Received: 28 September 2022 | Accepted: 3 April 2023

DOI: 10.1111/1460-6984.12888

International Journal of Language & Communication Disorders

REVIEW

Speech changes in old age: Methodological considerations for speech-based discrimination of healthy ageing and Alzheimer's disease

Olga Ivanova[1,3] | **Israel Martínez-Nicolás**[2,3] | **Juan José García Meilán**[2,3]

[1]Spanish Language Department, Faculty of Philology, University of Salamanca, Salamanca, Spain

[2]Department of Basic Psychology, Psychobiology and Behavioral Science Methodology, Faculty of Psychology, University of Salamanca, Salamanca, Spain

[3]Institute of Neuroscience of Castilla y León, Salamanca, Spain

Correspondence
Olga Ivanova, Departamento de Lengua Española, Facultad de Filología, Universidad de Salamanca, Plaza de Anaya s/n, E-37008 Salamanca, Spain.
Email: olga.ivanova@usal.es

Funding information:
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Abstract

Background: Recent evidence suggests that speech substantially changes in ageing. As a complex neurophysiological process, it can accurately reflect changes in the motor and cognitive systems underpinning human speech. Since healthy ageing is not always easily discriminable from early stages of dementia based on cognitive and behavioural hallmarks, speech is explored as a preclinical biomarker of pathological itineraries in old age. A greater and more specific impairment of neuromuscular activation, as well as a specific cognitive and linguistic impairment in dementia, unchain discriminating changes in speech. Yet, there is no consensus on such discriminatory speech parameters, neither on how they should be elicited and assessed.

Aims: To provide a state-of-the-art on speech parameters that allow for early discrimination between healthy and pathological ageing; the aetiology of these parameters; the effect of the type of experimental stimuli on speech elicitation and the predictive power of different speech parameters; and the most promising methods for speech analysis and their clinical implications.

Methods & Procedures: A scoping review methodology is used in accordance with the PRISMA model. Following a systematic search of PubMed, PsycINFO and CINAHL, 24 studies are included and analysed in the review.

Main Contribution: The results of this review yield three key questions for the clinical assessment of speech in ageing. First, acoustic and temporal parameters are more sensitive to changes in pathological ageing and, of these two, temporal variables are more affected by cognitive impairment. Second, different types of stimuli can trigger speech parameters with different degree of accuracy for the discrimination of clinical groups. Tasks with higher cognitive load are more precise in eliciting higher levels of accuracy. Finally, automatic speech analysis for the discrimination of healthy and pathological ageing should be improved for both research and clinical practice.

Conclusions & Implications: Speech analysis is a promising non-invasive tool for the preclinical screening of healthy and pathological ageing. The main current challenges of speech analysis in ageing are the automatization of its clinical assessment and the consideration of the speaker's cognitive background during evaluation.

KEYWORDS

Alzheimer's disease, automatic analysis, cognitive impairment, computational linguistics, dementia, healthy ageing, language, non-invasive assessment, speech

WHAT THIS PAPER ADDS

What is already known on the subject

- Societal aging goes hand in hand with the rising incidence of ageing-related neurodegenerations, mainly Alzheimer's disease (AD). This is particularly noteworthy in countries with longer life expectancies. Healthy ageing and early stages of AD share a set of cognitive and behavioural characteristics. Since there is no cure for dementias, developing methods for accurate discrimination of healthy ageing and early AD is currently a priority. Speech has been described as one of the most significantly impaired features in AD. Neuropathological alterations in motor and cognitive systems would underlie specific speech impairment in dementia. Since speech can be evaluated quickly, non-invasively and inexpensively, its value for the clinical assessment of ageing itineraries may be particularly high.

What this paper adds to existing knowledge

- Theoretical and experimental advances in the assessment of speech as a marker of AD have developed rapidly over the last decade. Yet, they are not always known to clinicians. Furthermore, there is a need to provide an updated state-of-the-art on which speech features are discriminatory to AD, how they can be assessed, what kind of results they can yield, and how such results should be interpreted. This article provides an updated overview of speech profiling, methods of speech measurement and analysis, and the clinical power of speech assessment for early discrimination of AD as the most common cause of dementia.

What are the potential or actual clinical implications of this work?

- This article provides an overview of the predictive potential of different speech parameters in relation to AD cognitive impairment. In addition, it discusses the effect that the cognitive state, the type of elicitation task and the type of assessment method may have on the results of the speech-based analysis in ageing.

INTRODUCTION

Many studies have reported significant speech changes in people with Alzheimer's disease (AD). Speech is a complex neurophysiological process that relies on both motor and cognitive functions and, as such, makes a unique window to the speaker's cognitive state (Fraser et al., 2016).

From the motor perspective, speech is an elaborate serial behaviour, which depends on motor timing and sequencing mechanisms to enable co-articulated speaking (Tremblay et al., 2017). Articulate speech is the most complex motor skill in humans: it is supported by an extensive cortical–subcortical network that regulates the automated and flexible coordination of more than 100 muscles (Ackermann, 2008; Sörös et al., 2011). Any change or deviation in this complex coordination system can trigger changes in speech production. From the cognitive perspective, speech is a two-stage psycholinguistic process connecting the conceptual network with the articulatory level. In connecting concepts resulting in articulation, speaking incorporates several intermediate stages, such as lemma representation and phonological adjustment. This way, speech holds for both low-level (motor) and high-level (conceptual) processes (cf. Hickok, 2012).

Unsurprisingly, human speech can be easily influenced by changes in physiology, cognition or both. Some of such changes lead to clinical pictures with direct, primary speech disorders (e.g., dysphasia, stuttering, apraxia, dysarthria, voice disorders, etc.). Others lead to speech impairments as a result of a main cognitive or motor condition (e.g., hormonal imbalance, hearing loss, genetic and congenital syndromes, neurodegenerations, etc.). Progressive cognitive and/or motor decline can also be responsible for changes in speech, as observed in healthy and pathological ageing.

In healthy ageing, speech changes are due to ageing-related alterations in speech-supporting anatomical bases (An Xue & Jianping Hao, 2003; Kuruvilla-Dugdale et al., 2020; Torre & Barlow, 2009) and neural mechanisms (Sörös et al., 2011; Tremblay et al., 2018, 2019). Growing difficulties in cognition (Tucker et al., 2021) and language (Burke & Shafro, 2004; Shafro & Tyler, 2014) also contribute to ageing-related speech changes. As a result, healthy older adults speak (and read) at a slower rate (Torre & Barlow, 2009), with longer pauses among syllables and halfway through utterances (Tremblay et al., 2017), with higher duration, less intelligibility (Kuruvilla-Dugdale et al., 2020) and lower accuracy (Bilodeau-Mercure & Tremblay, 2016). Ageing speech is susceptible to breakdowns and is generally characterized by the longer duration of language segments (Sörös et al., 2011). Acoustically, older adults phonate at a shorter rate and their voices show lower harmonic energy (Kuruvilla-Dugdale et al., 2020), being frequently perceived as rough (Schultz et al., 2021).

Many ageing-related neuropathologies also result in speech changes (e.g., Balci et al., 2009; Lee et al., 2011; Lowit et al., 2006). Significant changes in speech are reported in motor diseases, such as Parkinson's disease (PD) (Goberman & Elmer, 2005; Skodda & Schlegel, 2008), progressive supranuclear palsy (PSP) (Skodda et al., 2011), and multiple system atrophy (MSA) (Sachin et al., 2008). Cognitive decline is responsible for speech changes in such dementias as behavioural variant frontotemporal dementia (bvFTD) (Geraudie et al., 2021; Vogel et al., 2017), vascular dementia (VD) (Powell et al., 1988; Vuorinen et al., 2000) and, more sharply, in dementia with Lewy bodies (DLB) (Ash et al., 2012, 2016), and logopenic primary progressive aphasia (Poole et al., 2017; Marshall et al., 2018).

However, speech changes due to neurodegeneration are more consistently described for AD, the most common cause of dementia today. Although the term 'Alzheimer's disease' is recommended by the NIA-AA (the National Institute on Aging and Alzheimer's Association) Research Framework only in cases of in vivo presence of pathological biomarkers (cf. Jack et al., 2018), it will be used throughout to refer to all patients classified accordingly (either through biomarkers or symptomatologically). We consider it less misleading, and in this paper the term 'Alzheimer's disease' will not necessarily refer to confirmed biomarker presence in living people, but rather to what the NIA-AA Research Framework defines as Alzheimer's clinical syndrome or dementia.

AD is a cognitive–communication disorder, that is, a neurodegenerative disease in which impairments in cognition, mainly in memory, lead to deficits in communication and language (Badarunisa et al., 2015). Memory impairment, resulting from degenerative processes in the medial temporal lobe, is considered as the most prominent symptom of AD (Williams et al., 2021). Yet, many studies on AD also agree about significant impairments in language (cf. Fraser et al., 2016; Vincze et al., 2021). In fact, linguistic patterns, specifically lexical and semantic aspects of language, have been extensively studied as possible early markers of AD. A pioneering work by Bayles et al. (1992) already reported significant deficits in object description, picture description and superordinate naming as AD linguistic features. Successive studies confirmed AD-driven declines in semantic fluency task, action naming task and semantic verification (Pistono et al., 2018). Overall, language in early AD is characterized by a reduction in semantic units and categories (Ahmed et al., 2013; Cuetos et al., 2007; Taler & Phillips, 2008), reduced naming ability (Verma & Howard, 2012) and ability to reference people and actions (Ahmed et al., 2013), decreased lexical diversity, increased word frequency (Kavé & Dassa, 2018), and reduced discourse efficiency (Ahmed et al., 2013). Semantic perspectives on discourse also suggest that speakers with AD present with reduced information content (Kavé & Dassa, 2018), lower informativeness and higher modalization (Pistono et al., 2018). Several longitudinal studies associated the future onset of AD with repetitiveness, misspellings and telegraphic patterning in written tasks (Eyigoz et al., 2020). Importantly, language markers are highly relevant in early-onset AD, specifically in its phenotypic variant with logopenic PPA (Mendez, 2019). Unsurprisingly, the assessment of such language markers has been significantly targeted in recent years through natural language processing (NLP) and machine learning (ML).

Considering this insightful background, as well as the direct implication of cognitive and language processes in speech production, this paper will focus on speech markers of AD. First, ongoing deficits in cognition and memory can lead to deficits in neuromuscular motor coordination and in proprioceptive feedback (Voleti et al., 2020). Further, progressive deficits in language, mainly, word-finding and naming difficulties (Fraser et al., 2016; Jarrold et al., 2014; Mueller et al., 2018), reduced lexical–semantic repertoire, and impaired semantic processing (Jarrold et al., 2014), can predictively affect speech production in AD. We assume that such changes can be defined as early and discriminating for two reasons. First, since both episodic and semantic memory are early disrupted in AD, so does language (Williams et al., 2021) and, consequently, speech. Second, since early AD-related cognitive impairments can be traced through language assessment (Taler & Phillips, 2008), so can they be traced through speech assessment too. Speech production is modulated by the speaker's cognitive ability to correctly perform other stages of language planning and production, mainly conceptualization and formulation (Voleti et al., 2020). Consequently, speech can sensitively reflect any alteration in cognition and language.

Objectives and scope of the paper

The objective of this paper is twofold. On the one hand, we offer an update on the most important speech changes in AD. On the other, we overview the most salient methodological questions related to the measurement and clinical application of speech analysis in ageing. To achieve these goals, the paper follows a scoping review and discusses the next questions. First, it describes which speech parameters can differentiate AD from healthy ageing (healthy controls – HC) and mild cognitive impairment (MCI), the intermediate stage of cognitive ageing progressing to dementia in 5–10% of all cases. This aspect is particularly relevant since there is an existing need to differentiate between MCI that will evolve into AD from MCI that will not. Second, the paper discusses how different types of experimental stimuli can condition the predictive power of speech analysis, and which cognitive correlates can explain discriminating accuracy. Finally, it reviews the most promising methods for speech analysis and discusses the implications of its inclusion in clinical practice. In doing so, the following questions will be addressed:

- Which speech parameters are more systematically and reliably altered in AD compared with HC and MCI?
- Which experimental stimuli, with and without cognitive load, are more reliable for assessing speech as a clinical marker of AD?
- Can speech measures be used in clinical practice?

Methodology

To address the stated questions, this paper relies on a scoping review. Scoping reviews are specifically useful to give an account of the existing data and to identify existing research gaps (Roncarolo et al., 2017). They are based on a broader scope of topics and cover different study designs with the aim of rapidly mapping the key available evidence within a research area (Arksey & O'Malley, 2005). In approaching the literature review, we followed the stages suggested by Arksey and O'Malley (2005): (1) identification of the research question; (2) identification of relevant studies; (3) selection of studies; (4) charting of the data; and (5) collating, summarizing and reporting of the results. The sixth, optional, stage, consisting of the consultation exercise, was not carried out due to the still dominant basic nature of the speech analysis in ageing and the consequent difficulty to recruit clinical experts with sufficient expertise.

Identification of the research question

This research was guided by the following question: What speech parameters are reliable markers of AD and what experimental protocols and methodologies are particularly useful for assessing speech as an early marker of AD? Following suggestions from Levac et al. (2010), we articulated the scope of our search in a narrower way to assure a more effective search strategy. Considering the proposed objectives, we focused our research on the following aspects. We include major methodological considerations and analyse the usefulness and practicality of different types of experimental protocols and stimuli. Furthermore, we make an original—the first, to the best of our knowledge—overview of language relativity (the influence of one specific language) on discriminating speech changes in AD. We specifically focus on the universal AD-related speech changes across languages.

Identification of relevant studies

Studies for the present scoping review were searched for in PubMed, PsycINFO and CINAHL, the three most used databases in psychological sciences. Two filters were applied during the search: publication period and language of the paper. The publication period of eligible studies was established between 1 January 2010 and 31 December 2021; the start date set in view of the recent nature of research on speech as a preclinical marker of AD. Only studies published in English and Spanish were considered.

To identify the search terms, we followed the two first steps from the protocol proposed by Peters et al. (2015): (1) preliminary analysis of text words and index terms in related articles; and (2) application of identified keywords and index terms in searching across included databases. After conducting stage 1 and checking in DeCS (Descriptors in Health Sciences), we set the following search terms: speech AND Alzheimer AND (prodromal OR mild OR preclinical OR early stage OR probable). At stage 2, the search returned 309 publications in PubMed, 201 publications in PsycINFO and 112 publications in CINAHL; overall, 622 publications.

Selection of studies

All the studies found through the database search were referenced in separate documents. Duplicated studies were removed manually. The first screening step of the collected articles was done by assessing the title and the abstract. Studies that met one or more of the following criteria were excluded from the review: (1) studies published in languages other than English or Spanish; (2) studies not published as a peer-review journal article or collection; (3) studies that did not include any type of speech production analysis; (4) studies that did not include assessable data on methods and speech traits; (5) studies addressing disorders other than AD or not including AD as a contrasting group (e.g., only MCI); and (6) studies not reporting original research (e.g., protocol presentations or systematic reviews on similar topics were excluded; however, they were considered in the discussion of this work). Reference lists from the papers and grey literature were not considered.

The second screening step was done by assessing full-text papers. After relevance screening and deduplication, a total of 24 studies met the eligibility criteria. Table 1 shows how searching and selective evidence of this study complied with STARLITE criteria (Booth, 2006). The flow of papers from the identification stage to the final inclusion stage is shown in Figure 1, following criteria from Moher et al. (2015).

The estimated level of evidence is I. The general characteristics of the papers included in the review are given in Table 2.

Among possible limitations of our scoping review, we identify the possibility of having missed some relevant studies because of database selection and language limitations to English and Spanish. However, we believe that our review covers the most significant works in the field because the search was carried out in the most relevant databases for psychological sciences.

Charting the data

Data from selected papers were extracted regarding: authors, year of publication, country of origin, sample size, diagnosis, speech variables, methodology, type of stimuli, type of analysis, the language of the participants, key findings about speech variables, level of accuracy in detecting AD, and clinical relevance (if given). We did not organize the reviewed studies according to their descriptive or predictive scope.

Collating, summarizing and reporting

Considering the heterogeneity of the data reported in the reviewed studies with respect to the types of stimuli and methods of analysis, we chose a narrative approach to collating, summarizing and reporting the results. Most studies reach high and very high levels in the classification of clinical groups on the bases of speech analysis, although the types of stimuli and their relation to cognitive impairment vary widely. The main findings from our review are, thus, clustered into the next sections: (1) discriminant speech characteristics of HC and AD; (2) type of elicitation stimuli and their predictive/discriminating power; (3) experimental protocols and methods of measurement; and (4) variation across languages.

RESULTS

The overview of the literature search is summarized in Table 3. Several general conclusions can be drawn

TABLE 1 Search strategy according to STARLITE criteria

S: sampling strategy	Comprehensive—attempts to identify all relevant studies on the topic
T: Type of studies	Any study—either quantitative or qualitative
A: Approaches	Searching electronic databases
R: Range of years	January 2010–December 2021
L: Limits	English and Spanish language
I: Inclusion	Any type of speech production analysis
	Any type of speech production trait analysed
	Analysed groups: AD compared with MCI and/or HC
	(also see above)
T: Terms used	Speech AND Alzheimer AND (prodromal OR mild OR preclinical OR early stage OR probable)
E: Electronic databases	Pubmed, PsycINFO, CINAHL

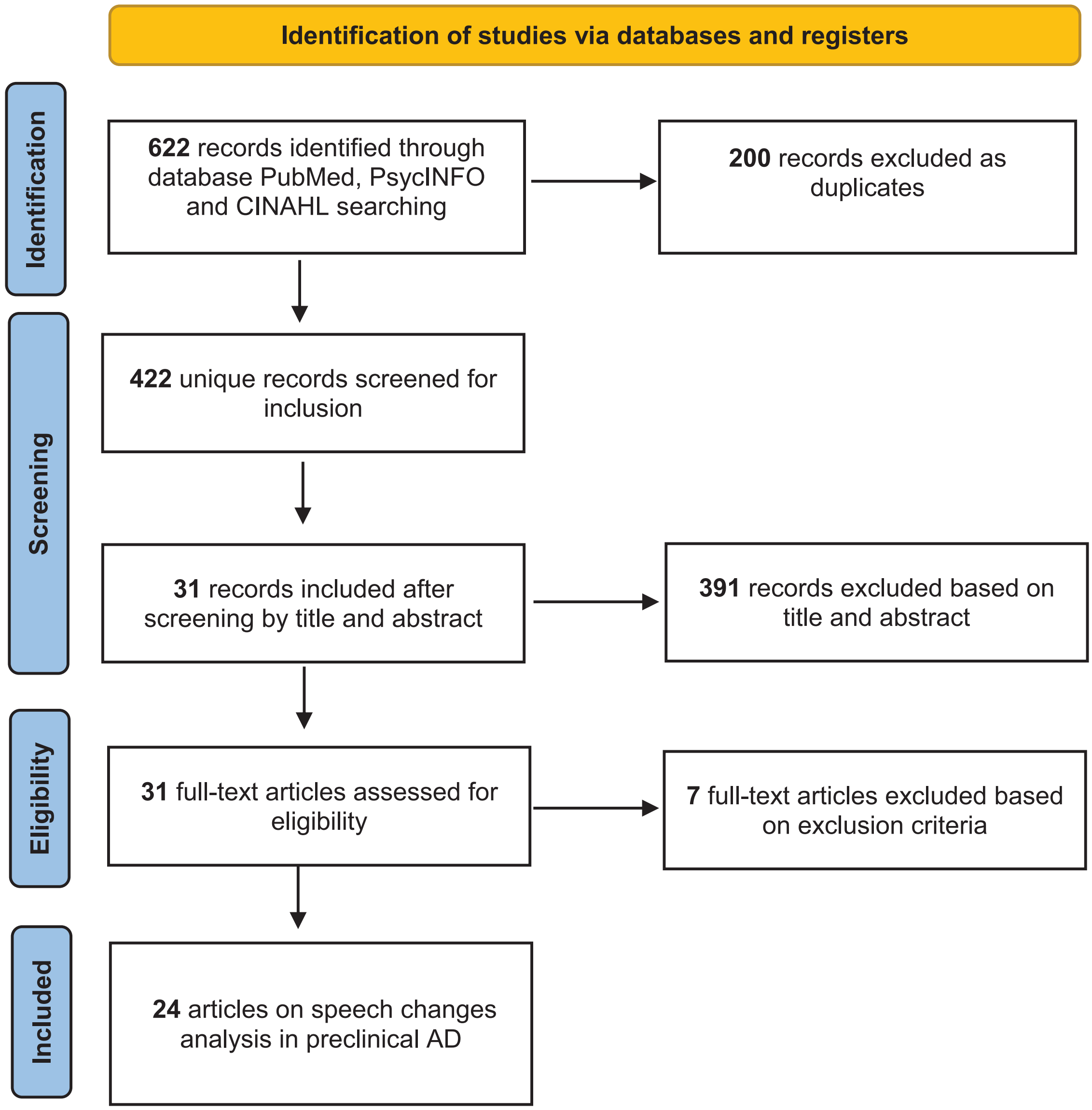

FIGURE 1 PRISMA decision flowchart [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 General characteristics of the reviewed studies

Characteristic	Number ($n = 24$)	(%)
Publication year		
2010–13	3	12.5%
2014–17	4	16.7%
2018–21	17	70.8%
Publication type		
Journal article	22	91.7%
Conference proceeding	2	8.3%

regarding speech assessment in HC and AD.

First, speech features can be more powerful for discriminating between healthy and pathological groups than other language features. Some of these speech features—mainly, temporal properties—have greater potential than others in such discrimination. Second, not all stimuli allow for the same levels of discrimination between groups: tasks with higher cognitive load are generally more sensitive to classification between healthy and pathological ageing. Third, automatic speech analysis may be a promising tool for experimental and clinical assessment of speech in ageing, but some caution is in order when applying and interpreting its results. Finally, there is a major gap in the exploration of the effect of language type on the specificity of speech changes in ageing. In the following sections, we explore these findings in more detail.

Discriminant speech characteristics of HC and AD

Speech changes are a powerful predictor of AD. In some cases, they can outperform other cognitive measurements in group discrimination. Crucially, speech parameters may gain predictive potential when they are jointly assessed with other language features, which significantly impair during the evolution to AD. In fact, some of the discriminatory speech features of AD are related to or are the result of such language impairments themselves. While this paper focuses on speech parameters only, correlations between other language features and speech features properly should not be neglected in future exploratory work.

Speech assessment can be based on the study of three interrelated components: acoustic properties, prosodic patterns and temporal features.

Acoustic properties are modulated airflow perturbations produced by laryngeal and vocal tract configurations, which define the physical and linguistically relevant aspects of speech (Kendall & Fridland, 2021). Acoustic properties of speech sounds (such as fundamental frequency (F0), harmonics, formants or amplitude) are determined by four main processes involved in speech production: the modulation of the sound source, the shape of the vocal tract filter, the characteristics of the energy losses, and the mode of sound radiating from the mouth (Harrington & Cassidy, 1999).

Prosodic patterns refer to the suprasegmental (or non-phonemic) characteristics of speech, that is, to the melodic contouring of oral uttering. Prosodic patterns (such as pitch (variation in F0), stress, intonation or rhythm) result from the variability of acoustic markers (Garcia-Toro et al., 2000). Prosody plays a central role at three levels of communication: grammatical (e.g., modulation of the utterance as a statement or a question), pragmatic (e.g., emphasis or stress on specific information) and affective (e.g., expression of attitudes or opinion) (Attwood, 2007).

Finally, within the prosodic patterns, a prominent place is given to temporal features. Temporal features refer to the timing structure of speech, segment duration and pauses (Zellner, 1994). Temporal features play an important role in linguistic contrasts (Rosen, 1992) and, thus, in the comprehension of spoken language, specifically, in its intelligibility (Greenberg et al., 2003).

Any component of speech can be affected because of changes in physiological and hormonal backgrounds, but also because of cognitive dysfunctions and impairment in the neurocognitive basis for language (Bóna, 2014), changes in the emotional spectrum (Garcia-Toro et al., 2000), sensory deficits (Leitman et al., 2011), neurologic lesions, and general disorders and diseases affecting the central nervous system (Vogel et al., 2011). In AD-related pathological ageing, cognitive and language deficits are responsible for generating discriminant changes in acoustic, prosodic and temporal variables. For ease of description, we include a small glossary of such variables in Table A1 in Appendix A.

Acoustic parameters

Basic acoustic variables are not significantly altered in AD. There are no significant differences between healthy and pathological speech in the fundamental frequency (F0) and in the average of frequencies height as measured by the spectral centre of gravity. It means that when compared with healthy older adults, speakers with AD do not produce speech sounds with significantly different brightness.

Yet, some acoustic abnormalities are among the four main language factors allowing for a precise classification of healthy and pathological ageing, in addition to semantic, syntactic and information impairment (Fraser et al., 2016). When correctly selected, such acoustic fea-

TABLE 3 Overview of the included studies on speech changes in healthy and pathological ageing

Reference and language	Sample and number of speakers per group	Experimental stimuli for speech elicitation	Assessed speech variables	Main findings
Beltrami et al. (2018) Italian	HC = 48 aMCI = 16 mdMCI = 16 mildDEM = 16	Three spontaneous speech tasks: • Narrative task (picture description) • Descriptive recall of a last dream • Description of a typical working day	• Acoustic • Prosodic • Temporal	• Speech features are more powerful, and less task dependent, in group classification than other language features (vocabulary, syntax or rhythm) • Narrative tasks are more powerful in predicting pathological ageing because of higher load on memory recall and structured discourse organization
De Looze et al. (2018) English	HC = 36 MCI = 16 m-mAD = 18	Reading aloud task of 90 sentences: • Three sets of 15 pairs of sentences; • Sentences with different length (2, 8, 16 and 32 syllables) • Sentences with different syntactic complexity (branching and non-branching)	• Temporal	• Patterns of temporal organization of speech during reading correlate with the level of cognitive impairment • Tasks including higher load over cognitive functions responsible for speech planning can elicit significant differences in the temporal organization of speech between HC, MCI and AD
De Looze et al. (2021) English	HC = 40 MCI = 20 m-mAD = 20	Three collaborative referential tasks: • Describe trial: description of shapes and indication of their position • Match trial: following instruction about description of shapes and indication of their position • Describe-and-match trial: joint discussion about shapes and their position	• Temporal	• Temporal changes in AD speech correlate with volume reduction in brain ROI related to language and speech planning, as well as timing control, working memory, and mental manipulation • Tasks with higher demands for language and speech planning are particularly sensitive to eliciting discriminating temporal traits of AD
Fraser et al. (2016) English	HC = 97 AD = 167	Cookie Theft picture description task	• Acoustic	• Acoustic abnormality is one of the four factors (in addition to semantic impairment, syntactic impairment and information impairment) explaining most of the variance between HC and AD • AD speakers show a high degree of heterogeneity/variability in their linguistic deficits
Hoffmann et al. (2010) Hungarian	HC = 15 mildAD = 10 modAD = 10 sevAD = 10	Three spontaneous speech tasks: • Question answering • Episodic recall • Description of daily activity/hobby	• Temporal	• Two temporal variables (speech tempo and hesitation ratio) significantly differentiate between HC and mild AD • Hesitation ratio is a powerful feature to differentiate between AD stages

(Continues)

TABLE 3 (Continued)

Reference and language	Sample and number of speakers per group	Experimental stimuli for speech elicitation	Assessed speech variables	Main findings
Jang et al. (2021) English	HC = 83 SMC = 9 MCI = 22 m-mAD = 48	Multimodal test including three language-based tasks: • Cookie Theft picture description task • Reading aloud of a standardized text • Recall of a positive life event	• Acoustic	• Acoustic measures are more important than other variables throughout different language tasks • Different language tasks individually show similar levels for group classification • The combination of different language tasks, plus eye measurement, significantly improves group classification performance
Kato et al. (2013) Japanese	HC = 20 MCI = 19 AD = 9	Battery of oral tasks: • Talking about hometown • Talking about childhood • Answering HDS-R questionnaire • Three reminiscence tasks: listening, talking, watching; • Three working memory tasks: category recall, reading span, face recall	• Prosodic	• There is a moderately significant correlation between neuropsychological scores and a set of prosodic features in predicting cognitive impairment • MCI is the most misleading group for accurate classification compared with HC and AD
Kato et al. (2015) Japanese	HC = 81 MCI = 81 mildAD = 81	Set of four oral tasks: • Talking about birthplace • Name of patient's elementary school • Time orientation • Repetition of three-digit numbers backward	• Prosodic	• Different oral tasks vary significantly in their discrimination power of AD • The most significant result for discriminating AD is obtained from answering to the question about time orientation. The most significant result for MCI is obtained from the repetition of three-digit numbers backward
Kato et al. (2018) Japanese	HC = 91 MCI = 91 mildAD = 91	Set of four oral tasks: • Talking about birthplace • Name of patient's elementary school • Time orientation • Repetition of three-digit numbers backward	• Prosodic	• Different oral tasks vary significantly in their discrimination power of AD • The most significant result for discriminating AD is obtained from answering to the question about time orientation. The most significant result for MCI is obtained from the repetition of three-digit numbers backward
König et al. (2015) French	HC = 15 MCI = 23 AD = 26	Set of four oral tasks: • Counting backward • Sentence repeating • Picture description • Semantic verbal fluency	• Acoustic • Temporal	• A set of approximately 20–25 speech features is good enough for classification accuracy • Different tasks (or combinations of tasks) prove to be more powerful for group pairwise classification (HC versus MCI, HC versus AD, MCI versus AD)

(Continues)

TABLE 3 (Continued)

Reference and language	Sample and number of speakers per group	Experimental stimuli for speech elicitation	Assessed speech variables	Main findings
König et al. (2018) French	HC (SCI) = 56 MCI = 44 AD = 27 MD = 38	Battery of short oral tasks: • Five sentence-repeating task • Picture denomination task • Picture description task • Phonemic verbal fluency task • Semantic verbal fluency task • Positive and negative story telling task • Episodic telling task • Counting down	• Temporal	• The combination of different oral tasks gives rise to combinations with different power for group classification • Fluency tasks are highly powerful for distinguishing between MCI and dementia • Tasks with minimum cognitive load could benefit speech-based detection of dementia at early stages
Martínez-Sánchez et al. (2018) Spanish	HC = 98 AD = 47	Reading aloud task of a text	• Acoustic	• Nine acoustic variables, plus speaker's age, build a highly significant discriminant function for automatic discrimination between HC and AD at 92.4% classification level
Meilán et al. (2012) Spanish	AD = 21	Delayed reading aloud task of stimuli with different cognitive and language load: • Sentence completion • Sentence reading and repetition • Paragraph reading aloud	• Acoustic	• From among different acoustic variables, the proportion of voiceless segments correlates with the cognitive impairment in AD
Meilán et al. (2018) Spanish	HC = 102 MCI = 38 AD = 42	Reading aloud task of a text with words with different semantic load	• Acoustic	• 96% of HC and 76.3% of AD can be successfully discriminated by a set of seven acoustic parameters
Meilán et al. (2020) Spanish	nodMCI = 73 preAD = 13	Reading aloud task of a text with words with different semantic load	• Acoustic • Temporal	• A set of temporal and acoustic properties making the rhythmic profiling of speech can allow for automated discrimination between non-degenerative MCI and preclinical AD
Nasrolahzadeh et al. (2018) Not declared (presumably, Persian)	HC = 30 AD = 30	Set of three oral tasks: • Conversation • Personal story telling • Feeling expression	• Acoustic	• Non-linear acoustic features allow for a high classification of HC and AD at different stages

(Continues)

International Journal of Language & Communication Disorders

TABLE 3 (Continued)

Reference and language	Sample and number of speakers per group	Experimental stimuli for speech elicitation	Assessed speech variables	Main findings
O'Malley et al.(2021) English	HC = 15 MCI = 15 FCD = 15 AD = 15	Set of open questions from a virtual clinician prompting conversation	• Acoustic	• An automated tool for interacting, recording and analysing speech allow for an 80% accuracy in classifying MCI or AD
Pistono et al. (2016) French	HC = 15 AD = 15	Guided oral discourse production task: free delayed recall of eight mini-events	• Temporal	• Some temporal variables (mainly, pauses) significantly correlate with difficulties in episodic memory and discourse planning in AD
Pistono et al. (2019) French	HC = 17 AD = 17	Set of two oral tasks: • Picture-based narrative • Memory-based narrative (free delayed recall of mini-events)	• Temporal	• Temporal variables obtained from picture-based narratives are more powerful in speech-based discrimination
Sluis et al. (2020) English	HC = 20 mildAD = 20 modAD = 20	Cookie Theft picture description task	• Temporal	• Mean and total pause duration significantly increases with dementia progression • Pause duration significantly differs between mild and moderate AD
Tröger et al. (2019) French	HC = 20 MCI = 55 AD = 20	Oral semantic fluency task	• Temporal	• Several temporal parameters significantly increase in AD group • Temporal enlargement in AD is related to the increasing inefficiency to access semantically unrelated terrain
Vincze et al. (2021) Hungarian	HC = 25 MCI = 25 mildAD = 25	Set of three oral tasks: • Immediate recall of film description • Delayed recall of film description • Previous day description	• Temporal	• Tasks with higher load over cognitive functions can elicit significant discriminating speech parameters between HC, MCI and AD • Tasks with load over different cognitive functions elicit different significant discriminating speech parameters among groups • Tasks with load over episodic memory (recall tasks) allow for more precise discrimination of AD
Yamada et al. (2021) Japanese	HC = 47 MCI = 45 AD = 26	Set of five oral tasks: • Counting backwards • Subtraction • Phonemic verbal fluency • Semantic verbal fluency • Cookie Theft picture description	• Acoustic • Prosodic	• Speech changes significantly correlate with impairments in episodic memory • Speech-based analysis is more powerful for classification between HC and MCI • A multimodal measurement of behavioral data (speech, gait and drawing) can significantly improve classification among groups (HC, MCI and AD) and in pairwise comparison (100% for AD/HC)

(Continues)

TABLE 3 (Continued)

Reference and language	Sample and number of speakers per group	Experimental stimuli for speech elicitation	Assessed speech variables	Main findings
Yeung et al. (2021) English	HC = 30 MCI = 30 AD = 30	Cookie Theft picture description task	• Acoustic • Temporal	• Speech variables observed through word-finding, incoherence and perseveration correlate with clinical group classification • Longer samples of speech might be more useful in speech-based assessment of MCI and AD

Note: AD, Alzheimer's disease; aMCI, amnesic mild cognitive impairment; FCD, functional cognitive disorder; HC, healthy controls; MCI, mild cognitive impairment; MD, mixed dementia; mdMCD, multiple domain mild cognitive impairment; mildAD, mild Alzheimer's disease; mildDEM, mild dementias (mixed group); m-mAD, mild-to-moderate Alzheimer's disease; modAD, moderate Alzheimer's disease; nodMCI, non-degenerative MCI; preAD, preclinical Alzheimer's disease; SCI, subjective cognitive impairment; sevAD, severe Alzheimer's disease; SMC, subjective memory complaint.
We only report language-oriented tasks used in the reviewed studies. Many of the reviewed studies also included other tasks relative to cognitive and executive screening and assessment, but we do not report them considering the scope of the present paper.

tures can be merged as a discriminating function with high levels of classification accuracy (Martínez-Sánchez et al., 2018; Meilán et al., 2018). Significant AD-related acoustic changes are mainly related to the spectral properties of articulated sounds, that is, to how acoustic modulation of speech sounds is organized in time. Speakers developing AD tend to articulate segments with higher asymmetry towards lower frequencies (Meilán et al., 2020) and show lower Mel-frequency cepstral coefficient 1 (MFCC1) when compared with HC (Yamada et al., 2021). In the long-term average spectrum (LTAS), speakers with AD show a lower intensity and a lower variation (Meilán et al., 2020). They also produce more acoustic signals lacking periodicity, and more acoustic signals with less amplitude despite trying to phonate appropriately. When compared with HC, speakers with AD produce segments with a higher percentage of unvoiced periods and voice breaks, lower Shimmer Apq11 (amplitude perturbation quotient), and higher variability in the standard deviation of the third formant (F3). Some of these parameters (mainly, the percentage of voiceless segments) correlate with the neuropsychological scoring and, thus, with the degree of cognitive impairment (Meilán et al., 2012).

Importantly, measuring selective acoustic variables offers a higher power in discriminating AD than other speech variables (Jang et al., 2021). One possible explanation is that acoustic variables are less dependent on cognitive load than prosodic or temporal variables, so they remain more stable during speaking. This assumption is supported by the results from the VAD-AD prototype (cf. Martínez-Sánchez et al., 2018), which classifies HC and AD with 92.4% classification accuracy based on the automatic analysis of nine acoustic variables.

Prosodic parameters

Prosodic parameters are important in speech since they convey information about the emotional state of the utterer or about his/her cognitive impairment (Kato et al., 2018). Yet, prosody is probably the least explored aspect of speech in ageing, and findings on prosodic parameters in AD are varying. Some research suggests that prosody does not change significantly in pathological ageing. For example, speech rhythm and vowel–consonant alternation patterns do not show significant variation between HC and AD (Beltrami et al., 2018). At the same time, other studies observe prosodic impairments in speakers with dementia when compared with healthy older adults. Impaired prosodic variables include lower pitch variability (Yamada et al., 2021) and higher rhythmic variability as measured by normalized pairwise variability index (nPVI). It means that AD speakers show greater durational contrast among adjacent elements, and their speech sounds less expressive. Overall, correlations between prosodic features and neuropsychological scores only show moderately significant power (Kato et al., 2013).

Temporal parameters

Temporal parameters are among the most explored speech parameters in ageing. Timing in speech is cognitively conditioned, with cognitive agility being an important predictor of the temporal organization of speech and language. In AD, temporal aspects of speech can be significantly altered as a result of deficits in executive functions, working and episodic memory, mental manipulation, timing control, attention, language decision-making, semantic retrieval, discourse planning and topic adjustment (De Looze et al., 2021; Sluis et al., 2020). Indeed, AD results in an increased slowness of speech, primarily observed in the increased speech/reading rate, longer phonation, and longer and more frequent pauses. AD speakers produce fewer words per minute (Pistono et al., 2016) and shorter/more speech chunks (De Looze et al., 2018). It means that they segment their discourse more and do so in shorter pieces of speech. In general, with the progress of AD, speakers tend to produce fewer phonemes per timeslot and, consequently, more absences of speech (Hoffmann et al., 2010).

However, the most important temporal variable for the classification of dementia is pausing. In AD, pauses become significantly longer in duration, though not necessarily more frequent (Pistono et al., 2016; Pistono et al., 2019; Sluis et al., 2020). It is important to note that such significant pause lengthening is recorded at the speech points with the highest cognitive load, that is, when speech production puts more pronounced demands on recall or discourse planning (Pistono et al., 2016). The characteristics of pauses (in particular, their increased frequency) may draw up the degree of ongoing cognitive impairment. A higher appearance of pauses during lexical–semantic search may point to a better preservation of semantic memory and, thus, to a lower degree of cognitive impairment (i.e., to more mild dementia). Here, pauses themselves can be regarded as a cognitive compensatory strategy for improving mental time travel and consequent memory retrieval (Pistono et al., 2016, 2019).

Overview of discriminant speech characteristics of healthy and pathological ageing

Different speech parameters (acoustic, prosodic and temporal) significantly change in pathological ageing. Yet, one of the most promising approaches is based on the combination of their different types. Combining acoustic and prosodic (Yamada et al., 2021) or acoustic and temporal variables (Meilán et al., 2020; Yeung et al., 2021) can potentially profile cognitive states in ageing in a more precise way. Meilán et al. (2020) explored such a combination of temporal and acoustic measures under the perspective of rhythmic profiling, showing AD specificity in temporal variability and vocalization of syllabic intervals.

Overall, speech production in persons with AD is characterized by slower speech rate, slower turn-taking, shorter transition overlaps, a larger number of gaps, longer gap durations, shorter interpausal units (De Looze et al., 2021), general longer word duration, and more and longer pauses (Yeung et al., 2021). As a result, speakers with AD stammer more during reading, produce more and less appropriate pauses, utter with higher rhythmic variability, speak with lower speech volume, sound hesitant and monotonous, and with no oscillations in intensity.

The predictive power of elicitation stimuli for speech analysis in ageing

Studies use different types of tasks to elicit speech in ageing people. The most common tests are picture description tasks (41.7%) (usually it is the Cookie Theft picture from the Boston Diagnostic Aphasia Examination set), recall tasks (37.5%), reading aloud tasks (29.1%) and activity/place description tasks (25%).

Importantly, most studies (62.5%) combine several tasks to obtain more complex speech samples. The most common combinations are two: (1) batteries including tasks based on the involvement of different cognitive functions (mainly, different types of memory); and (2) batteries including tasks based on the involvement of different types of language production (spontaneous production, guided production, reading, etc.).

For the first, the assumption is that in cognitively different and specifically demanding tasks, speech production would provide more reliable and precise evidence of ongoing cognitive impairment. Thus, the specificity of different stimuli may involve specific cognitive functions more intensively (Table 4).

Indeed, despite not explicitly commenting on it, results from many experiments suggest that the specificity of the cognitive load from elicitation tasks may lead to more accurate classification results. A higher cognitive load is sensitive to the classification of groups with greater cognitive impairment, mainly non-degenerative MCI versus preclinical AD (Meilán et al., 2020) or HC versus AD (Kato et al., 2015, 2018). In these studies, an association is found between speech changes and scoring in episodic memory

TABLE 4 Cognitive functions involved into different speech elicitation tasks

Speech elicitation task	Related cognitive function	Reference
Picture description task	Episodic memory	Fraser et al. (2016)
	Information processing	Sluis et al. (2020)
	Visual recognition memory	Pistono et al. (2019)
Immediate recall task	Episodic memory	Vincze et al. (2021)
Delayed recall task	Episodic memory	Vincze et al. (2021)Pistono et al. (2016)
Reading aloud	Working memory Attention Planning	De Looze et al. (2018)
Description task	Working memory	Vincze et al. (2021)
Conversation	Episodic memory Working memory Sustained attention Executive functions	De Looze et al. (2021)

(Yamada et al., 2021) or in global cognitive effort (Meilán et al., 2020). In turn, a lower cognitive load would allow for better discrimination between groups with more severe initial cognitive impairment, for example, HC versus early AD, as it balances the cognitive load across subjects (cf. König et al., 2015, 2018, for a discussion).

For the second, the assumption is that different characteristics of oral speaking (spontaneous versus non-spontaneous; unguided versus guided; semantically free versus semantically conditioned) may better reflect the specificity of the underlying cognitive and linguistic impairments. For example, pathological ageing triggers sensitive speech predictors in such cognitively demanding tasks as semantically conditioned production (Tröger et al., 2019) and exposition to longer utterances (De Looze et al., 2019).

Overall, there are three important points to note. First, different tasks are not systematic in eliciting discriminating speech variables across clinical groups. Furthermore, their predictive power for group classification is variable. Second, multi-class classification (HC versus MCI versus AD) is more accurate than binary-class classification. In binary-class analyses, the pairwise classification HC/AD is usually more precise than the pairwise classification HC/MCI (Table 5). Finally, not all speech variables elicited by different tasks show progressive, or linear change from HC to MCI to AD. Some speech variables are specific for MCI or AD stages only.

Eventually, different types of tasks may be more precise for eliciting specific discriminating speech parameters (acoustic, prosodic or temporal) in a more sensitive way. Acoustic parameters are more predictive when elicited through reading tasks and discriminate more precisely between HC and AD. Prosodic variables are more predictive in guided speech production tasks (counting or reading) and only discriminate between HC and AD. However, temporal parameters are sensitive to both HC/AD and MCI/AD discrimination, and the relationship between the elicitation task and the predictive power of temporal variables is more complex. Speech, phonation and voicing rates discriminate all clinical groups independently of the type of task. Yet, speech timing is more sensitive to reading tasks and is more reliable for MCI/AD classification. Pausing, in its different varieties (silent versus filled), allows to discriminate HC and AD through almost all types of tasks, including those based on memory load (picture description or recall) or language load (reading). However, MCI and AD can be classified through pausing (number of pauses, ratio of pauses to utterances, and length of pauses) only through tasks with a higher load on memory. Assessment of progression from HC to MCI to AD based on pausing is better measured by reading tasks.

Table 6 summarizes findings on the correlation of speech variables/stimuli for both binary and multi-class discrimination.

Methodological approaches to speech assessment in healthy and pathological ageing

Speech analysis for clinical purposes is aimed at extracting relevant speech parameters for group classification. The exploration of speech as a biomarker of AD is a novel research line, so most studies are still focused on the identification of discriminatory sets of parameters with the highest level of predictive accuracy. Therefore, two key methodological aspects can be highlighted: (1) approaches to speech analysis for group classification; and (2) approaches to testing their clinical application.

In performing speech analysis for group classification, researchers rely on different software. The most commonly

International Journal of Language & Communication Disorders

TABLE 5 Classification accuracy of speech variables in binary and multi-class classification

Speech variables	HC versus MCI	HC versus AD	Multi-class	Reference
Acoustic	80%	80.0%	65%	O'Malley et al. (2021)
		91.2%		Meilán et al. (2018)
		97.7%		Nasrolahzadeh et al. (2018)
Prosodic			85.4%	Kato et al. (2013)
	70.9−76.4%	67.2−89.5%		Kato et al. (2015, 2018)
Temporal	63%	74%	84%	De Looze et al. (2021)
	86%	92%	78%	König et al. (2018)
Acoustic and temporal	80%	87%		König et al. (2015)
	92.4%			Martínez-Sánchez et al. (2018)

used is PRAAT: an open-source program developed by Paul Boersma and David Weenink (Boersma, 2001; Boersma & Weenink, 2022).

One of the advantages of PRAAT is that it allows to develop ad-hoc scripts and algorithms. It is systematically used for primary speech data processing and analysis, with the subsequent application of inferential and regression statistics. In most cases, the goal is to identify and describe powerful patterns of speech for the discrimination of healthy and pathological ageing.

The qualitative change in speech analysis comes with the application of the results obtained in PRAAT (or other speech analysis software, e.g., Audacity) to the automated evaluation of clinical groups. The obtained algorithms are tested and used for training specifically developed devices or software programs. This will usually include training automatic classifiers and testing their detection accuracy (König et al., 2018). Insightful descriptions of how speech analysis is prepared for further applications, specifically in clinics, can be found in König et al. (2018), López-de-Ipiña et al. (2020), and, in general, in papers derived from the ADReSSo Challenge (https://dementia.talkbank.org/ADReSS-2021/).

Although it is not the objective of this paper to describe the whole process and steps of application of speech analysis to clinical assessment, let us make some comments in this respect. Speech technologies can be very precise and provide objective, automated, and reliable quantitative analysis of clinical groups (Pappagari et al., 2021; Vincze et al., 2021). It is, furthermore, reasonable that much of the research focuses on the development of automatic methods that would relate speech-based assessment with confirmed diagnoses, neuropsychological scores or other dementia-related symptomatology. Most of the studies rely on ML algorithms. The better the predictive (or classification) power of the algorithm, the better the match between speech profiling and the clinical condition. Some research groups have already developed their own software or have either adapted the existing software to their goals. This is the case of *Calpy*, an open-source speech processing toolkit aimed at automated identification and coding of pauses in a pre-recorded speech in ageing speakers (Sluis et al., 2020).

Among the most widely used tools for the development of automatic speech-assessing algorithms are automatic speech recognition (ASR) and automatic speech analysis (ASA). ASR allows for the automatic conversion of speech to text and can be useful in the automatization of speech sample transcription. ASA allows for the automatic analysis of speech samples and their identification in accordance with the given classification parameters. Some research groups are currently working on developing ASA-based tools for their further use in clinical settings. König et al. (2018) developed a mobile application that interacts with patients by producing verbal stimuli and recording their answers. Martínez-Sánchez et al. (2018) developed the VAD-AD device that records patients while they read a short text and immediately analyses their speech as pathological or not. O'Malley et al. (2021) developed a virtual clinician tool, known as *CognoSpeak*, that records and progressively assesses patients' speech productions.

Yet, caution should be exercised regarding the use of ASA/ASR in clinics. First, most of the current ASA/ASR protocols in ageing are mainly descriptive or correlationally predictive. They do not consider causal models, which would rigorously link cognitive and behavioural alterations in pathological ageing to speech parameters. Second, we cannot disregard the important bias from the sample sizes on which the ASA protocols are usually trained. Intersubject variability is an important factor to consider. Cognition is a highly complex process, and small samples cannot properly reflect such complexity when identifying clinically relevant speech features and generating accurate performance ranges (Berisha et al., 2022). Third, since AD is heterogeneous, so is its symptomatology. The major drivers of AD heterogeneity are risk factors (mainly, biological factors, such as age, sex or the polymorphism of APOE (apolipoprotein gene), protec-

TABLE 6 Speech variables significantly discriminating in binary-class through different experimental stimuli

Speech variable	Type of experimental stimuli		Reference
	Significant HC versus AD	Significant MCI versus AD	
Acoustic parameters			
Asymmetry (skewness, Hz)		Reading	Meilán et al. (2020)
	Reading		Martínez-Sánchez et al. (2018)
Mel frequency cepstral coefficients			
• MFCC1	Picture description		Yamada et al. (2021)
• MFCC8 (variance)	Memory recall		Jang et al. (2021)
• MFCC12 (variance)	Reading		Jang et al. (2021)
Centre of gravity (SD)		Reading	Meilán et al. (2020)
Long-term average spectrum/LTAS		Reading	Meilán et al. (2020)
Voice's minimum amplitude (mean)	Reading		Martínez-Sánchez et al. (2018)
Maximum amplitude difference value (mean)	Reading		Martínez-Sánchez et al. (2018)
Shimmer Apq11	Reading		Meilán et al. (2018)
F1 (SD)	Reading		Martínez-Sánchez et al. (2018)
F3 (SD)	Reading		Meilán et al. (2018)
F3 bandwidth (B3)	Reading		Martínez-Sánchez et al. (2018)
Harmonics-to-noise ratio (SD)	Reading		Martínez-Sánchez et al. (2018)
AVQI harmonics-to-noise ratio	Reading		Martínez-Sánchez et al. (2018)
Prosodic parameters			
Pitch variability	Counting backwards		Yamada et al. (2021)
Pitch trajectory (in syllabic nuclei)	Reading		Martínez-Sánchez et al. (2018)
Normalized pairwise variability index (nPVI)		Reading	Meilán et al. (2020)
	Reading		Martínez-Sánchez et al. (2018)
Temporal parameters			
Speech rate/speaking rate	Picture description		Beltrami et al. (2018); De Looze et al. (2021); Pistono et al. (2019); Yeung et al. (2021)
		Picture description	Beltrami et al. (2018)
	Memory recall		Pistono et al. (2016)
		Memory recall	Vincze et al. (2021)
	Reading	Reading	De Looze et al. (2018)
Articulation rate/speech tempo	Reading		De Looze et al. (2018)
	Conversation		Hoffmann et al. (2010)
Reading time	Reading		De Looze et al. (2018)
		Reading	De Looze et al. (2018): Meilán et al. (2020)
Phonation time/phonation rate	Picture description	Picture description	Beltrami et al. (2018)
		Reading	Meilán et al. (2020)
Duration of voice segments	Picture description	Picture description	Beltrami et al. (2018)
	Memory recall	Memory recall	König et al. (2018)

(Continues)

International Journal of Language & Communication Disorders

TABLE 6 (Continued)

Speech variable	Type of experimental stimuli		Reference
	Significant HC versus AD	Significant MCI versus AD	
Phrase and speech duration	Picture description		Sluis et al. (2020)
Word duration	Picture description		Yeung et al. (2021)
Duration of voiced versus unvoiced segments	Memory recall	Memory recall	König et al. (2018)
	Reading	Reading	Meilán et al. (2018)
Time distribution between words		Phonemic fluency task	König et al. (2018)
Number of syllabic intervals		Reading	Meilán et al. (2020)
Voice breaks (%)	Reading	Reading	Meilán et al. (2018)
Number of total pauses (silent and filled)	Picture description		De Looze et al. (2018); Yeung et al. (2021)
	Memory recall		Pistono et al. (2019)
		Memory recall	Vincze et al. (2021)
	Reading	Reading	Meilán et al. (2018, 2020)
Number of silent pauses		Memory recall	Vincze et al. (2021)
Number of long pauses (over 2 s)	Picture description		Sluis et al. (2021)
Length of total pauses (s)	Picture description		Beltrami et el. (2018) et al. (2018); Yamada et al. (2021); Yeung et al. (2021)
		Picture description	Beltrami et al. (2018)
	Memory recall		Pistono et al. (2016)
Length of total pauses (s)	Semantic fluency		Yamada et al. (2021)
	Counting backwards		Yamada et al. (2021)
		Conversation	Vincze et al. (2021)
Length of silent pauses (s)	Picture description		Sluis et al. (2021)
	Counting backwards		König et al. (2018)
Ratio of silent pauses/utterance length (%)		Memory recall	Vincze et al. (2021)
Between-utterance pause (rate and %)	Memory recall		Pistono et al. (2016)
Number of speech chunks/ interpausal units	Reading	Reading	De Looze et al. (2018, 2021)
Transition overlaps in speech chunks	Description task		De Looze et al. (2021)
Disfluencies (hesitations, corrections, truncations and repetitions)	Reading	Reading	De Looze et al. (2018)
Hesitation ratio	Memory recall		Hoffmann et al. (2010)
	Description task		

Note: SD, standard deviation.

tive factors (mainly, cognitive and neurocognitive factors, such as cognitive reserve or brain resistance), and concomitant non-AD pathologies (Ferreira et al., 2020). Such heterogeneity should be borne in mind when assessing the reliability of any kind of results from the non-invasive assessment of AD.

Overall, there is a consensus that the automatization of speech discriminant analysis is currently one of the research priorities in clinical phonetics. However, two aspects must be considered: (1) the definition of protocols for accurate speech-based and cognitive-related classification of clinical groups; and (2) the importance of complex testing. In the latter respect, the application of multimodal tests, combining automatic speech analysis with the study of other factors, may improve the predictive potential in the classification of clinical groups. Speech analysis can be combined with the assessment of gait and drawing, allowing for 100% classification accuracy of HC and AD (Yamada et al., 2021), or of eye movements, significantly altered in pathological ageing (Jang et al., 2021). Overall, artificial intelligence (AI) techniques could be a promising, unexpensive, and accessible set of techniques for early discrimination of clinical groups. Some studies (Li et al., 2022) suggest that tools based on AI have a great potential for developing digital markers of pathological ageing. Yet, there are still some shortcomings in using AI in discriminating ageing itineraries based on speech. The major challenges are frequent inconsistency, instability, low quality of data, limitations of samples and limitations of sample diversity (Dashwood et al., 2021). Furthermore, some caution is still needed to correctly assess the impact and the preciseness of AI methods in the clinical evaluation of speech in ageing.

Speech changes in ageing and cross-linguistic variation

One of the most intriguing, yet little explored questions is how speech changes in ageing are systematic (or variable) across languages, and to what extent automated tools for speech assessment are valid cross-linguistically. To begin with, let us say that very little work focused intentionally on cross-linguistic variation in pathological ageing. In a recent position paper, Ivanova (2023) considered several isolated studies with contradictory results regarding a universal pattern of language involution in ageing (some say yes, others say no). Regarding speech, two studies suggest cross-linguistic differences in healthy ageing: in articulation rate (Gerstenberg et al., 2018) and in phonological clustering (Ardila, 2020), respectively.

Regarding pathological ageing, research is even more limited. Still, understanding the extent of the cross-linguistic universality of AD-related speech changes bears important methodological and clinical consequences. Studies included in our review offer very little discussion on the effect of a particular language on discriminatory speech parameters in AD. Only one of them suggests that their results are language specific (cf. Kato et al., 2018), which is consistent with previous observations on language specificity of speech in healthy ageing (cf. Tremblay et al., 2019).

However, the question makes a major gap in clinical phonetics. There is an urgent need to determine how acoustic, prosodic and temporal parameters of speech vary between speakers of different languages. The lack of speech databases for most of the World languages (cf. Zhang et al., 2017) is a major impediment to our understanding of how typological variation affects clinical parameters. By now, the available evidence suggests that there should be a significant effect of language typology on the patterns of speech change in AD speakers of different languages.

Let us first consider the crosslinguistic variation at other language levels. In general, the pattern of language decline is common among patients with AD independently of the language they speak. Yet, some typological properties affect this pattern cross-linguistically. For example, in morphology speakers with AD show different patterns of regularization bias depending on the morphological structure of their language. Regularization bias is less pronounced in speakers of morphologically richer languages (e.g., Italian versus English) (Walenski et al., 2009). Similar tendencies are observed in syntax, with pronoun overuse being significant only in speakers of non-pro-drop languages (e.g., English versus Bengali) (Bose et al., 2022).

Although there are no data on typological variation in AD speech, we can draw parallels with other neuropathologies. Indeed, recent findings in clinical phonetics suggest that speech parameters can also be determined cross-linguistically. For example, a comparative study of speakers of five Indo-European languages (Czech, English, German, French and Italian) performing sustained phonation of /a/, /pa-ta-ka/ syllable repetition, reading task and oral narration task, did not find any difference in overall severity of acoustic and perceptual changes in people suffering from PD. But, at the same time, the authors found different effects of language typology on how speakers changed in voice quality, consonant articulation, articulation rate, loudness variability or sequential motion rate (Rusz et al., 2021). An important conclusion of this work was that PD was defined by both universal and language-specific speech traits. Interestingly, cross-linguistic phonetic variation was also found in autism spectrum disorder (Lau et al., 2022).

Overall, these very limited data suggest that language and, specifically, speech markers of AD may vary (and vary significantly) in different languages. There is, furthermore, an urgent need to expand the availability of clinical data for under-explored languages. Beveridge and Bak (2011) already mentioned the extreme bias of clinical language samples towards European languages, mainly English, and the consequent limitation in the applicability of clinical findings.

DISCUSSION

Speech analysis is one of the most promising research lines in assessing ageing as healthy or potentially pathological. The results of this review show that temporal speech parameters and elicitation stimuli with higher cognitive load are more powerful for discriminating HC and AD. Although acoustic parameters of speech are more stable and less dependent on the typology of stimuli, temporal parameters are more sensitive to cognitive and language decline. Thus, speech timing, as measured by patterns of speech production and pausing, shows higher levels of group discrimination because of its stronger dependence on AD itinerary. We, therefore, conclude that not all elicitation tasks reveal significant speech impairments between healthy and pathological ageing, nor do all speech traits change systematically and lineally in HC, MCI and AD.

One of the most interesting questions is what type of load allows for more reliable speech marking between HC and AD. At the beginning of this paper, we mentioned the relevance of linguistic patterns for early discrimination of AD. Thus, one could assume that tasks focused on language, rather than on speech, would be more cognitively complex and, consequently, more sensitive to group differentiation. Yet, a recent study assessing binomial speech–cognitive load relation (MacPherson, 2019) showed that increased cognitive load, as measured by embedding incongruent Stroop condition in sentence repetition task, was directly associated with changes in articulatory and temporal aspects of speech. These changes were related to overload in selective attention, working memory and inhibition, rather than in language itself. Importantly, the study observed a higher association for older speakers. Works of this kind let us conclude that, while the cognitive load imposed by language tasks may be an important source of data, speech parameters themselves can also offer valuable insights into cognitive impairment.

Still, the importance to consider the correlation between speech changes and other language deficits in AD is undeniable. As mentioned previously, many studies point to a high reliability of some language features for early identification of AD, and its discrimination not only against HC but also against other neurodegenerative profiles. What we consider important here is to point out the combinatorial capacity of measuring language and speech for the same purpose. Recent advances in computational approaches suggest that combining speech and lexical (Farrús & Codina-Filbà, 2020) or speech and semantic features (Gosztolya et al., 2019) leads to high levels of classification accuracy of clinical groups. Considering these important observations, but also the fact that computational analyses still need to be tested, we suggest that at this point in time speech analysis can potentially offer higher reliability because of its minor dependence on speaker's individual idiosyncrasy. As the linguistic profiling assessment progresses, speech analysis already allows to minimize the effect of linguistic and sociolinguistic variation, cognitive reserve, communication, and reading experience, and other factors influencing overall language performance.

The automatization of speech analysis is a priority for the transfer of speech-based tools to the clinics. Yet, as this review shows, attempts to generate prototypes based on ASR and ASA are still few and far between. Some of the protocols described in recent literature are trained on small-size samples and, thus, overperform in classification accuracy. Furthermore, the development of ASA/ASR protocols must necessarily consider two important factors: inter-subject variability among speakers and the effect of cognitive impairment.

That said, the current state of speech analysis for the discrimination of healthy and pathological ageing is already yielding important practical and clinical implications.

First, neuropsychological assessment is not always precise in the discrimination of clinical groups, specifically if the diagnosis is made presymptomatically (Meilán et al., 2020). This is particularly true when discrimination is aimed to be done between non-degenerative MCI, a highly common age-related clinical condition, and preclinical AD, a clinical condition that will derive from between 10% and 15% of all MCI (Ataollahi Eshkoor et al., 2015). In addition, neuropsychological tools do not always dismiss the effect of social variables (Kato et al., 2013, 2015), such as education level, individual sociocultural development, or socio-communicative habits. Consequently, the first, and the utmost clinical implication of speech analysis is its significant contribution to the identification of pathological ageing in a non-invasive and highly precise way. Indeed, speech analysis, specifically in its automated version, can serve as a highly valuable and powerful supportive tool for early-stage discrimination of dementia (König et al., 2015).

Second, speech analysis can be used in clinics, but can clinicians be trained to identify AD development by how patients express themselves? Several speech traits can be indeed perceptually noticeable, specifically the temporal ones (cf. De Looze et al., 2018), such as pauses, speech rate or articulation rate. But, to what extent all speakers with pathological ageing, and speakers with pathological ageing at different stages, can be expected to produce such noticeable characteristics? Truly indeed, correlations between clinicians' assessment of older adults as belonging to a clinical group, and their automatic speech-based classification as a such clinical group, can be quite precise in some cases. Yet, such correlations do not always distinguish between the most difficult cases, for example, between MCI and AD (Yeung et al., 2021). This observation makes evident that automatic analysis of speech can be a highly useful, and necessary tool for more reliable identification of dementia progression. Improved discriminating algorithms and speech analysis models will allow for an accurate non-invasive analysis of language in ageing.

Third, speech analysis in ageing can be automatized. Automatic analysis of speech variables is currently one of the most promising methods for identifying age-related clinical pictures in a non-invasive way (Hall et al., 2019). Indeed, most of the reviewed studies underline the need for an early, non-invasive, and effective mean of diagnosis of AD. Automatic recognition of speech (ARS) and ASA seem to be one of the most promising tools in this respect. The usefulness of automatic analysis is particularly significant if we consider that neuropsychological symptoms are not always discriminant between clinical groups and, in addition, are highly susceptible to subjective opinions from patients and caregivers. Moreover, automated analysis is necessary considering the difficulty that the manual study of speech productions can make for therapists. At present, one of the main challenges of automatization of speech analysis is to find out which speech factors generate the most accurate combination for the assessment of speech, as some of the many speech parameters can generate noise, and even decrease classification accuracy within a parameter set (cf. König et al., 2015). Furthermore, as discussed previously, protocols of speech analysis must be improved by passing from descriptive/correlational to causal, involving bigger samples, and, importantly, considering background cognitive impairment in ageing.

Overall, it should be borne in mind that the aim of speech analysis is to discriminate between groups, not to diagnose. Let us not forget either that, according to the NIA-AA Research Framework (Jack et al., 2018), the evolution from HA to MCI to AD is defined as a continuum, rather than as three separate stages. Indeed, the true value of speech analysis does not refer to diagnosing AD, which can (and should) be confirmed by other clearly evident cognitive symptoms assessable through neuropsychological tests, but rather to screening subjective or mild cognitive impairment in older adults as pathological or non-pathological (cf. Beltrami et al., 2018; Jarrold et al., 2014). One cannot therefore forget that speech analysis is emerging as a very promising clinical method, not only for the discrimination of AD as the most common cause of dementia, but also for identifying other clinical conditions with diffuse symptomatology, including depression or risk of suicide (Jarrold et al., 2014).

CONCLUSIONS

This scoping review on speech analysis for discriminating healthy and pathological ageing shows that some speech variables, in particular temporal features, are more sensitive to neuropathological progress in ageing. It also confirms that elicitation tasks with higher cognitive load may be more precise in displaying these features. The predictive potential of different speech parameters correlates with different variables, including: type of stimuli, clinical diagnosis (HC, MCI or AD), combinatory organization of speech parameter sets, methods of data analysis, methods of automatization and typology of the native language.

Although caution is in order in suggesting a reliable guide for the clinical application of speech analysis as things state at present (the reviewed studies show that this research line still needs to improve in accuracy, protocols and predictive methods), we can make some relatively safe suggestions based on our review. They can be formulated as follows: (1) temporal parameters, specifically speech tempo, hesitation ratio, number of pauses and mean/total pause duration, are the strongest predictors of cognitive impairment; acoustic parameters are less powerful, but they increase in predictive value when analysed as set of features (from 7 to 25); (2) narrative tasks are more powerful for speech sample collection because they lead to higher load on memory; tasks specifically loading on episodic memory (e.g., recall tasks) are also good candidates for clinical assessment of dementia; (3) combining several language elicitation tasks (e.g., description, repetition, fluency and counting) can improve the discriminative power of the collected speech parameters; and (4) collecting samples which are long enough is crucial for improving speech sample quality for clinical assessment.

Our review also confirms the necessity to standardize the measuring protocols for assessing speech as an early marker of AD. Such a standardized way of measuring speech properties in ageing can be of high relevance to speech and language therapy practice. As stated elsewhere, speech is a highly promising biomarker of pathological ageing with great potential for clinical practice: it is easy to collect and measure over time and across different disease stages (De La Fuente García et al., 2020). To achieve this potential, it will be necessary to replicate the results in an independent sample through clinical trials in order to generalize them to practice.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Olga Ivanova https://orcid.org/0000-0002-9657-5380

REFERENCES

Ackermann, H. (2008) Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, 31(6), 265–272.

Ahmed, S., De Jager, C.A., Haigh, A.-M. & Garrard, P. (2013) Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology*, 27(1), 79–85.

An Xue, S. & Jianping Hao, G. (2003) Changes in the human vocal tract due to aging and the acoustic correlates of speech production. *Journal of Speech, Language, and Hearing Research*, 46(3), 689–701.

Ardila, A. (2020) A cross-linguistic comparison of category verbal fluency test (ANIMALS): a systematic review. *Archives of Clinical Neuropsychology*, 35, 205–217.

Arksey, H. & O'Malley, L. (2005) Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32.

Ash, S., McMillan, C., Gross, R.G., Cook, P., Gunawardena, D., Morgan, B., Boller, A., Siderowf, A. & Grossman, M. (2012) Impairments of speech fluency in Lewy body spectrum disorder. *Brain and Language*, 120(3), 290–302.

Ash, S., York, C., Bisbing, T., Min, N., Kofman, O. & Grossman, M. (2016) Longitudinal decline in speech fluency in Lewy body spectrum disorder (S35.005). *Neurology*, 86(Suppl. 16).

Ataollahi Eshkoor, S., Aizan Hamid, T., Yoke Mun, C. & Kyun Ng, C. (2015) Mild cognitive impairment and its management in older people. *Clinical Interventions in Aging*, 10, 687–693.

Attwood, T. (2007) *The complete guide to Asperger's syndrome*. London & Philadelphia: Jessica Kingsley Publishers.

Badarunisa, M.B., Sebastian, D., Rangasayee, R.R. & Kala, B. (2015) ICF-based analysis of communication disorders in dementia of Alzheimer's type. *Dementia and Geriatric Cognitive Disorders Extra*, 5, 459–469.

Balci, F., Meck, W., Moore, H. & Brunner, D. (2009) Timing deficits in aging and neuropathology. In Bizon J. & Woods, A. (Eds.) *Animal models of human cognitive aging. Aging medicine*. Totowa, New Jersey: Humana Press.

Bayles, K.A., Tomoeda, C.K. & Trosset, M.W. (1992) Relation of linguistic communication abilities of Alzheimer's patients to stage of disease. *Brain and Language*, 42(4), 454–472.

Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F. & Calzà, L. (2018) Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, 10, 369.

Berisha, V., Ktantsevich, C., Stegmann, G., Hahn, S. & Liss, J. (2022) Are reported accuracies in the clinical speech machine learning literature overoptimistic? *Interspeech 2022*, 691, 2453–2457.

Beveridge, M.E.L. & Bak, T.H. (2011) The languages of aphasia research: bias and diversity. *Aphasiology*, 25(12), 1451–1468.

Bilodeau-Mercure, M. & Tremblay, P. (2016) Age differences in sequential speech production: articulatory and physiological factors. *Journal of the American Geriatrics Society*, 64(11), 177–182.

Boersma, P. (2001) Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10), 341–345.

Boersma, P. & Weenink, D. (1992-2022) Praat: doing phonetics by computer [Computer program]. Version 6.3.09, retrieved 1 March 2023 from http://www.praat.org/

Bóna, J. (2014) Temporal characteristics of speech: the effect of age and speech style. *The Journal of the Acoustical Society of America*, 136(2), EL116.

Booth, A. (2006) "Brimful of STARLITE": towards standards for reporting literature searches. *Journal of the Medical Library Association*, 94(4), 421–429.

Bose, A., Dutta, M., Dash, N.S., Nandi, R., Dutt, A. & Ahmed, S. (2022) Importance of task selection for connected speech analysis in patients with Alzheimer's disease from an ethnically diverse sample. *Journal of Alzheimer's Disease*, 87(4), 1475–1481.

Burke, D.M. & Shafro, M.A. (2004) Aging and language production. *Current Directions in Psychological Science*, 13(1), 21–24.

Cuetos, F., Arango-Lasprilla, J., Uribe, C., Valencia, C. & Lopera, F. (2007) Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, 13(3), 433–439.

Dashwood, M., Churchhouse, G., Young, M. & Kuruvilla, T. (2021) Artificial intelligence as an aid to diagnosing dementia: an overview. *Progress in Neurology and Psychiatry*, 25(3), 42–47.

De La Fuente García, S., Ritchie, C.W. & Luz, S. (2020) Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review. *Journal of Alzheimer's Disease*, 78(4), 1547–1574.

De Looze, C., Kelly, F., Crosbt, L., Vourdanou, A., Coen, R.F., Walsh, C., Lawlor, B.A. & Reilly, R.B. (2018) Changes in speech chunking in reading aloud is a marker of mild cognitive impairment and mild-to-moderate Alzheimer's disease. *Current Alzheimer Research*, 15(9), 828–847.

De Looze, C., Kelly, F., Crosbt, L., Vourdanou, A., Coen, R.F., Walsh, C., Lawlor, B.A. & Reilly, R.B. (2021) Cognitive and structural correlates of conversational speech timing in Mild Cognitive Impairment and mild-to-moderate Alzheimer's disease: relevance for early detection approaches. *Frontiers in Aging Neuroscience*, 13, 637404.

Eyigoz, E., Mathur, S., Santamaria, M., Cecchi, G. & Naylor, M. (2020) Linguistic markers predict onset of Alzheimer's disease. *eClinicalMedicin*, 28, 100583.

Farrús, M. & Codina-Filbà, J. (2020) Combining prosodic, voice quality and lexical features to automatically detect Alzheimer's disease. arXiv:2011.09272v1

Ferreira, D., Nordberg, A. & Westman, E. (2020)Biological subtypes of Alzheimer disease. A systematic review and meta-analysis. *Neurology*, 94 (10), 436–48.

Fraser, K.C., Meltzer, J.A. & Rudzicz, F. (2016) Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49, 407–422.

Garcia-Toro, M., Talavera, J.A., Saiz-Ruiz, J. & Gonzalez, A. (2000) Prosody impairment in depression measured through acoustic analysis. *The Journal of Nervous and Mental Disease*, 188(12), 824–829.

Gerstenberg, A., Fuchs, S., Kairetn, J.M., Frankenberg, C. & Schröder, J. (2018) A cross-linguistics, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. Paper presented at 9th International Conference on Speech Prosody 2018, Poznan, 13–16 Jun (2018).

Geraudie, A., Battista, P., García, A.M., Allen, I.E., Miller, Z.A., Gorno-Tempini, M.L. & Montembeault, M. (2021) Speech and language impairments in behavioral variant frontotemporal dementia: a systematic review. *Neuroscience & Biobehavioral Reviews*, 131, 1076–1095.

Goberman, A.M. & Elmer, L.W. (2005) Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. *Journal of Communication Disorders*, 38(3), 215–230.

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J. & Hoffmann, I. (2019) Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, 53, 181–197.

Greenberg, S., Carvey, H., Hitchcock, L. & Chang, S. (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, 31(3–4), 465–485.

Hall, A.I., Shinkawa, K., Kosugi, A., Takase, T., Kobayashi, M., Nishimura, M., Nemoto, M., Watanabe, R., Tsukada, E., Ota, M., Higashi, S., Nemoto, K., Arai, T. & Yamada, Y. (2019) Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: preliminary results. *AMIA Joint Summits on Translational Science*, 2019, 34–43.

Harrington, J. & Cassidy, S. (1999) *Techniques in speech acoustics*. Dordrecht: Kluwer.

Hickok, G. (2012) Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135–145.

Hoffmann, I., Nemeth, D., Dye, C.D., Pákáski, M., Irinyi, T. & Kálman, J. (2010) Temporal parameters of spontaneous speech in Alzheimer's disease. *International Journal of Speech–Language Pathology*, 12(1), 29–34.

Ivanova, O. (2023) Aging language and language typology, or why we should pay attention to cross-linguistic features when studying aging. *Journal of Language and Aging Research* (in press).

Jack Jr., C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Budd Haeberlein, S.B. et al. (2018) NIA-AA research framework: Towards a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535–562.

Jang, H., Soroski, T., Rizzo, M., Barral, O., Harisinghani, A., Newton-Mason, S., Granby, S., Monnerat Stutz Da Cunha Vasco, T., Lewis, C., Tutt, P., Carenini, G., Conati, C. & Field, T.S. (2021) Classification of Alzheimer's disease leveraging multi-task machine learning analysis of speech and eve-movement data. *Frontiers in Human Neuroscience*, 15, 716670.

Jarrold, W., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M.L. & Ogar, J. (2014) Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: from Linguistic Signal to Clinical Reality, 27–37.

Kato, S., Endo, H., Homma, A., Sakuma, T. & Watanabe, K. (2013) Early detection of cognitive impairment in the elderly based on Bayesian mining using prosody and cerebral blood flow activation. 35th Annual International Conference of the IEEE EMBS, 2013, 581–5816.

Kato, S., Homma, A. & Sakuma, T. (2018) Easy screening for mild Alzheimer's disease and mild cognitive impairment from elderly speech. *Current Alzheimer Research*, 15(2), 104–110.

Kato, S., Homma, A., Sakuma, T. & Nakamura, M. (2015) Detection of Mild Alzheimer's disease and Mild Cognitive Impairment from elderly speech: binary discrimination using logistic regression. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, 556–5572.

Kavé, G. & Dassa, A. (2018) Severity of Alzheimer's disease and language features in picture descriptions. *Aphasiology*, 32(1), 27–40.

Kendall, T. & Fridlang, V. (2021) *Sociophonetics*. Cambridge: Cambridge University Press.

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P.H. & David, R. (2015) Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, *Assessment & Disease Monitoring*, 1, 112–124.

König, A., Satt, A., Sorin, A., Hoory, R., Derreumaux, A., David, R. & Robert, P.H. (2018) Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. *Current Alzheimer Research*, 15, 120–129.

Kuruvilla-Dugdale, M., Dietrich, M., Mckinley, J.D. & Deroche, C. (2020) An exploratory model of speech intelligibility for healthy aging based on phonatory and articulatory measures. *Journal of Communication Disorders*, 87, 105995.

Lau, J.C.Y., Patel, S., Kang, X., Nayar, K., Martin, G.E., Choy, J., Wong, P.C.M. & Losh, M. (2022) Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study. *PLoS ONE*, 17(6), e0269637.

Lee, H., Gayraud, F., Hirsch, F. & Barkat-Defradas, M. (2011) Speech dysfluencies in normal and pathological aging: a comparison between Alzheimer patients and healthy elderly subjects. *ICPhS*, 17, 1174–1177.

Leitman, D.I., Wolf, D.H., Laukka, P., Ragland, J.D., Valdez, J.N., Turetsky, B.I., Gur, R.E. & Gur, R.C. (2011) Not pitch perfect: sensory contributions to affective communication impairment in schizophrenia. *Biological Psychiatry*, 70(7), 611–618.

Levac, D., Colquhoun, H. & O'brien, K. (2010) Scoping studies: advancing the methodology. *Implementation Science*, 5, 69.

Li, R., Wang, X., Lawler, K., Garg, S., Bai, Q. & Alty, J. (2022) Applications of artificial intelligence to aid early detection of dementia: a scoping review on current capabilities and future directions. *Journal of Biomedical Informatics*, 127, 104030.

López-De-Ipiña, K., Martinez-De-Lizarduy, U., Calvo, P.M., Beitia, B., García-Melero, J., Fernández, E., Ecay-Torres, M., Faundez-Zanuy, M. & Sanz, P. (2020) On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment. *Neural Computing & Applications*, 32, 15761–15769.

Lowit, A., Brendel, B., Dobinson, C. & Howell, P. (2006) An investigation into the influences of age, pathology and cognition on speech production. *Journal of Medical Speech–Language Pathology*, 14, 253.

Macpherson, M.K. (2019) Cognitive load affects speech motor performance differently in older and younger adults. *Journal of Speech, Language, and Hearing Research*, 62(5), 1258–1277.

Martínez-Sánchez, F., Meilán, J.J.G., Carro, J. & Ivanova, O. (2018) A prototype for the voice analysis diagnosis of Alzheimer's disease. *Journal of Alzheimer's Disease*, 64, 473–481.

Meilán, J.J.G., Martínez-Sánchez, F., Carro, J., Sánchez, J.A. & Pérez, E. (2012) Acoustic markers associated with impairment in language processing in Alzheimer's disease. *The Spanish Journal of Psychology*, 15(2), 487–494.

Meilán, J.J.G., Martínez-Sánchez, F., Carro, J., Carcavilla, N. & Ivanova, O. (2018) Voice markers of lexical access in Mild Cognitive Impairment and Alzheimer's Disease. *Current Alzheimer Research*, 15, 111–119.

Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E. & Carro, J. (2020) Changes in the rhythm of speech difference between people with nondegenerative Mild Cognitive Impairment and with preclinical dementia. *Behavioural Neurology*, 2020, 4683573.

Marshall, C.R., Hardy, C.J.D., Volkmer, A., Russell, L.L., Bond, R.L., Fletcher, P.D., Clark, C.N., Mummery, C.J., Schott, J.M., Rossor, M.N., Fox, N.C., Crutch, S.J., Rohrer, J.D. & Warren, J.D (2018)Primary progressive aphasia: a clinical approach. *Journal of Neurology*, 265(6), 1474–1490.

Mendez, M.F. (2019) Early-onset alzheimer disease and its variants. *Continuum (Minneap Minn)*, 25(1), 34–51.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L.A. (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P). *Systematic Reviews*, 4(1), 1.

Mueller, K.D., Hermann, B., Mecollari, J. & Turksta, L.S. (2018) Connected speech and language in mild cognitive impairment and Alzheimer's disease: a review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, 40(9), 917–939.

Nasrolahzadeh, M., Mohammadpoory, Z. & Haddadnia, J. (2018) Higher-order spectral analysis of spontaneous speech signals in Alzheimer's disease. *Cognitive Neurodynamics*, 12, 583–596.

O'Malley, R., Mirheidari, B., Harkness, K., Reuber, M., Venneri, A., Walker, T., Christensen, H. & Blackburn, D. (2021) Fully automated cognitive screening tool based on assessment of speech and language. *Journal of Neurology, Neurosurgery & Psychiatry*, 92(1), 12–15.

Pappagari, R., Cho, J., Joshi, S., Moro-Velazquez, L., Zelasko, P., Villalba, J. & Dehak, N. (2021) Automatic detection and assessment of Alzheimer Disease using speech and language technologies in low-resource scenarios. *INTERSPEECH* 2021, 3825–3829.

Peters, M., Godfrey, C., Khalil, H., Mcinerney, P., Parker, D. & Soares, C. (2015) Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, 13(3), 141–146.

Pistono, A., Jucla, M., Barbeau, E.J., Saint-Aubert, L., Lemesle, B., Calvet, B., Köpke, B., Puel, M. & Pariente, J. (2016) Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *Journal of Alzheimer's Disease*, 50, 687–698.

Pistono, A., Jucla, M., Bézy, C., Lemesle, B., Le Men, J. & Pariente, J. (2018) Discourse macrolinguistic impairment as a marker of linguistic and extralinguistic functions decline in early Alzheimer's disease. *International Journal of Language & Communication Disorders*, 54(3), 390–400.

Pistono, A., Pariente, J., Bézy, C., Lemesle, B., Le Men, J. & Jucla, M. (2019) What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. *Neuropsychologia*, 124, 133–143.

Poole, M.L., Brodtmann, A., Darby, D. & Vogel, A.P. (2017) Motor speech phenotypes of frontotemporal dementia, primary progressive aphasia, and progressive apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 60, 897–911.

Powell, A.L., Cummings, J.L., Hill, M.A. & Benson, D.F. (1988) Speech and language alterations in multi-infarct dementia. *Neurology*, 38, 717–719.

Roncarolo, F., Boivin, A., Denis, J.L., Hébert, R. & Lehoux, P. (2017) What do we know about the needs and challenges of health systems? A scoping review of the international literature. *BMC Health Services Research*, 17, 636.

Rosen, S., (1992), Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B*, 33(1278), 367–373.

Rusz, J., Hlavnicka, J., Novotny, M., Tykalová, T., Pelletier, A., Montplaisir, J., Gagnon, J.F., Dusek, P., Galbiati, A., Marelli, S., Timm, P.C., Teigen, L.N., Janzen, A., Habibi, M., Stefani, A., Holzknecht, E., Seppi, K., Evangelista, E., Rassu, A.L., Dauvilliers, Y., Högl, B., Oertel, W., St Louis, E.K., Ferini-Strambi, L., Ruzicka, E., Postuma, R.B. & Sonka, K. (2021) Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Annals of Neurology*, 90(1), 62–75.

Sachin, S., Shukla, G., Goyal, V., Singh, S., Aggarwal, V., Gureshkumar & Behari, M. (2008) Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy. *Neurology India*, 56(2), 122–126.

Schultz, B.G., Rojas, S., ST John, M., Kefalianos, E. & Vogel, A.P. (2021) A cross-sectional study of perceptual and acoustic voice characteristics in healthy aging. *Journal of Voice*, S0892-1997(21)00193-4.

Shafro, M.A. & Tyler, L.K. (2014) Language in the aging brain: the network dynamics of cognitive decline and preservation. *Science*, 346, 583–587.

Skodda, S. & Schlegel, U. (2008) Speech rate and rhythm in Parkinson's disease. *Movement Disorders*, 23(7), 985–992.

Skodda, S., Visser, W. & Schlegel, U. (2011) Acoustic analysis of speech in progressive supranuclear palsy. *Journal of Voice*, 25(6), 725–731.

Sluis, R.A., Angus, D., Wiles, J., Back, A., Gibson, T., Liddle, J., Worthy, P., Copland, D. & Angwin, A.J. (2020) An automated approach to examining pausing in the speech of people with dementia. *American Journal of Alzheimer's Disease & Other Dementias*, 35, 1–8.

Sörös, P., Bose, A., Guttman Sokoloff, L., Graham, S.J. & Stuss, D.T. (2011) Age-related changes in the functional neuroanatomy of overt speech production. *Neurobiology of Aging*, 32(8), 1505–1513.

Taler, V. & Phillips, N.A. (2008) Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5), 501–556.

Torre, P. & Barlow, J.A. (2009) Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42(5), 324–333.

Tremblay, P., Sato, M. & Deschamps, I. (2017) Age differences in the motor control of speech: an fMRI study of healthy aging. *Human Brain Mapping*, 38(5), 2751–2771.

Tremblay, P., Deschamps, I., Bédard, P., Tessier, M.-H., Carrier, M. & Thibeault, M. (2018) Aging of speech production, from articulatory accuracy to motor timing. *Psychology and Aging*, 33(7), 1022–1034.

Tremblay, P., Poulin, J., Martel-Sauvageau, V. & Denis, C. (2019) Age-related deficits in speech production: from phonological planning to motor implementation. *Experimental Gerontology*, 126, 110695.

Tröger, J., Linz, N., König, Z., Robert, P., Alexandersson, J., Peter, J. & Kray, J. (2019) Exploitation vs. exploration—computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, 131, 53–61.

Tucker, B.V., Ford, C. & Hedges, S. (2021) Speech aging: production and perception. *WIREs Cognitive Science*, 12(5), e1557.

Verma, M. & Howard, R.J. (2012) Semantic memory and language dysfunction in early Alzheimer's disease: a review. *International Journal of Geriatric Psychiatry*, 27, 1209–1217.

Vincze, V., Szatlóczki, G., Tóth, L., Gosztolya, G., Pákáski, M., Hoffmann, I. & Kálmán, J. (2021) Telltale silence: temporal speech parameters discriminate between prodromal dementia and mild Alzheimer's disease. *Clinical Linguistics & Phoneics*, 35(8), 727–742.

Vogel, A.P., Fletcher, J., Snyder, P.J., Fredrickson, A. & Maruff, P. (2011) Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency. *Journal of Voice*, 25(2), 137–149.

Vogel, A.P., Poole, M.L., Pemberton, H., Caverlé, M.W.J., Bonstra, F.M.C., Low, E., Darby, D. & Brodtmann, A. (2017) Motor speech signature of behavioral variant frontotemporal dementia. Refining the phenotype. *Neurology*, 89(8), 837–844.

Voleti, R., Liss, J.M. & Berisha, V. (2020) A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 282–298.

Vuorinen, E., Laine, M. & Rinne, J. (2000) Common pattern of language impairment in vascular dementia and in Alzheimer disease. *Alzheimer Disease and Associated Disorders*, 14(2), 81–86.

Walenski, M., Sosta, K., Cappa, S. & Ullman, M.T. (2009) Deficits on irregular verbal morphology in Italian-speaking Alzheimer's disease patients. *Neuropsychologia*, 47(5), 1245–1255.

Williams, E., Mcauliffe, M. & Theys, C. (2021) Language changes in Alzheimer's disease: a systematic review of verb processing. *Brain and Language*, 223, 105041.

Yamada, Y., Shinkawa, K., Kobayashi, M., Caggiano, V., Nemoto, M., Nemoto, K. & Arai, T. (2021) Combining multimodal behavioral data of gait, speech, and drawing for classification of Alzheimer's disease and Mild Cognitive Impairment. *Journal of Alzheimer's Disease*, 84(1), 315–327.

Yeung, A., Iaboni, A., Rochon, E., Lavoie, M., Santiago, C., Yancheva, M., Novikova, J., Xu, M., Robin, J., Kaufman, L.D. & Mostafa, F. (2021) Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech–language changes in mild cognitive impairment and Alzheimer's dementia. Alzheimer's *Research & Therapy*, 13, 109.

Zellner, B. (1994) Pauses and the temporal structure of speech. In Keller E. (Ed.) *Fundamentals of speech synthesis and speech recognition*. Chichester: John Wiley, pp. 41–62.

Zhang, Z., Cummins, N. & Schuller, B. (2017) Advanced data expoitation in speech analysis: an overview. *IEEE Signal Processing Magazine*, 34(4), 107–129.

How to cite this article: Ivanova, O., Martínez-Nicolás, I. & Meilán, J.J.G. (2024) Speech changes in old age: Methodological considerations for speech-based discrimination of healthy ageing and Alzheimer's disease. *International Journal of Language & Communication Disorders*, 59, 13–37. https://doi.org/10.1111/1460-6984.12888

International Journal of Language & Communication Disorders

APPENDIX A: GLOSSARY

TABLE A1 Most relevant speech parameters for clinical assessment of healthy and pathological ageing

Speech variable	Description
Acoustic voice quality index (AVQI)	Multiparametric combination of six acoustic parameters quantifying overall voice quality (in clinics it is mainly aimed at measuring the level of dysphony).
Articulation rate/speech tempo	Number of units (e.g., phonemes) per unit of time (e.g., s/min), not considering pause intervals.
Asymmetry (skewness)	The shape of the spectrum below the centre of gravity comparing with the shape of the spectrum above the mean frequency.
Centre of gravity	The measure of the height of frequencies in a spectrum on average
Disfluency	Any type of disruption in the flow of spoken language during uttering.
Filled pause	Sound articulation which does not correspond to a word nor part of a word. The speaker articulates without speaking.
Long-term average spectrum (LTAS)	Power spectrum of the frequencies comprising speech sample and reflecting voice quality.
Mel frequency cepstral coefficients (MFCC)	Short-term power spectra of a sound.
Normalized pairwise variability index (nPVI)	Measure of average variation of a set of durations obtained from successive orders of events, measuring vowel length.
Pitch	Fundamental frequency of vibration of the vocal folds in the speakers, conditioning high or low voice.
Phonation time	The amount of time a speaker can sustain phonation.
Shimmer	Average absolute difference between the amplitude of consecutive periods, responsible for the brightness of the voice.
Silent pause	Interruption of articulation with no filling of the timing period.
Speech chunk	Units of several words separated by brief pauses during speaking.
Speech rate/speaking rate	Number of units (e.g., phonemes) per unit of time (e.g., s/min), including pause intervals.
Unvoiced segment	Segment of speech defined by non-periodic and noise-like nature, and low acoustic energy.
Voice break	Interruption during speaking driven by interruption in the air vibration in the vocal folds.
Voiced segment	Segment of speech defined by periodic nature and high acoustic energy.