4 Intelligibility Impairment

KATHERINE C. HUSTAD[1] AND
STEPHANIE A. BORRIE[2]

[1] University of Wisconsin–Madison, WI, USA
[2] Utah State University, Logan, UT, USA

4.1 Introduction

Intelligibility is a critical concern in speech-language pathology, impacting a wide range of individuals across populations, with ages ranging across the lifespan. For example, intelligibility is an important developmental concern for children who are acquiring speech, including those with and without risk factors for speech impairment (Hustad, Mahr, & Rathouz, 2020). Speech intelligibility impairments in children can stem from speech sound disorders, childhood apraxia, conditions associated with neuromotor involvement or disease (cerebral palsy, childhood brain injury or stroke), genetic etiologies (Down syndrome, cleft palate), or sensory involvement (hearing impairment). In adults, intelligibility remains an important concern for individuals with many of the aforementioned childhood onset etiologies as well as for individuals with adult onset etiologies. Adult onset etiologies include those that have a degenerative course affecting speech motor control (e.g., amyotrophic lateral sclerosis, Parkinson's disease, Huntington's disease), those with a recovering course (e.g., stroke, traumatic brain injury during the post-onset recovery window), and those with a persistent stable course (stroke, traumatic brain injury, cerebral palsy, and other chronic conditions). The field of motor speech disorders has had a particularly notable and longstanding interest in speech intelligibility, in part because reductions in intelligibility are very frequently associated with dysarthria (Darley, Aronson, & Brown, 1969). Improving intelligibility has long been considered a key goal of treatment for this population (Ansel & Kent, 1992).

Clinically, intelligibility measures provide an important metric that can be used for treatment decision-making. Intelligibility measures are often used as a basis of comparison for documenting and monitoring change in speech performance

The Handbook of Language and Speech Disorders, Second Edition. Edited by Jack S. Damico, Nicole Müller, and Martin J. Ball.

(Yorkston, Beukelman, Strand, & Bell, 1999), as an index of severity of the speech disorder (Weismer & Martin, 1992), and as an indicator of functional ability (or disability) relative to "normal" performance (Yorkston et al., 1999). Critical clinical concerns such as justifying treatment for third-party payers, and deciding whether or not to discontinue intervention, sometimes may be made on the basis of intelligibility measures.

4.2 Intelligibility Defined

Intelligibility is a complex and multifaceted construct that is dynamic in nature. It has been defined as the extent to which an acoustic signal, generated by a speaker, can be correctly recovered by a listener (Kent, Kent, et al., 1989; Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1980). Intelligibility is a dyadic construct that is not solely attributable to a speaker or to a listener. Rather, it is a product of the joint efforts of the speaker (who produces the signal) and the listener (who interprets the signal) as communication *partners*. To be intelligible, speech does not need to be perfect or even "normal." In fact, productions may be characterized by a range of different errors and still be readily recoverable to listeners. The key issue in intelligibility is whether listeners are able to map acoustic cues onto the intended linguistic representations in spite of any deviant production patterns.

Intelligibility is influenced by a host of variables related to the speaker and his or her impairment(s), the listener and his or her ability to make sense of a distorted speech signal, and contextual factors such as the communicative environment, and shared knowledge between the speaker and the listener. Studies have shown that intelligibility can be markedly affected when different variables are manipulated. Examples include the length and nature of speech being produced (single words, individual sentences, narrative discourse, conversational discourse) (Hustad, Mahr, & Rathouz, 2020; Miller, Heise, & Lichten, 1951), semantic predictability of messages (Kent, Miolo, & Bloedel, 1994), availability of visual information (Borrie, 2015; Hustad & Cahill, 2003; Hustad, Dardis, & McCourt, 2007), and listener familiarity with the speaker (Borrie et al., 2012; Liss, Spitzer, Caviness, & Adler, 2002), to name but a few. Because so many variables influence intelligibility, no one measure can accurately and adequately provide a complete index of it. Kent and colleagues have suggested that "a particular talker has a range of intelligibility potentials, depending on listener familiarity, nature of the linguistic message, physical setting, motivation, effort level, and so on" (Kent et al., 1994, p. 81). Thus, any given measure of intelligibility is best considered a snapshot of performance under a specific set of circumstances. An intelligibility estimate must be interpreted cautiously and within the context it was obtained.

Within the World Health Organization's International Classification of Functioning, Disability, and Health (ICF-2), intelligibility has been regarded as an index of a person's ability to engage in the activity of speaking (Dykstra, Hakel, & Adams, 2007). Intelligibility is influenced by body structures and functions, particularly the speech subsystems of articulation, resonance, phonation, and respiration, and their constituent components. However, intelligibility is separate from speech subsystem structures and functions because intelligibility is a product of the subsystems working together as a collective whole to produce speech. Further, intelligibility requires the contribution of a listener. It is also influenced by contextual factors. A person's ability to participate in his or her life situations (the participation level of the ICF-2) is impacted by intelligibility as well as the broader notion of comprehensibility.

4.3 Measurement of Intelligibility

Intelligibility is challenging to measure, in part because of its complexity. A considerable body of research has demonstrated that there are many variables that may influence intelligibility, as noted above. Operationally, there are two main approaches to measuring intelligibility: objective measures and subjective measures.

4.3.1 Objective Measures of Intelligibility

Objective measures of intelligibility can be obtained and scored in numerous ways, but the primary commonality is that a specific speech sample is directly analyzed to yield a quantitative score. Objective measures often involve transcription of speech, using traditional orthography, broad phonetic transcription, narrow phonetic transcription, or forced-choice recognition of target items. These approaches typically yield a percentage of items identified correctly relative to the targets that the speaker intended to produce (Tikofsky & Tikofsky, 1964; Yorkston & Beukelman, 1978, 1980). Objective measures have been considered the "gold standard" for measuring intelligibility clinically because quantification is straightforward: units are either correct or incorrect. However, in order to score items as correct or incorrect, targets produced by the speaker must be known so that they can be scored accordingly. For this reason, elicited words and sentences are typically used for measuring intelligibility via transcription or forced-choice recognition approaches. Standard clinical tools such as the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Tice, 1996), the Assessment of Intelligibility of Dysarthric Speech (Yorkston, Beukelman, & Traynor, 1984), and the Test of Children's Speech (TOCS; Hodge & Daniels, 2007) employ this type of transcription method. Of particular note is that typically listeners are unfamiliar with the speaker and the stimulus material and are thus "naïve," potentially representing an unfamiliar listener that a speaker may encounter in daily life. However, there are several aspects of intelligibility measurement that are less ecologically representative. For example, speakers are recorded in a quiet environment, and listening tasks also usually take place in a quiet environment, which may be unlike many speaking situations. Indeed, background noise has been shown to further reduce intelligibility for listeners transcribing dysarthric speech (Yoho & Borrie, 2018). In addition, the language produced by speakers is not spontaneously generated in elicitation or recitation tasks, thus any interactions between language formulation, language production, and speech motor variables (for the better or for the worse) may not be reflected in elicited intelligibility measures. Finally, the speaker and listener do not have the opportunity to interact, making intelligibility contrived and without a real communicative purpose.

There is a body of evidence showing that adult speakers without communication disorders entrain their speech production behaviors to one another during conversation, essentially becoming more acoustically and perceptually similar (Borrie, Barrett, Willi, & Berisha, 2019; Giles & Powesland, 1975; Pardo, 2006). These interdependent adjustments to speech production behavior occur at a seemingly unconscious level during spoken dialog and are considered to reduce the computational load of spoken language processing and improve the effectiveness and efficiency with which information is exchanged. While this is a new area of investigation for individuals with intelligibility impairment, work has begun to examine entrainment of speech behaviors in the conversations that occur between individuals with dysarthria and adults without communication disorders. Preliminary evidence suggests that, while substantially reduced relative to the entrainment that occurs between two adults without communication disorders, entrainment of some speech behaviors may transpire in conversations with this clinical population (Borrie, Barrett, Liss, & Berisha, 2020; Borrie, Lubold, & Pon-Barry, 2015). While the link between entrainment and traditional measures of intelligibility has not yet received attention, entrainment of speech behavior, even in conversations with individuals with dysarthria, has been linked with objective measures of improved communicative efficiency.

The quantification of intelligibility of spontaneous speech, whether in conversational or narrative discourse, is clearly desirable because of its ecological validity. However, it is difficult, if not impossible, to score orthographic transcriptions of spontaneous speech when the number and nature of lexical targets are not definitively known. Nonetheless, the literature on children's speech and language has reported intelligibility of spontaneous speech using speech/language samples that have been transcribed for language sample analysis. For example, the number of complete and intelligible utterances divided by the total number of utterances in a transcript has been reported as a percentage of intelligible utterances (Binger, Ragsdale, & Bustos, 2016; Rice et al., 2010; Yoder, Woynaroski, & Camarata, 2016). Similarly, Flipsen (2006) proposed methods for quantifying intelligibility of conversational speech using speech samples that were transcribed for the evaluation of phonetic development using narrow phonetic transcription by expert transcribers. Although estimates of intelligibility of spontaneous speech obtained from language sample analysis or from phonetic analysis are ecologically valid, measurement procedures have some critical limitations that may inflate estimates of intelligibility. First, typically only one individual transcribes a child's speech, and this person is an expert in child speech and/or language and is thus not representative of an everyday communication partner that a child might encounter. Second, transcribers are usually allowed to play back recorded speech samples multiple times—a convenience unavailable in real-life listening situations. Third, there is typically a communication partner (clinician or parent) who interacts with the child during a speech and language sample who provides considerable contextual information and who may even gloss the child's utterances, which aids the transcriber in making sense of the speech signal. Finally, in a spontaneous speech sample context, the speaker's intended message is not definitively known a priori because it is spontaneously generated and the content of the speech/language sample is accepted as accurate if the transcriber assigned words (possibly the wrong words) to the spoken message. For these reasons, intelligibility measures obtained from speech and language samples may provide an inflated estimate of intelligibility.

One alternative that has been used in the literature is a hybrid approach, combining elements of language sample analysis, the procedures described by Flipsen (2006), and transcription intelligibility of elicited utterances. Hodge and Gotzke (2014) measured intelligibility of spontaneously generated speech by having experts create a master transcript against which unfamiliar listener responses could be scored. They then employed unfamiliar listeners who completed orthographic transcription tasks like those described for elicited words and sentences, above, to yield a percentage intelligibility score (Hodge & Gotzke, 2014). Findings indicate that intelligibility of elicited sentences from the TOCS did not differ from intelligibility of spontaneous speech samples. This convergence of findings may be related to similarities in methods, including use of unfamiliar listeners and use of a listening task that was constrained and decontextualized. Although this type of hybrid approach may not be fully reflective of the rich context available in dynamic interaction between speaker and listener, results provide important construct validity for the use of measures such as the TOCS for understanding intelligibility in children.

4.3.2 *Subjective Measures of Intelligibility*

The second main approach to measurement of speech intelligibility involves subjective measures. Subjective measures of intelligibility generally require listeners to quantify their perception of a speaker's intelligibility by assigning a number to, or scaling, what they heard (Weismer & Laures, 2002). Direct magnitude estimation (DME) procedures have been used frequently for the study of contributors to intelligibility in dysarthria. DME procedures require listeners to scale the intelligibility of speech relative to a modulus or exemplar. In contrast, use of Likert ratings, or equal appearing interval scales, requires listeners to assign numbers based on perceived similarity to anchor point descriptors (e.g., 1 = very good; 7 = very poor). Although there are a variety of problems with the use of Likert-type ratings, there is a long history in speech-language pathology of quantifying subjective phenomena using this approach (i.e., (Darley et al., 1969). One advantage of subjective measures is that a variety of dimensions of speech may be reflected in the numerical rating, and it is possible that intelligibility measures obtained from such a rating may reflect a more holistic view of intelligibility than transcription-based scores.

Perhaps the most widely referenced screening guidelines and the most widely used clinical tools for speech intelligibility development in children are based on subjective ratings of intelligibility made by parents or other familiar communication partners. For example, Coplan and Gleason (1988) identified cut points for typical intelligibility development in children between 12 months and 5 years of age by asking parents to make a forced-choice categorical rating regarding how much of their child's speech they thought a stranger would be able to understand. However, these subjective parent ratings have not been validated with objective measures of intelligibility, thus the extent to which parents rate their child accurately relative to some objective standard is unknown.

More recently, McLeod and colleagues (McLeod, Crowe, & Shahaeian, 2015; McLeod, Harrison, & McCormack, 2012) developed the Intelligibility in Context Scale (ICS) to characterize intelligibility of children across different communication partners and contexts, as revealed by parent ratings. The ICS asks parents to rate their perception of their child's intelligibility on a 5-point scale across seven different contexts. Studies of the ICS have examined its relationship with segmental measures such as percentage consonants correct, percentage vowels correct, and percentage phonemes correct, as scored on standardized tests. The ICS has not been examined relative to other measures of intelligibility to our knowledge. It is widely used and has been translated into more than 60 languages (McLeod et al., 2015), however, normative data are limited and growth curves have not been developed.

In a recent study, Natzke and colleagues (Natzke, Sakash, Mahr, & Hustad, 2020) examined measures of intelligibility including percentage of intelligible utterances, parent ratings of intelligibility, and multiword transcription intelligibility scores from elicited utterance, all obtained from the same children at three longitudinal time-points. Their results found weak associations between measures, suggesting that different measures of intelligibility are not reflective of one another even for the same children. Results also showed that not all measures were sensitive to growth over time. Large-scale studies of intelligibility using consistent methods across the full range of development are currently underway.

4.4 Intelligibility from a Developmental Perspective

Adult speakers without communication disorders are generally assumed to be fully intelligible. However, for children, acquisition of adult-like intelligible speech is a protracted developmental process, beginning early in the first year of life with vocal play, babbling, and word approximations, and continuing through childhood. Segmental development (acquisition of speech sounds) has been well documented in the literature. Expected age of acquisition for consonants and vowels in single words has been characterized using expert perceptual techniques (McLeod & Crowe, 2018; Smit, Hand, Freilinger, Bernthal, & Bird, 1990b). These data have been very useful to clinicians for assessing children's speech sound development and identifying children with speech sound disorders. Developmental data indicate that English-speaking children produce most speech sounds accurately by about 5–6 years, with adult-like mastery expected at about 8 years (Sander, 1972; Smit, Hand, Freilinger, Bernthal, & Bird, 1990a). However, studies suggest that *intelligibility* is not readily predictable from phoneme data (Ertmer, 2010; Weismer, 2008; Whitehill, 2002). For example, studies have indicated that measures such as percentage of consonants correct (PCC) have a weak relationship with intelligibility (Ertmer, 2010). Generally, the number of articulation errors is negatively correlated with intelligibility; however, individuals can have significant articulation errors and still be highly intelligible (Whitehill, 2002).

Although it is clear that children acquire intelligible speech gradually, the precise course of development of intelligibility in typical children and the range of expected variability over the full course of development is not well understood. Problems that have plagued the historical literature include methodological differences among studies, such as whether intelligibility was measured objectively or subjectively, whether listeners were "experts" (e.g., speech-language pathologists or phoneticians or naïve listeners), and the nature of speech material (elicited vs. spontaneous; single words vs. sentences vs. discourse or conversation). These differences among studies have led to conflicting reports on intelligibility development. Across studies, findings are discrepant and difficult to reconcile, for example, intelligibility of 3-year-old children has varied for different studies between about 53% and 96% (Chin, Tsai, & Gao, 2003; Flipsen, 2006; Morris, Wilcox, & Schooling, 1995; Weiss, 1982). From the existing literature it is impossible to know whether these values reflect the range of variability among typical children or whether they are a result of methodological differences between studies. Data for children at 4 years of age and older are similarly discrepant. However, one consistent and important finding is that intelligibility increases with age.

Growth curves for intelligibility development based on a large sample of typical children producing elicited utterances (single word and multiword) transcribed by unfamiliar listeners have recently been published for children up to 47 months of age (Hustad, Mahr, & Rathouz, 2020). Results indicate that there is a very wide range of variability among children at 2 years of age, with 5th and 95th percentile single word intelligibility scores of 18% and 74%, respectively. However, variability reduces somewhat with age, with 5th and 95th percentile single word intelligibility scores of 55% and 86% respectively at 4 years of age. This variability sheds some light on previous studies showing very discrepant results, suggesting that the range of typical intelligibility development is very wide, particularly for younger children. Results from Hustad and colleagues (Hustad, Mahr, & Rathouz, 2020) indicate that there is an intelligibility advantage for single word production prior to 41 months of age; after 41 months of age there is an intelligibility advantage for multiword production. Otherwise, the range of variability for typical children is similar for single word utterances and multiword utterances. Notably, typically developing children are not 100% intelligible as indicated by objective measures at 4 years of age. In addition, ongoing preliminary work comparing intelligibility development in children who speak different native languages is revealing important convergences. These results suggest that there may be some intelligibility development universals across languages. Such a finding would have critical implications for early identification of functional speech deficits in children.

A key issue is the ability to differentiate between children whose intelligibility falls within the range of age-level expectations from those whose performance is delayed or disordered with regard to age-level milestones. For many children, intelligibility reductions beyond age-expectations may have a significant detrimental impact on functional communication and on social participation, leading to important negative educational consequences, since speech is a primary modality through which children in the early grades demonstrate their learning. Accurate differential diagnosis of intelligibility deficits, early identification, and treatment to improve intelligibility is critical for these children.

Studies are currently underway that seek to identify cut points for typical intelligibility development and to validate the diagnostic accuracy of intelligibility cut points for separating children who have mild or subtle speech motor disorders from those who are in the lower percentiles of typical development. Recent work employing receiver operating characteristic (ROC) curves (DeLong, DeLong, & Clarke-Pearson, 1988) suggests that intelligibility scores differentiate between children with cerebral palsy (CP) who have speech motor impairment and typically developing peers with nearly perfect certainty (area under curve = .99). These data further suggest that at 5 years of age nearly all typically developing children had intelligibility scores above 87%, while the vast majority of children with CP and speech motor impairment had intelligibility below 72% at the same age (Hustad, Sakash, Broman, & Rathouz, 2019). These findings are consistent with earlier work suggesting that the range of intelligibility between 75% and 85% may represent a "gray area" for determining whether a 5-year-old child was performing at an age-appropriate level (Hustad, Oakes, & Allison, 2015). The determination of cut points of this nature for children across the age span is currently ongoing, with a focus on children who have CP and are thus at considerable risk for speech motor disorders.

In other work focused on intelligibility development in children, studies have shown that children with CP experience their greatest growth in single word intelligibility between the ages of 36–60 months, but that intelligibility is still developing through 8 years (Hustad, Sakash, Natzke, Broman, & Rathouz, 2019). Results indicate that growth is impacted by speech and language profile characteristics, such that children with CP who do not show evidence of speech motor involvement have their greatest growth at earlier ages and reach higher intelligibility levels at 8 years of age than those who do have speech motor involvement. Further, children with comorbid receptive language impairment and speech motor impairment lag behind their peers who do not have receptive language impairment but do have speech motor involvement (Hustad, Mahr, Broman, & Rathouz, 2020). Clearly, many attributes of the speaker impact intelligibility and its development.