

AMERICAN NATIONAL STANDARD

Method for Evaluation of the Intelligibility of Text-to-Speech Synthesis Systems

Secretariat:

Acoustical Society of America

Approved on May 6, 2013 by:

American National Standards Institute, Inc.

Abstract

This Standard is to be used for testing the speech intelligibility of text-to-speech systems, providing a measure of human listeners' recovery of words that correspond to the intended phonemic content of speech created by the system. Listeners are tasked to record the words or sentences they hear. Scoring may be either at the word or segment level. A normalized edit distance of the response from the intended message is the measure of the system's speech intelligibility. This Standard specifies methods for selecting test material, which may depend on the purpose and constraints of the test. The Standard also specifies methods for selecting and training the listeners; for designing, controlling, and reporting the test conditions; and for analyzing and reporting the test results. The Standard also provides background material, important for designing the test. Informative software is provided to assist the user in creating stimuli and scoring the test results. Use of the software is not mandatory.

AMERICAN NATIONAL STANDARDS ON ACOUSTICS

The Acoustical Society of America (ASA) provides the Secretariat for Accredited Standards Committees S1 on Acoustics, S2 on Mechanical Vibration and Shock, S3 on Bioacoustics, S3/SC 1 on Animal Bioacoustics, and S12 on Noise. These committees have wide representation from the technical community (manufacturers, consumers, trade associations, organizations with a general interest, and government representatives). The Standards are published by the Acoustical Society of America as American National Standards after approval by their respective Standards Committees and the American National Standards Institute (ANSI).

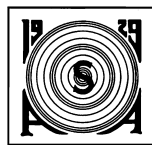
These standards are developed and published as a public service to provide standards useful to the public, industry, and consumers, and to Federal, State, and local governments.

Each of the Accredited Standards Committees (operating in accordance with procedures approved by ANSI) is responsible for developing, voting upon, and maintaining or revising its own Standards. The ASA Standards Secretariat administers Committee organization and activity and provides liaison between the Accredited Standards Committees and ANSI. After the Standards have been produced and adopted by the Accredited Standards Committees, and approved as American National Standards by ANSI, the ASA Standards Secretariat arranges for their publication and distribution.

An American National Standard implies a consensus of those substantially concerned with its scope and provisions. Consensus is established when, in the judgment of the ANSI Board of Standards Review, substantial agreement has been reached by directly and materially affected interests. Substantial agreement means much more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered and that a concerted effort be made towards their resolution.

The use of an American National Standard is completely voluntary. Their existence does not in any respect preclude anyone, whether he or she has approved the Standards or not, from manufacturing, marketing, purchasing, or using products, processes, or procedures not conforming to the Standards.

NOTICE: This American National Standard may be revised or withdrawn at any time. The procedures of the American National Standards Institute require that action be taken periodically to reaffirm, revise, or withdraw this Standard.



Acoustical Society of America
ASA Secretariat
35 Pinelawn Road, Suite 114E
Melville, New York 11747-3177
Telephone: 1 (631) 390-0215
Fax: 1 (631) 390-0217
E-mail: asastds@aip.org

© 2013 by Acoustical Society of America. This Standard may not be reproduced in whole or in part in any form for sale, promotion, or any commercial purpose, or any purpose not falling within the provisions of the U.S. Copyright Act of 1976, without prior written permission of the publisher. For permission, address a request to the Standards Secretariat of the Acoustical Society of America.

Contents

1	Scope.....	1
2	Normative references	1
3	Terms and definitions	1
4	Description of a text-to-speech synthesis system	2
5	General guidance for experimental design and testing	3
6	Requirements (Methods)	4
6.1	TTS system description and specification	4
6.2	Listeners.....	5
6.3	Selection and design of test materials	6
6.4	Intelligibility test procedures.....	7
6.5	Measurements and analysis of results	8
	Annex A (informative) Rationale for the recommendations concerning intelligibility test materials	9
A.1	Introduction	9
A.2	Acoustic cues to linguistic units vary from context to context.....	9
A.3	Systems vary in the algorithms they use and the types of errors they produce	10
A.4	Conclusion	12
	Annex B (normative) Methodological considerations for stimuli and responses: Considerations for test material containing names and nonsense words	13
B.1	Stimuli preparation	13
B.2	Response scoring	14
	Annex C (informative) Example software to create stimuli and score results in conformity with the method described in ANSI/ASA S3.50-2013	15
C.1	Disclaimer	15
C.2	Example software	15
	Bibliography.....	18

Figures

Figure 1 — Block diagram of a typical TTS system. This Standard primarily evaluates processing below the dotted line.	3
Figure A.1 — Spectrograms of (a) <i>Miss Peak</i> , (b) <i>Miss Beak</i> , and (c) <i>misspeak</i>	10

Tables

Table A.1 — Sample responses for one listener to <i>fake ill</i>	11
Table A.2 — Sample responses for one listener to <i>dock, cat, dock, bird</i>	12
Table A.3 — Sample responses for one listener to <i>Jupiter eyebrows</i>	12
Table C.1 — An example grammar for the susgen program showing sentence frames with Part of Speech (POS) tags, and the total number of syllables in the non-variable content of each frame.	16
Table C.2 — Example lexicon entries. Each row specifies a word, the POS tag to which that word can be assigned within grammar frames, and a syllable count.	16

Foreword

[This Foreword is for information only, and is not a part of the American National Standard ANSI/ASA S3.50-2013 American National Standard Method for Evaluation of the Intelligibility of Text-to-Speech Synthesis Systems. As such, this Foreword may contain material that has not been subjected to public review or a consensus process. In addition, it does not contain requirements necessary for conformance to the Standard.]

This Standard comprises a part of a group of definitions, standards, and specifications for use in bioacoustics. It was developed and approved by Accredited Standards Committee S3, Bioacoustics, under its approved operating procedures. Those procedures have been accredited by the American National Standards Institute (ANSI). The Scope of Accredited Standards Committee S3 is as follows:

Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, shock and vibration, which pertain to biological safety, tolerance and comfort.

The software provided with this American National Standard is entirely informative and is provided for the convenience of the user. Use of the provided software is not required for conformance with the Standard.

The Acoustical Society of America (ASA) and the owners of the copyright to the software provided with this American National Standard make no other representation or warranty or condition of any kind, whether express or implied (either in fact or by operation of law) with respect to any part of the product, including, without limitation, with respect to the sufficiency, accuracy or utilization of, or any information or opinion contained or reflected in, any of the product. ASA and the owners expressly disclaim all warranties or conditions of merchantability or fitness for a particular purpose. No officer, director, employee, member, agent, representative, or publisher of the copyright holder is authorized to make any modification, extension, or addition to this limited warranty.

At the time this Standard was submitted to Accredited Standards Committee S3, Bioacoustics, for approval, the membership was as follows:

C.J. Struck, *Chair*
G.J. Frye, *Vice-Chair*
S.B. Blaeser, *Secretary*

Acoustical Society of America	C.J. Struck
.....	M.D. Burkard (Alt.)
American Academy of Audiology	C. Schweitzer
.....	T. Ricketts (Alt.)
American Academy of Otolaryngology, Head and Neck Surgery, Inc.	R.A. Dobie
.....	L.A. Michael (Alt.)
American Industrial Hygiene Association	T.K. Madison
.....	D. Driscoll (Alt.)
American Speech-Hearing-Language Association	L.A. Wilber
.....	N. DiSarno (Alt.)
Beltone/GN Resound	S. Petrovic
Council for Accreditation in Occupational Hearing Conservation	L.D. Hager

ETS-Lindgren Acoustic Systems	S. Dunlap
.....	D. Winker (Alt.)
Etymotic Research, Inc.	M.C. Killian
.....	J.K. Stewart (Alt.)
Food and Drug Administration	S-C Peng
Frye Electronics, Inc.	G.J. Frye
.....	K.E. Frye (Alt.)
G.R.A.S. Sound and Vibration	J. Soendergaard
.....	B. Schustrich (Alt.)
Hearing Industries Association	VACANT
.....	C.M. Rogin (Alt.)
National Electrical Manufacturers Association, Signaling Protection and Communication Section (3SB)	J. McNamara
.....	R. Reiswig (Alt.)
National Hearing Conservation Association	G.L. Poling
National Institute for Occupational Safety and Health	M. Stephenson
.....	W.J. Murphy (Alt.)
National Institute of Standards and Technology	V. Nedzelnitsky
.....	R. Wagner (Alt.)
National Park Service	M. McKenna
.....	K. Fristrup (Alt.)
Natus Medical, Inc.	Y. Hekimoglu
.....	P. Becke (Alt.)
Ocean Conservation Research	M. Stocker
Starkey Laboratories	D.A. Preves
.....	T.H. Burns (Alt.)
U.S. Army Aeromedical Research Lab	W. Ahroon
U.S. Army CERL	D. Delaney
.....	M.J. White (Alt.)
U.S. Army Human Research & Engineering Directorate	P. Henry
.....	M. Ericson (Alt.)
U.S. Department of the Air Force	R. McKinley
.....	B.D. Simpson (Alt.)
University of Cincinnati Animal Audiology Clinic/Bioacoustics Lab	P.M. Scheifele
.....	D.K. Brown (Alt.)

Individual Experts of the Accredited Standards Committee S3, Bioacoustics, were:

J.R. Bareham	K.D. Kryter	C.J. Struck
A.J. Brammer	R. McKinley	H. Teder
R.F. Burkard	P.D. Schomer	L.A. Wilber
A.J. Campanella		W.A. Yost

Working Group S3/WG 91, Text-to-Speech Synthesis Systems, which assisted Accredited Standards Committee S3, Bioacoustics, in the development of this Standard, had the following membership:

	A.K. Syrdal, Chair	
C.A. Bickley	J. Lilley	D.M. Rekart
H.T. Bunnell	M.J. Makashay	M. Spiegel
S.R. Hertz	T. Mishra	

Suggestions for improvements of this Standard will be welcomed. They should be sent to Accredited Standards Committee S3, Bioacoustics, in care of the Standards Secretariat of the Acoustical Society of America, 35 Pinelawn Road, Suite 114E, Melville, New York 11747-3177. Telephone: 631-390-0215; FAX: 631-390-0217; E-mail: asastds@aip.org.

American National Standard

Method for Evaluation of the Intelligibility of Text-to-Speech Synthesis Systems

1 Scope

This American National Standard specifies an experimental method for evaluation of the intelligibility of synthetic speech, in English, generated by text-to-speech (TTS) synthesis systems. It is intended to be used by developers of applications that incorporate TTS technology, such as e-mail and SMS readers, talking kiosks, e-learning systems, navigation systems, automated messaging services, screen readers for people who are blind, and assistive devices for people who have difficulty speaking.

Although this Standard is targeted toward English, many of the recommendations and requirements concerning experimental design, listener selection and training, test materials and procedures, and measurement and analysis of results are sufficiently general to be valid for evaluating the intelligibility of synthetic speech in languages other than English.

This Standard describes methodology that is applicable both for comparisons of different TTS systems, and for comparisons of different versions of the same TTS system.

2 Normative references

The following referenced documents are indispensable for the application of this Standard. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ANSI/ASA S3.2-2009, *American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems*

3 Terms and definitions

For the purposes of this Standard, the terms and definitions given below apply:

3.1 speech synthesis. The generation of speech output from data input, which may include plain text; marked-up text; or parametric input, such as acoustic properties or articulatory configurations.

3.2 text-to-speech (TTS) synthesis. The generation of speech output from plain text or marked-up text.

3.3 intelligibility. That property which allows a human listener to identify words that correspond to the intended phonemic units of speech.

3.4 closed-response test. Evaluation in which participants, for each trial in the test, make a selection from a subset (termed a “closed set”) of possible responses. This procedure is exemplified by the familiar “multiple-choice” test format.

3.5 open-response test. Evaluation in which participants’ responses are not constrained to a subset of response alternatives, but instead are open to the full range of possible responses.

3.6 text pre-processing. The application-specific handling of text applied before input to a TTS system (e.g., re-ordering of words in telephone listings; adjustments for non-standard pronunciations