
Artificial Speech Test Stimulus (ASTS)

Product Description

Prepared by:
Psytechnics Limited

23 Museum Street
Ipswich, Suffolk
United Kingdom
IP1 1HN

t: +44 (0) 1473 261 800
f: +44 (0) 1473 261 880
e: info@psytechnics.com

Date January 2004

Table of Contents

Introduction.....	1
Features.....	1
How ASTS works.....	2
Background	3
Performance	4
Integration and Resources	5
Applications	5

Introduction

ASTS (Artificial Speech Test Stimulus) is a test signal which accurately represents the full range of sounds and transitions that occur in natural speech. The signal is ideal for performing like-for-like test comparisons and is recommended to be used with PESQ™ and PAMS.

At 30 seconds duration, ASTS is practical for both field and laboratory use and may easily be loaded onto test equipment.

ASTS is a software product which may be loaded onto test equipment or built in to network infrastructure, and is available as part of the PESQ Suite for equipment manufacturers to include on their products.

Features

- 30 second stimulus including both male and female talkers
- Represents 4 hours of conversation
- Currently available in two variants: British English and American English
- Major improvement over ITU-T P.50

P50	ASTS
Designed to match short and long-term spectral properties of speech, as well as amplitude distribution and syllabic envelope	Minimally redundant model of speech
Lacks clear formant structure for voiced events	Includes most important phonemes and phonetic transitions
Unvoiced speech events have unrepresentative (too broad) spectrum	Fully representative of short-term and long-term spectrum of speech
Temporal structure of P.50 different to English language structure	Natural amplitude distribution
Not suitable for testing systems with non-linear components e.g. speech codecs, voice activity detectors, etc	Clear speech-like phonetic and temporal structure
	Ideal for testing systems with non-linear components e.g. speech codecs, voice activity detectors, etc

Table 1: Comparison of ASTS and P.50

How ASTS works

It is essential to test systems with representative signals. Telephone networks are primarily designed to carry speech - in fact, low bit-rate coders are specifically optimized for speech. This means that measurements made using non-speech signals, such as tones or white noise, are unrepresentative and misleading.

Natural recorded speech can of course be used for network testing. Unfortunately speech has a lot of redundancy - for example, it contains a lot of silence and repetition. Thus to do a representative test can require 10 minutes of speech.

ASTS combines the representative properties of speech in a signal with much lower redundancy. Thus with 15s for each of two talkers (one male, one female), 30s in total, it is possible to make much faster, cheaper tests with all of the key characteristics of speech.

Background

Speech consists of a limited number of sounds and although subject to wide speaker dependency are characterized into groups of 'equivalent sounds' called phonemes. The specific sound of a phoneme is affected by its association with other phonemes and the context in which it is produced.

Natural English speech contains approximately 40 to 50 phonemes, depending on specific dialect, and conversational speech contains many repeated phonemes. This is a principle reason why real speech as a test signal results in a long test duration.

ASTS removes phonemic repetitions and combines a minimal selection and their first order transitions into linguistically legal sequences. These are then passed through BT's Laureate text-to-speech system to generate a speech signal. The resulting sequence is thus representative of the full range of sounds and transitions that occur in natural speech, with a practical duration of 30 seconds.



Figure 1: Portion of British English ASTS

Performance

With its emphasis on the characteristics of "real speech" ASTS is a considerable improvement on the ITU-T standard. To that end an application has been made to make ASTS part of ITU-T recommendation P.501.

ITU-T recommendation P.50 (1993) is a mathematically defined signal which is intended to reproduce certain properties of human speech. ASTS and P.50 are compared below.

	ASTS	P50
Derivation	Phonetic analysis of 4 hours of conversational speech.	Average spectrum of speech.
	Same short-term and long-term average spectrum as speech.	Same short-term and long-term average spectrum as speech.
Advantages	Same temporal distribution of amplitude as speech.	Approximate amplitude distribution and syllabic structure of speech.
	Similar phonetic structure (voiced/unvoiced) as speech.	
Disadvantages	Separate versions for different languages: American English - for US markets British English - for UK and European markets. Mandarin Chinese is currently under development in Psytechnics.	No proper formant structure in voiced events. Spectrum too broad during unvoiced events. Temporal structure not representative of English or other major languages.

Table 2: Performance of ASTS versus P.50

Integration and Resources

- Available as 16-bit audio sampled at 8 kHz and at 16 kHz sample rate
- Signal duration approx. 15s male, 15s female
- Variety of formats including PCM and Wave
- IRS send handset pre-processed version available for telephony testing applications
- Tailored conversational version with low memory requirement also available

Applications

- ASTS is the ideal test signal for use with PESQ in all network testing
- Also suitable for telephonometric testing of handsets and networks e.g. levels, transfer function estimation, echo canceller testing



Psytechnics, the Psytechnics logo and PESQ are trademarks of Psytechnics Ltd. Information subject to change without notice. Psytechnics assumes no responsibility for any errors or omissions that may appear in this document.