



Towards a comprehensive assessment of speech intelligibility for pathological speech

W. Xue¹, V. Mendoza Ramos², W. Harmsen³, C. Cucchiari¹, R.W.N.M. van Hout³, H. Strik^{1,3,4}

¹ Centre for Language and Speech Technology (CLST), Radboud University Nijmegen

² Department of Otorhinolaryngology, Head and Neck Surgery and Communication Disorders,
University Hospital of Antwerp

³ Centre for Language Studies (CLS), Radboud University Nijmegen

⁴ Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

w.xue@let.ru.nl, Viviana.MendozaRamos@uza.be, w.harmsen@student.ru.nl,
c.cucchiari@let.ru.nl, r.vanhout@let.ru.nl, w.strik@let.ru.nl

Abstract

Speech intelligibility is an essential though complex construct in speech pathology. It is affected by multiple contextual variables and it is often measured in different ways. In this paper, we evaluate various measures of speech intelligibility based on orthographic transcriptions, with respect to their reliability and validity. For this study, different speech tasks were analyzed together with their respective perceptual ratings assigned by five experienced speech-language pathologists: a Visual Analogue Scale (VAS) and two types of orthographic transcriptions, one in terms of existing words and the other in terms of perceived segments, including nonsense words. Six subword measures concerning graphemes and phonemes were derived automatically from these transcriptions. All measures exhibit high degrees of reliability. Correlations between the six subword measures and three independent measures, VAS, word accuracy, and severity level, reveal that the measures extracted automatically from the orthographic transcriptions are valid predictors of speech intelligibility. The results also indicate differences between the speech tasks, suggesting that a comprehensive assessment of speech intelligibility requires materials from different speech tasks in combination with measures at different granularity levels: utterance, word, and subword. We discuss these results in relation to those of previous research and suggest possible avenues for future research.

Index Terms: speech intelligibility, dysarthric speech, speech therapy, computational paralinguistic

1. Introduction

Speech disorders in general, and dysarthria especially, lead to decreased speech intelligibility. This can have a severe impact on the patients' quality of life because they can lose social contact and eventually become isolated from society. Most of the time, degraded speech intelligibility can be improved through speech therapy. However, the effects of intensive therapy are not always evident. For monitoring a possible evolution, pre- and post-therapy evaluations in which intelligibility scores play an important role, are necessary. Thus, intelligibility requires a clear definition and a robust operationalization.

A clear definition has been proposed by Hustad [1] "Intelligibility refers to how well a speaker's acoustic signal can be accurately recovered by a listener". In line with this definition, intelligibility can be measured in various ways. One

of them is based on orthographic transcriptions of sentences, words or phonemes [2,3,4]. The percentage of words or phonemes correctly identified is employed as a measure of intelligibility and it is used in the Sentence Intelligibility Test [4]. Besides, intelligibility has also been measured by collecting scalar ratings from human judges [7, 8, 9] through an equal-appearing interval scales like the Likert scale [9], or by placing a point on a horizontal line like the visual analogue scale (VAS) [10].

All the methods described above rely on perceptual judgments. It is common practice to collect different measures from multiple judges and check their reliability before the usage for future research purposes [5, 6, 11], since these measures can be influenced by, for example, the type of judge or listener. As pointed out in [18], measures collected from inexperienced ('lay' or 'naive') listeners showed larger variances than those collected from well-trained expert judges such as speech-language therapists [18]. In addition, the measures can also be influenced by the judges' familiarity with the speech tasks. Specifically, Beukelman and Yorkston [19] reported that the estimates of speech intelligibility increased as the judges became familiar with the reading passage. Given that all these operations are time-consuming, costly and laborious in practice, there is a need for obtaining valid measures of speech intelligibility in a more objective way.

In line with this need, several studies have investigated the relationship between perceptual ratings of intelligibility and various automatically calculated measures. These automated measures are normally obtained using computational models such as Automatic Speech Recognition (ASR) [22, 23, 30] or neural networks [16, 24–28]. Very high correlations have been reported. For example, the magnitude of correlations between ASR outputs (e.g. word accuracy and word error rate) and intelligibility ratings measured by a 5-point Likert scale reached 0.9 for patients with cancer of oral cavity [22] and 0.92 for children with cleft lip [23]. Typical models can also be found in [24–28] and the performance is comparable to the results presented above [16]. In general, these measures are not detailed enough to be used by therapists to diagnose the problems that led to decreased intelligibility.

For this reason, a semi-automatic approach was proposed in [11]. In the study, a set of intelligibility ratings of disordered speech assigned by lay listeners were investigated to obtain measures at three different levels of granularity: utterance, word, and subword level including grapheme and phoneme levels. Utterance level evaluations were obtained using subjective

rating scales (i.e. VAS and Likert scale) while word and subword level evaluations, i.e. distance scores, were obtained automatically from human-generated orthographic transcriptions using automatic alignment and grapheme-phoneme conversion algorithms. The results indicated that the distance measure at the phoneme level was feasible and reliable, and it was a more sensitive measure to changes within patients, thus providing an informative measure of intelligibility.

In the present paper, we extend this semi-automatic approach based on orthographic transcriptions and its automatically derived metrics as measures of pathological speech intelligibility on a number of important points. First, we collect measures for a larger number of samples, including both pathological and normal, covering different speech tasks. Intelligibility measures are explored in relation to speaker types and speech tasks. Second, we collect ratings from experts as opposed to lay listeners. Third, we ask the raters to provide two types of transcriptions, one in terms of existing words and the other in terms of literal or perceived segments. More detailed measures are calculated automatically from them. Fourth, we evaluate the reliability and validity of measures obtained from transcriptions in relation to other measures such as VAS and severity level of dysarthria.

In the remainder of the paper, we first describe the experimental design and explain how measures of speech intelligibility were computed at different levels of granularity in Section 2. In Section 3, we present the results and in Section 4, we discuss our findings and present ideas for future research.

2. Method

2.1. Experimental design

2.1.1. Speakers and speech materials

The speech material used for this study covers different speech tasks that are often used in clinical practice to perceptually assess speech intelligibility. The speakers were selected to cover different types of dysarthria and severity levels (SL), at different ages and were gender-balanced.

The investigation comprised three experiments. Experiment 1 included 36 speakers (10 control and 26 speakers with dysarthria). For each speaker, the same four sentences were selected from the Dutch phonetically balanced text ‘Papa en Marloes’. In Experiment 2, 18 speakers (4 control and 14 speakers with dysarthria) read the 50 existing and non-existing consonant-vowel-consonant words of the Dutch Intelligibility Assessment (DIA) task [13]. The recordings used in the first two experiments were selected from the Dutch Corpus of Pathological and Normal Speech (COPAS) [14]. In Experiment 3, 23 speakers with dysarthria were involved and for each speaker six Semantically Unpredictable Sentences (SUS) with different lengths were selected from the Dutch Sentence Intelligibility Assessment [13].

The severity levels of the speakers had been assigned by speech-language pathologists on a four-category scale (normal-mild-moderate-severe), and they were already available in the speech corpora. Specifically, for Experiments 1 and 2, the ‘normal’ refers to the healthy control speakers while for Experiment 3, the ‘normal’ category includes two speakers with mild dysarthria whose speech was classified as clear as control speakers. The frequency plot of the severity level of speakers in all three experiments is shown in Figure 1.

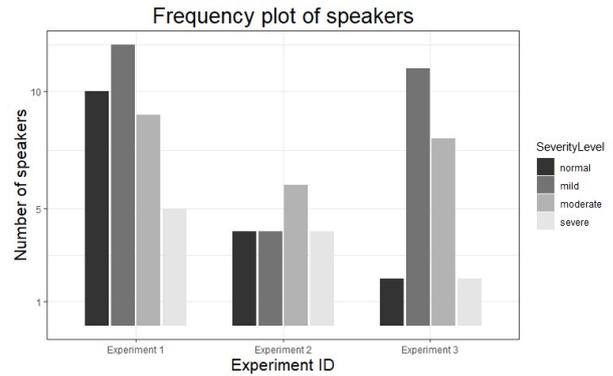


Figure 1: Frequency plot of the severity level of speakers in all three experiments.

2.1.2. Procedure and raters

Five expert listeners perceptually judged intelligibility. They were not familiar with the materials used in Experiments 2 and 3. However, they were familiar with the sentences in Experiment 1 since these sentences are part of a phonetically balanced text commonly used in their clinical practice.

For each recording, the experts perceptually rated intelligibility on the utterance level through a Visual Analogue Scale (VAS) and made two types of orthographic transcriptions which are Word – using only meaningful words, and Literal – also allowing nonsense words, to reflect the speech segments the experts perceived. Since Experiment 2 contains nonsense words, they only made Literal transcriptions. They were allowed to listen only once to the sample and subsequently score it (i.e. assigning a VAS score and making a Literal transcription). In the other two experiments, they were allowed to listen twice to each sample to perform the task (i.e. assigning a VAS score and making two types of transcriptions). Limiting the number of listening times was done to reduce the eventual impact of familiarity. The samples were also randomized to prevent any systematic order effect.

2.2. Intelligibility measures

Intelligibility measures at different levels of granularity were collected and calculated. The same eight intelligibility measures at the utterance, word and subword level were obtained in all three experiments.

2.2.1. Intelligibility measures at speaker and utterance level

The severity levels of dysarthria were used as speaker-level measures. Utterance level intelligibility ratings were obtained using VAS ranging from 0 to 100. Expert listeners were instructed that 0 represented “not intelligible” and 100 “intelligible”.

2.2.2. Intelligibility measures at the word level

The orthographic transcriptions were compared to the reference transcriptions after removing punctuation and symbols indicating missing words. Note that for this measure, nonsense words were treated as words. Based on these transcriptions, the accuracy (Acc) of words was computed as follows:

$$Acc = (N_{total} - N_d - N_s) / N_{total} \times 100 \quad (1)$$

where N_{total} denotes the total number of words in the reference transcriptions, and N_d and N_s denote the number of deletions

Table 1: Mean and standard deviations of eight different measures in our three experiments. For the utterance level, we use Visual Analogue Scales (VAS). For the word (W) level, we use accuracy (Acc). For the grapheme (G) and phoneme (P) level, we use Acc as well as Distance (Dist) and number of Changes (Ch). Specifically, for the last three levels, we have two types of transcription in Experiment 1 and Experiment 3: Literal (left part of the cell) and Word (right part of the cell).

Levels	Measures	Experiment1		Experiment 2	Experiment 3	
Utterance	VAS	84.25 (25.52)		68.12 (25.29)	79.9 (24.48)	
Word	Acc-W	79.99 (29.31)	88.09 (24.50)	41.68 (24.52)	71.27 (29.01)	76.74 (29.00)
Grapheme	Acc-G	92.29 (15.85)	95.15 (13.03)	80.34 (12.65)	88.86 (15.53)	89.02 (16.90)
	Dist-G	4.78 (10.13)	3.24 (8.70)	30.28 (19.35)	6.89 (9.72)	6.82 (10.39)
	Ch-G	3.89 (8.10)	2.65 (6.87)	18.77 (12.08)	5.11 (7.31)	4.98 (7.74)
Phoneme	Acc-P	91.13 (17.73)	94.47 (14.68)	77.75 (13.92)	87.73 (17.16)	88.14 (18.27)
	Dist-P	10.59 (21.58)	7.17 (18.27)	37.64 (26.25)	12.28 (18.12)	11.86 (19.12)
	Ch-P	3.81 (7.76)	2.56 (6.62)	16.39 (10.58)	4.69 (6.82)	4.53 (7.19)

and substitutions in the corresponding orthographic transcriptions, respectively.

2.2.3. Intelligibility measures at subword level

At subword level (grapheme and phoneme), in addition to Acc, another two measures, the distance (*Dist*) between two transcriptions and the number of changes (*Ch*), were extracted as follows:

$$Dist = N_d \times C_d + N_i \times C_i + N_s \times C_s \quad (2)$$

$$Ch = N_d + N_i + N_s \quad (3)$$

where N_d , N_i and N_s denote the number of deletions, insertions and substitutions, respectively. C_d , C_i and C_s denote their corresponding costs. For grapheme, $C_d = 1$, $C_i = 1$ and $C_s = 2$. For phonemes, $C_d = 3$, $C_i = 3$ and C_s was calculated by employing matrices with articulatory features [15, 21]. These measures were calculated by the software ADAPT. More details about ADAPT, the extracting procedures and the grapheme to phoneme conversion can be found in [11, 15].

3. Results

In this section, we present the experimental results of the different measures concerning reliability (Section 3.1), mean and standard deviations (Section 3.2), and correlations (Section 3.3).

3.1. Reliability of measures

The reliability of all eight measures in the three experiments was calculated using ICC (2, k) (items and raters random with $k = 5$) in psych package in R [17] since all the raters assigned judgments to all the recordings from all the speakers.

The analyses reveal that for the eight measures, the reliability coefficients for the five listeners together are generally very high, above 0.90, except for a relatively lower reliability (0.82) for accuracy at the word level (Acc-W) in Experiment 1 in Word transcription.

3.2. Mean and standard deviation of measures

The mean and standard deviation (SD) of the intelligibility measures are shown in Table 1. For Experiment 2, we observed worse intelligibility, probably because of the isolated nonsense words. Experiments 1 and 3 show better scores. Moreover, the scores for Experiment 1 are higher compared to Experiment 3. Possible explanations are that Experiment 1 has a higher num-

Table 2: Correlations between VAS, Accuracy of word (Acc-W) and speakers' Severity Level of dysarthria (SL) in our three experiments. For Experiments 1 and 3, we have two types of transcription: Literal (left part of the cell) and Word (right part of the cell). The Multiple R was computed when SL was involved, treating SL as a nominal variable.

Correlation	VAS vs Acc-W	VAS vs SL	Acc-W vs SL
Experiment 1	0.85 / 0.73	0.70	0.60 / 0.70
Experiment 2	0.75	0.69	0.70
Experiment 3	0.71 / 0.74	0.51	0.46 / 0.48

ber of normal speakers than Experiment 3, and the latter used semantically unpredictable utterances. Besides, in both Experiments 1 and 3, better intelligibility scores are observed for Word than for Literal transcription.

3.3. Intelligibility measures and correlations

The median values of the correlations between each pair of the subword-level measures turned out to be always higher than 0.9, with a lowest minimum correlation of 0.87.

Table 2 shows the correlations between SL at the speaker level, VAS at the utterance level and Acc at the word level, the three measures we want to use as criterion variables for the six subword-level measures. We can observe moderate to strong correlations between the three measures. Probably these three measures reflect different aspects of intelligibility, and thus, they constitute an interesting combination to investigate the validity of the subword-level measures. Besides, the SL correlations do not reflect intra-speaker variation with the consequence that these correlations are systematically lower than the other correlations.

Table 3 gives the correlations of the six subword-level measures and our three criterion variables mentioned above. Table 3 shows that the Acc measures (17 times the highest correlation) outperform the Dist and Ch measures (8 times the highest correlation). Within the Acc measures, Acc at phoneme level (Acc-P) performs better (13 times) than Acc at grapheme level (Acc-G) (4 times). We also looked at non-parametric Spearman correlations, but that did not change the overall strength of the correlations nor the overall correlational pattern.

Table 3: Correlations between measures at subword level and measures at utterance, word and speaker level. For Experiments 1 and 3, we have two types of transcription: Literal (left part of the cell) and Word (right part of the cell). The Multiple R was computed when SL was involved, treating SL as a nominal variable. For each row, the bold numbers are the highest ones in each type of transcription (Literal / Word).

Correlation			Grapheme			Phoneme		
			Dist	Ch	Acc	Dist	Ch	Acc
Experiment 1	Utterance	VAS	0.83 / 0.75	0.81 / 0.73	0.88 / 0.78	0.81 / 0.72	0.83 / 0.74	0.89 / 0.79
	Word	Acc-W	0.77 / 0.79	0.77 / 0.79	0.88 / 0.89	0.78 / 0.79	0.78 / 0.80	0.89 / 0.90
	Speaker	SL	0.58 / 0.52	0.56 / 0.51	0.59 / 0.52	0.56 / 0.50	0.57 / 0.51	0.60 / 0.53
Experiment 2	Utterance	VAS	0.78	0.78	0.79	0.76	0.77	0.79
	Word	Acc-W	0.90	0.90	0.90	0.83	0.87	0.90
	Speaker	SL	0.69	0.68	0.70	0.67	0.69	0.71
Experiment 3	Utterance	VAS	0.72 / 0.72	0.72 / 0.72	0.73 / 0.75	0.71 / 0.71	0.72 / 0.72	0.73 / 0.75
	Word	Acc-W	0.81 / 0.86	0.80 / 0.85	0.88 / 0.91	0.79 / 0.84	0.80 / 0.85	0.89 / 0.92
	Speaker	SL	0.55 / 0.55	0.56 / 0.55	0.43 / 0.46	0.55 / 0.55	0.56 / 0.55	0.45 / 0.47

4. Discussion and conclusions

In this paper, we have conducted an extended evaluation of a semi-automatic approach to measuring the intelligibility of dysarthric speech in which we have investigated the reliability and validity of several descriptors of intelligibility for normal and pathological speech. The analyses reveal that for the eight measures we acquired, the reliability coefficients were very high in the different experiments. This supports the usability of these measures, in particular, because it means that a limited number of raters might be sufficient to obtain highly reliable ratings, which is of course very important in a clinical setting. The measures used show differences between the two types of transcription, word and literal, with the latter displaying more variability, as would be expected in a non-lexically driven context. As a consequence, all measures indicating accuracy at the various granularity levels are lower in the literal condition than in the word condition. All the measures appear to be sensitive to the different severity levels of the speakers with dysarthria.

We were interested in whether we could derive more detailed information from the expert transcriptions than merely computing the percent accuracy at the word level, as is often done [see e.g. 11]. We computed six additional measures at the subword level, three based on graphemes and three based on phonemes. In both cases, we computed Accuracy (Acc), distance (Dist) and the number of changes (Ch).

At the subword level, the mean values of Acc were very similar to Ch, with the phoneme-level results always slightly lower than those at the grapheme level. This is understandable because a phoneme may be associated with more than one grapheme and then its overall correctness requires correctness in its associated graphemes. The six automatically calculated subword-level measures are strongly correlated with each other, which could be explained by the fact that they are all based on the same orthographic transcriptions. However, it is worth noting that they are strongly related and that it does not make much difference using one or the other. For instance, in terms of use in clinical practice, a grapheme-level measure may be easier to apply than a phoneme-level one, but both will yield accurate results.

The results also showed that these orthography-based measures are strongly correlated with an independent measure, the VAS ratings, which are based on the listeners' perceptual judgments. To test the external validity of the subword-level

measures, we included the accuracy at the word level and the dysarthria severity level at the speaker level as additional evaluation criteria. The correlations between the above three measures are moderate to strong (see Table 2), presumably showing that evaluative components involved in estimating intelligibility are different.

Correlations between these three measures and the six subword-level measures (see Table 3) indicate that the phoneme measures outperformed the grapheme measures and that the best phoneme measure seems to be accuracy. This suggests that the orthography-based subword-level measures investigated are not only reliable indicators of speech intelligibility, but that they can also be considered as valid descriptors of speech intelligibility in pathological speech.

These results show the possibility of using orthographic transcriptions and the automatically derived phoneme measures to determine which mispronounced phonemes cause decreased speech intelligibility. In other words, these measures have potentially additional diagnostic value and can, therefore, be applied in speech therapy. Compared with other automatic measures [22, 23], these measures can provide more detailed information. In addition, they can be easily obtained from orthographic transcriptions without time-consuming human annotations at the phoneme level.

Future work will explore the possibility to fully automate intelligibility evaluation without any human-generated orthographic transcriptions. This could be achieved with the help of Automatic Speech Recognition (ASR) technology and the increasing availability of dysarthric speech data [12, 29, 31, 32]. Another option for prompted speech would be to use ASR in forced alignment mode, which is one of the methods we intend to investigate in future research.

5. Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 766287. We would like to thank our colleagues from the Department of Otorhinolaryngology, Head and Neck Surgery and Communication Disorders of the University Hospital of Antwerp, Gwen Van Nuffelen and Marc De Bodt, for their cooperation and feedback.

6. References

- [1] K. C. Hustad, "The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers with Dysarthria," *J. Speech Lang. Hear. Res.*, vol. 51, no. 3, 2008, pp. 562.
- [2] D. Kempler, D. Van Lancker, "Effect of speech task on intelligibility in dysarthria: a case study of Parkinson's Disease," *Brain Lang.*, vol. 80, 2002, pp.:449-64.
- [3] K. M. Yorkston, D. R. Beukelman, "A comparison of techniques for measuring intelligibility of dysarthric speech," *Journal Communication Disorders*, vol. 11, 1978, pp.:499-512.
- [4] K. Yorkston, D. R. Beukelman, and R. Tice, "Sentence intelligibility test [Measurement instrument]," *Lincoln, NE: Tice Technologies*, 1996.
- [5] K. C. Hustad, "Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy," *Folia Phoniatrica et Logopaedica*, vol. 59, no. 6, 2007, pp.: 306-317.
- [6] K. L. Stipancic, K. Tjaden, and G. Wilding, "Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 2, 2016, pp.: 230-238.
- [7] S. S. Barreto and K. Z. Ortiz, "Intelligibility measurements in speech disorders: a critical review of the literature," *Pró-Fono Revista de Atualização Científica*, vol. 20, no. 3, 2008, pp.: 201–206.
- [8] N. Miller, "Measuring up to speech intelligibility," *International Journal of Language & Communication Disorders*, vol. 48, no. 6, 2013, pp.: 601–612.
- [9] K. M. Yorkston and D. R. Beukelman, "A comparison of techniques for measuring intelligibility of dysarthric speech," *Journal of Communication Disorders*, vol. 11, 1978, pp.: 499–512.
- [10] C. Finizia, J. Lindstrom, and H. Dotevall, "Intelligibility and perceptual ratings after treatment for laryngeal cancer: laryngectomy versus radiotherapy," *Laryngoscope*, vol. 108, no. 1, 1998, pp.: 138–143.
- [11] M. Ganzeboom, M. Bakker, C. Cucchiarini, and H. Strik, "Intelligibility of Disordered Speech: Global and Detailed Scores," in *INTERSPEECH*, 2006, pp.: 2503–2507.
- [12] E. Yilmaz, M. Ganzeboom, C. Cucchiarini and H. Strik, "Combining Non-pathological Data of Different Language Varieties to Improve DNN-HMM Performance on Pathological Speech," in *INTERSPEECH*, 2016, pp.: 218-222.
- [13] M. de Bodt, C. Guns, and G. van Nuffelen, "NSVO: handleiding. Vlaamse Vereniging voor Logopedie: Herentals," 2006.
- [14] C. Middag, "Automatic analysis of pathological speech," dissertation, Ghent University, 2012.
- [15] B. Elffers, C. van Bael, and H. Strik, "ADAPT: Algorithm for Dynamic Alignment Transcriptions," *Internal report. Nijmegen, Netherlands: Department of Language & Speech*, University of Nijmegen, 2013.
- [16] G. Van Nuffelen, C. Middag, M. De Bodt, and J. Martens, "Speech technology-based assessment of phoneme intelligibility in dysarthria," *International journal of language & communication disorders*, vol. 44, no. 5, 2009, pp.: 716-730.
- [17] W. Revelle, "psych: Procedures for Psychological, Psychometric, and Personality Research," Northwestern University, Evanston, Illinois. R package version 1.9.12. 2019
- [18] E. O. Mencke, G. J. Ochsner, and E. W. Testut, "Listener judges and the speech intelligibility of deaf children," *Journal of Communication Disorders*, vol. 16, no. 3, 1983, pp.: 175-180.
- [19] D. R. Beukelman, and K. M. Yorkston, "Influence of passage familiarity on intelligibility estimates of dysarthric speech," *Journal of Communication Disorders*, vol. 13, no. 1, 1980, pp.: 33-41.
- [20] J. van de Weijer and I. Slis, "Nasalitymeting met de nasometer," *Logop Foniatr*, vol. 63, 1991, pp.: 97-101.
- [21] C. Cucchiarini, "Assessing transcription agreement: methodological aspects," *Clinical Linguistics & Phonetics*, vol. 2, 1996, pp.:131-155.
- [22] A. Maier, M. Schuster, A. Batliner, E. Nöth, and E. Nkenke, "Automatic scoring of the intelligibility in patients with cancer of the oral cavity," in *Eighth Annual Conference of the International Speech Communication Association*, 2007.
- [23] M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, and E. Nöth, "Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition," *International Journal of Pediatric Otorhinolaryngology*, vol. 70, no. 10, 2006, pp.: 1741-1747.
- [24] C. Middag, G. Van Nuffelen, J. Martens, and M. De Bodt, "Objective intelligibility assessment of pathological speakers," in *9th Annual conference of the International Speech Communication Association (Interspeech)*, 2008, pp.: 1745-1748.
- [25] G. Van Nuffelen, M. De Bodt, J. Vanderwegen, P. Van de Heyning, and F. Wuyts, "Effect of rate control on speech production and intelligibility in dysarthria," *Folia Phoniatrica et Logopaedica*, vol. 62, no. 3, 2010, pp.: 110-119.
- [26] C. Middag, T. Bocklet, J. Martens, and E. Nöth, "Combining phonological and acoustic ASR-free features for pathological speech intelligibility assessment," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [27] C. Middag, J. Martens, G. Van Nuffelen, and M. De Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Advances in Signal Processing*, no. 1, 2009.
- [28] C. Middag, Y. Saeys, and J. Martens, "Towards an ASR-free objective analysis of pathological speech," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.
- [29] E. Yilmaz, M. Ganzeboom, C. Cucchiarini and H. Strik, "Multi-stage DNN training for Automatic Recognition of Dysarthric Speech," in *Proc. INTERSPEECH*, 2017, pp.: 2685-2689.
- [30] D. Le, K. Licata, C. Persad, E. M. Provost, "Automatic assessment of speech intelligibility for individuals with aphasia," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 24, no. 11, 2016, pp.: 2187-2199.
- [31] D. Martinez, P. Green, and H. Christensen, "Dysarthria intelligibility assessment in a factor analysis total variability space," In *Proceedings of INTERSPEECH*, 2013.
- [32] D. Martinez, E. Lleida, P. Green, H. Christensen, A. Ortega, and A. Miguel, "Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 6, no. 3, 2015, pp.: 1-21.