the most effective, and that the effects of the global and local transforms would be additive was not borne out, as it was supported only by the results for the *angry* transforms. For *happy*, the local stimulus was the most effective; for *sad*, the global stimulus emerged as the most effective; for *bored*, the local *sad*-targeting stimulus yielded best results. At this point, we can only speculate that the most effective type of transform will be affect-dependent.

The transforms used here were intentionally kept relatively simple, as they are framed within a goal of allowing practical affect control to the synthetic voices of an Irish-language AAC device. Therefore, they did not involve further voice source features that could be important, such as voice quality settings for creaky and whispery voice, found in earlier studies to be relevant in signalling low-activation states [6, 7]. Broadening the scope of these affect-transforms to include additional source adjustments would also be worth exploring in future work.

As mentioned in the introduction, it is hoped that the approach adopted here will not only yield insights into affective prosody, but also allow exploration of linguistic prosody and interaction of affective and linguistic prosody – as a step towards a fuller, more holistic understanding of the role of prosody in speech communication.

5. Acknowledgements

The research conducted in this publication was funded by the Irish Research Council under award number GOIPG/2021/561 to the first author, and through the Róboglór-ABAIR project, awarded by the Dept. of Tourism, Culture, Arts, Gaeltacht, Sports and Media, with funding from the National Lottery, as part of the *Straitéis 20 Bliain don Ghaeilge*.

6. References

[1] R. Banse and K. R. Scherer, “Acoustic profiles in vocal emotion expression,” *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614-636, 1996.

[2] P. N. Juslin and P. Laukka, “Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?,” *Psychological Bulletin*, vol. 129, no. 5, pp. 770-814, 2003.

[3] K. R. Scherer, T. Johnstone and G. Klasmeyer, “Vocal expression of emotion,” in *Handbook of Affective Sciences*, R. J. Davidson, K. R. Scherer and H. H. Goldsmith, Oxford: Oxford University Press, 2003, Ch. 23, pp. 433-456.

[4] K. R. Scherer, “Vocal Affect Expression: A Review and a Model for Future Research,”, *Psychological Bulletin*, vol. 99, no. 2, pp. 143-165, 1986.

[5] S. Mozziconacci, “Speech variability and emotion: production and perception,” Ph.D. dissertation, Institute for Perception Research, Technische Universiteit Eindhoven, Eindhoven, 1998.

[6] C. Gobl and A. Ní Chasaide, “The role of voice quality in communicating emotion, mood and attitude,” *Speech Communication*, vol. 40, no. 1-2, pp. 189-212, 2003.

[7] I. Yanushevskaya, C. Gobl, and A. Ní Chasaide, “Cross-language differences in how voice quality and f0 contours map to affect,” *Journal of the Acoustic. Society of America*, vol. 144, no. 5, pp. 2730-2750, 2018.

[8] C. Gobl, E. Bennett, and A. Ní Chasaide, “Expressive synthesis: how crucial is voice quality?,” in *IEEE Workshop on Speech Synthesis*, Santa Monica, California, USA, paper 53, pp. 1-4, 2002.

[9] M. Heldner, “On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish,” *Journal of Phonetics,* vol. 31, no 1, pp. 39-62, 2003.

[10] M. Heldner, “Spectral emphasis as an additional source of information in accent detection,” in *Proceedings of ITRW on Prosody in Speech Recognition and Understanding*, paper 10, 2001.

[11] M. Heldner, E. Strangert, and T. Deschamps, “A focus detector using overall intensity and high frequency emphasis” in *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 1999),* San Francisco, 1999, pp. 1491-1494.

[12] A. Ní Chasaide, I. Yanushevskaya, J. Kane, and C. Gobl, “The voice prominence hypothesis: the interplay of F0 and voice source features in accentuation,” in *Proceedings of INTERSPEECH 2013*, Lyon, France, 2013, pp. 3527-3531.

[13] I. Yanushevskaya, A. Murphy, C. Gobl and A. Ní Chasaide, “Global waveshape parameter R_d in signaling focal prominence: Perceptual salience in the absence of f0 variation. *Front. Commun.*, 7, pp. 1-23, 2022.

[14] A. Murphy, I. Yanushevskaya, A. Ní Chasaide and C. Gobl, “Voice Source Contribution to Prominence Perception: Rd Implementation,” in *Proceedings of INTERSPEECH 2018*, Hyderabad, India, 2018, pp. 217-221.

[15] A. Ní Chasaide, I. Yanushevskaya and C. Gobl, “Prosody of voice: declination, sentence mode and interaction with prominence,” in *Proceedings of the 18th International Congress of Phonetic Sciences (ISPhS 2015),* Glasgow, United Kingdom, 2015, pp. 1-5.

[16] E. Barnes, J. Cummins, R. Errity, O. Morrin, H. Berthelsen, C. Wendler, A. Murphy, H. Husca, N. Ní Chiaráin, A. Ní Chasaide, “Geabaire, the First Irish AAC System: Voice as a Vehicle for Change,” in *Proceedings of ELRA/ISCA SIG on Under-resourced Languages (SIGUL 2023),* Dublin, Ireland, 2023.

[17] A. Ní Chasaide, A. Ní Chiaráin, H. Berthelsen, A. Murphy, Lonergan, J. Sloan, C. Wendler, C. McCabe, E. Barnes, and C. Gobl, “The language communities as active partners in technology provisions: the Irish ABAIR experience,” in *Proc. 2nd Annual Meeting of the ELRA/ISCA SIG on Under-resourced Languages (SIGUL 2023)*, Dublin, Ireland, 2023, pp. 116-120.

[18] A. Murphy, I. Yanushevskaya, A. Ní Chasaide, C. Gobl, “Affect Expression: Global and Local Control of Voice Source Parameters,” in *Proceedings of Speech Prosody 2022*, Lisbon, Portugal, 2022, pp. 525-529.

[19] M. Dalton, “The Phonetics and Phonology of Intonation in Three Irish Dialects,” unpublished Ph.D. Thesis, Trinity College, Dublin, 2008.

[20] M. Dalton and A. Ní Chasaide, A. (2007). Nuclear accents in four Irish (Gaelic) dialects,” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007),* Saarbrücken, Germany, pp. 965-968, 2007.

[21] A. Ní Chasaide, C. Gobl, and P. Monahan, “A technique for analysing voice quality in pathological and normal speech,” *Journal of Clinical Speech and Language Studies*, vol. 2, no. 1, pp. 1-16, 1992.

[22] C. Gobl and A. Ní Chasaide, “Voice source variation and its communicative functions,” in *Handbook of Phonetic Sciences*, 2nd ed., W. J. Hardcastle, J. Laver and F. E. Gibbon, Oxford: Blackwell, pp. 378-425, 2010.

[23] G. Fant, J. Liljencrants, and Q. Lin, “A four-parameter model of glottal flow,” *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 4, pp. 1-13, 1985.

[24] G. Fant, “The LF-model revisited. Transformations and frequency domain analysis,” *STL-QPSR*, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, vol. 36 (2-3), pp. 119-156, 1995.

[25] G. Fant, “The voice source in connected speech,” *Speech Communication*, vol. 22, pp. 125-139, 1997.

[26] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [Computer Program], Version 6.4.01, 2023.

[27] A. E. Rodgers, plugin_drawimportedParameters (Ver. 1.2.0) [Source Code]. https://github.com/AERodgers/plugindraw ImportedParameters#readme-ov-file, 2022.

[28] Z. Wang and C. Gobl, “Perceptual effects of aliasing distortion in glottal flow modelling,” in *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, Prague, Czech Republic, 2023, pp. 1801-1805.

[29] Z. Wang and C. Gobl, “A System for Generating Voice Source Signals that Implements the Transformed LF-model Parameter Control,” in *Proceedings of INTERSPEECH* 2023, Dublin, Ireland, 2023, pp. 4738-4742.

[30] MATLAB, Version 9.12.0.2170939 (R2022a), Natick, Massachusetts: The Mathworks Inc., 2022.

[31] C. Gobl, “The LF Model in the Frequency Domain for Glottal Airflow Modelling without Aliasing Distortion,” in *Proceedings of INTERSPEECH 2021,* Brno, Czech Republic, 2021, pp. 2651-2655.

[32] I. Torre, “Production and perception of smiling voice,” in *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2013)*, York, United Kingdom, 2014, pp. 100-117.

[33] E. Moulines and F. Charpentier, “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,” *Speech Communication*, vol. 9, no. 5-6, pp. 453-467, 1990.

[34] J. Haan, V. J. Van Heuven, J. Pacilly, R. Van Bezooijen, “An anatomy of Dutch question intonation,” *Linguistics in the Netherlands*, vol. 14, pp. 97-108. 1997