

SECCIÓN 8G: OBJETIVOS DE DISPONIBILIDAD Y CALIDAD E INTERFUNCIONAMIENTO CON REDES TERRENALES

INFORME 751 *

**MÉTODOS PARA LA EVALUACIÓN SUBJETIVA DE LA CALIDAD DE LA CONVERSACIÓN
EN EL SERVICIO MÓVIL MARÍTIMO POR SATÉLITE**

(Programa de Estudios 17A/8)

(1978)

1. Introducción

La evaluación de la calidad de servicio de una conexión telefónica que incluya un sistema del servicio móvil marítimo por satélite que puede estar conectado a circuitos nacionales o internacionales con conmutación, presenta algunas complicaciones en función de la diversidad de las condiciones y degradaciones que probablemente se encuentren en una conexión típica. Por lo menos uno de los enlaces introducirá siempre un largo tiempo de propagación del orden de 270 ms (valor medio en un sentido); serán necesarios supresores de eco y posiblemente otros equipos con conmutación vocal. Los diversos elementos del circuito que forman la conexión de la que es parte el sistema del servicio móvil marítimo, también contribuirán a otras formas de degradación tales como atenuación, ruidos en el circuito, distorsión «atenuación/frecuencia» y posiblemente distorsión de cuantificación. Debe tenerse en cuenta también que los niveles medios de la señal vocal y del ruido en el enlace por satélite pueden ser diferentes en los dos sentidos de la transmisión.

En teoría, el criterio último de la calidad de servicio de esa conexión es su calidad en servicio durante comunicaciones entre abonados tomados de la población de posibles usuarios. En estas condiciones estarán presentes todas las degradaciones de la palabra durante la escucha y la conversación. Sin embargo, este procedimiento puede no ser practicable por diversas razones por lo que la eficiencia de comunicación del sistema debe probarse por otros métodos. La simulación en laboratorio de comunicaciones en ambos sentidos para las pruebas de conversación, es factible siempre que las conexiones sean verdaderamente representativas y las pruebas se efectúen de forma metódica. Las pruebas de escucha únicamente tienen cabida en un procedimiento de evaluación, pero no sería prudente intentar predecir la calidad de servicio de sistemas marítimos por satélite en condiciones de conversación, fundándose exclusivamente en los resultados de esa clase de pruebas.

La finalidad del presente Informe es resumir los diversos métodos que podrían utilizarse para las pruebas de escucha únicamente y para las pruebas de conversación. El Informe sugiere los métodos más adecuados para estudiar los efectos de los diversos tipos de degradación que pueden encontrarse al establecer un servicio móvil marítimo por satélite y los más adecuados para establecer las diversas normas de calidad.

2. Objetivos de transmisión

Un objetivo primordial del establecimiento de un servicio móvil marítimo por satélite es permitir que se establezcan, entre los barcos y los abonados en tierra, conexiones tan buenas o casi tan buenas como las conexiones a través de las redes normales terrenales. Deben tomarse varias decisiones al desarrollar el servicio, y a continuación, figura una lista de los aspectos respecto de los cuales los métodos de evaluación influyen en la adopción de tales decisiones:

- determinación del sistema de modulación más adecuado de la señal vocal;
- en el caso de los sistemas de modulación que producen ruido correlacionado con la palabra, determinación del nivel de ruido blanco subjetivamente equivalente al ruido correlacionado con la palabra (véase el Informe 750, Kyoto, 1978);
- determinación de las normas mínimas de calidad de las comunicaciones de socorro;
- determinación del efecto del retardo, del eco y de la conmutación de las portadoras activadas por la voz;
- determinación de la calidad de funcionamiento requerida del sistema completo;
- investigación, en el curso de la explotación del servicio, de la medida en que éste satisface el nivel requerido de calidad y de los puntos en que pueden surgir dificultades concretas.

3. Pruebas de escucha únicamente

Existen muchos tipos de pruebas de escucha únicamente pero pueden dividirse en dos grandes categorías: los que miden la inteligibilidad de la conversación recibida por el canal que se prueba y los que miden la calidad global de la conversación recibida.

Para la primera finalidad se emplean las pruebas de **nitidez** que son relativamente fáciles de realizar aunque requieren personas entrenadas y el análisis es directo.

Entre las pruebas realizadas para las últimas figuran las notas de opinión y las pruebas isopreferenciales. Por lo general, para estos métodos son necesarias personas previamente entrenadas.

* Se ruega al Director del CCIR que señale este Informe a la atención del CCITT.

3.1 Pruebas de nitidez

3.1.1 Principios generales

La finalidad de las pruebas de **nitidez**, es conseguir la medida de la inteligibilidad de la conversación recibida por el canal sometido a prueba.

Por lo general las pruebas se realizan leyendo **textos** normalizados por el canal que se prueba. El porcentaje de sonidos de conversación recibidos correctamente por un grupo de escuchas se toma como medida de la inteligibilidad del canal. Puede expresarse como **nitidez** de palabras, logatomos, sonidos, consonantes o vocales.

Existe una gran variedad de métodos prácticos que difieren principalmente en el material utilizado y en las condiciones generales de los experimentos y son muy adecuados para la determinación de las normas mínimas de calidad de las comunicaciones de socorro.

La labor de las personas empleadas es sencilla y las pruebas de este tipo son relativamente fáciles de realizar. El análisis de los resultados es también directo. La interpretación de los resultados dependerá del método experimental específico utilizado.

3.1.2 Método de nitidez de logatomos

Los logatomos son sílabas carentes de significado compuestas de tres elementos:
consonante(s)-vocal-consonante(s).

Se elige cada uno de los sonidos que constituyen el logatomo aleatoriamente de una lista de fonemas representativos de diferentes lenguajes.

El procedimiento para formar logatomos se describe detenidamente en la Recomendación P.45, Tomo V del Libro Rojo del CCITT, página 74.

En este método, una persona entrenada escucha en el extremo receptor del canal sometido a prueba una lista de logatomos. En el otro extremo se enuncian los logatomos (no incluidos en ninguna frase) bien directamente por los locutores o mediante una cinta magnética previamente grabada (una prueba completa puede constar de 500 logatomos, pronunciados por varios locutores). La persona que interviene en la prueba escribe los sonidos que oye, confeccionando de esta forma una lista de logatomos «recibidos» que seguidamente se comparan con los de la lista leída.

El porcentaje de logatomos correctamente recibidos se denomina «~~nitidez~~ para los logatomos».

3.1.3 Métodos de ~~nitidez~~ de sonidos

Una prueba completa consta de 100 sílabas (cada una compuesta de uno o más sonidos) elegidas aleatoriamente en un idioma (por ejemplo, japonés).

Un locutor con experiencia pronuncia estas sílabas a un nivel constante de voz y una persona escribe las sílabas que oye. El porcentaje de sonidos correctamente recibidos se denomina «**nitidez** de sonidos».

3.1.4 Método de ~~nitidez~~ del CCITT

Se llama la atención sobre el método especificado en la Recomendación P.45 del CCITT, Libro Rojo, Tomo V, páginas 69-114.

Este método estuvo en principio concebido para medir equivalentes de referencia de **nitidez** pero puede utilizarse para otros fines.

3.1.5 Método de palabras fonéticamente equilibradas

Se señalan también las mediciones sobre inteligibilidad de la voz utilizando 50 palabras fonéticamente equilibradas (método de prueba PB-50 de la American National Standard, S3.2-1960: Método para la medición de la inteligibilidad de palabras monosilábicas).

Los experimentos han revelado que pueden modificarse los procedimientos normalizados de pruebas con 50 palabras fonéticamente equilibradas sin efectuar apreciablemente su confiabilidad y precisión [Milner, 1973; Milner y Golab, 1975]. Con objeto de llevar a cabo en un periodo limitado de tiempo un amplio programa de pruebas de inteligibilidad en experimentos marítimos y aeronáuticos con el ATS-6, el ritmo de lectura se incrementó a 2,5 segundos por palabra, suprimiéndose la frase clave. Las pruebas se limitaron a 400 palabras, estableciéndose 68 listas distintas de palabras. Estas modificaciones de los procedimientos de prueba de la American National Standard con palabras fonéticamente equilibradas no resultaron molestas para el oyente. Se evitó la memorización por parte del oyente y se hizo uso de una extensa gama de condiciones paramétricas al mismo tiempo que se aprovechaba al máximo el periodo de utilización del satélite y de la aeronave. (Véase el Informe 599.)

3.2 *Notas de opinión*

El método de notas de opinión intenta caracterizar los canales de transmisión telefónica utilizando una serie limitada de descripciones verbales (escala de opinión) de la calidad del canal. Después de escuchar la conversación transmitida por el canal sometido a prueba, se pide a personas que no han recibido adiestramiento que indiquen la descripción de calidad más apropiada para ese canal.

Los resultados de las pruebas de ese tipo pueden expresarse como porcentaje de respuestas en cada una de las diversas categorías adoptadas.

Además, pueden agruparse conjuntamente ciertas categorías, por ejemplo, en caso de utilizarse una escala de siete puntos, sumando los grupos 4, 5 y 6 se obtendrá el porcentaje de usuarios satisfechos y sumando los grupos 3, 4, 5 y 6 el porcentaje de los usuarios bastante satisfechos. En caso de utilizarse una escala de cinco puntos, los grupos correspondientes son A y B; A, B y C.

Asimismo, puede hacerse el cómputo de una nota media de opinión si se asignan valores numéricos a las categorías.

Por lo general, las pruebas de opinión son fáciles de realizar y el análisis de los resultados es directo. Sin embargo, en este método, la labor de la persona que escucha es difícil en el sentido de que han de expresar su opinión eligiendo una de una serie de descripciones de calidad. Los resultados dependen de la descripción de la escala, que debe reflejar adecuadamente la gama de calidad que abarca la prueba. Cuando se utilizan las formulaciones adecuadas, la interpretación de los resultados es directa.

Además, podrían utilizarse los resultados para calcular por interpolación el nivel de ruido blanco gaussiano que sería subjetivamente equivalente al nivel de ruido correlacionado con la palabra. Las personas que intervienen en la prueba escucharían por canales que tuviesen ruido correlacionado con la palabra y por canales con ruido blanco gaussiano (canales de referencia). Los niveles de ruido introducidos en los canales de referencia deberían cubrir una gama de valores de nota media de opinión más amplia que los canales con ruido correlacionado con la palabra.

A continuación se exponen dos escalas de opinión características:

3.2.1 Escala de 7 puntos

<i>Puntuación</i>	<i>Descripción de la calidad</i>
6	Circuito ideal.
5	Circuito excelente. Posibilidad de relajación completa durante la conversación. Muy agradable.
4	Circuito bueno. Necesidad de prestar atención, pero no es necesario hacer un esfuerzo especial. Circuito agradable.
3	Circuito bastante bueno, es necesario un cierto esfuerzo aunque no demasiado grande. No es un circuito muy agradable.
2	Circuito regular. Se puede oír, pero algunas veces con dificultad. Es desagradable de escuchar.
1	Circuito malo. Únicamente puede utilizarse con gran dificultad. Es muy desagradable de escuchar.
0	Circuito muy malo. Prácticamente inutilizable.

3.2.2 Escala de 5 puntos (esfuerzo de escucha) (Referencia: anexo 5 a la Cuestión 1/XII, parte C, Libro Verde del CCITT, Tomo V)

A	Posibilidad de una relajación completa: no se necesita esfuerzo alguno.
B	Es necesario prestar atención: se necesita un cierto esfuerzo.
C	Necesario un esfuerzo moderado.
D	Necesario un considerable esfuerzo.
E	No se entiende el significado ni haciendo considerables esfuerzos.

Uno de los problemas de las anteriores escalas es que no son fáciles de traducir a otros idiomas y por ello, las personas que participan en la prueba tienen dificultades para comprender la diferencia entre las notas de la escala. Una consecuencia posible es que las pruebas realizadas en idiomas diferentes pero que tengan, por lo demás, los mismos parámetros pueden no ser comparables. Para salvar ese problema idiomático puede utilizarse una escala sencilla, fácil de traducir a diversos idiomas, como por ejemplo: muy bueno, bueno, bastante bueno, mediocre, muy mediocre. Es, sin embargo, muy importante que al utilizar una escala de ese tipo se plantee al participante la pregunta correcta, pues de lo contrario puede no quedar claro si se está juzgando la palabra o el canal, y carecer de coherencia, en consecuencia, los resultados.

Los métodos de notas de opinión son adecuados para evaluar el efecto de la mayoría de los tipos de degradación, especialmente cuando la transmisión aspira a una buena calidad más que a la mera inteligibilidad. No permiten evaluar los efectos del retardo en la propagación y de los ecos.

3.3 *Métodos de comparación para la determinación del ruido equivalente*

3.3.1 *Consideraciones generales*

El empleo de ruido equivalente se debe a la necesidad de evaluar el efecto que ejercen en la calidad global de transmisión los circuitos que introducen distorsiones correlacionadas con la palabra. Para ello puede determinarse el ruido equivalente del sistema «desconocido» para este sistema únicamente, sin introducir ninguna degradación adicional (Informe 750, Kyoto, 1978). Puede calcularse entonces la calidad global de diversas conexiones por los métodos normales de cálculo de la calidad de transmisión (véase el Libro Naranja del CCITT, anexos a la Cuestión 7/XII).

Puesto que los métodos de evaluación subjetiva empleados para determinar el ruido equivalente conciernen a las conexiones telefónicas de carácter esencialmente normal, los mismos debieran basarse en estimaciones de la calidad global. Asimismo, si se quiere obtener resultados coherentes, este proceso de estimación debiera obedecer a la ley de transitividad, o sea, si $A = B$ y $B = C$, entonces $C = A$.

Las funciones que han de realizar las personas que intervienen en la prueba difieren ligeramente en los métodos que se describen a continuación, que establecen la preferencia, la similitud y la ordenación por rango respectivamente. Se cree, no obstante, que todos los métodos se basan en estimaciones de la calidad global.

3.3.2 *Método isopreferencial*

En el método isopreferencial se evalúa el canal sometido a prueba mediante la comparación con un canal de referencia que tiene volúmenes variables de ruido. En cada par de referencia y en cada canal de prueba se pide a la persona que interviene en las pruebas que indique cuál de las dos muestras de conversación prefiere escuchar. El nivel isopreferencial (de ruido) es pues el valor cuando el 50% de las personas que escuchan prefieren un canal y el 50% el otro. Dado que en la práctica no es probable que se obtenga el nivel isopreferencial, se utiliza un procedimiento de interpolación para llegar a él; las pruebas han demostrado que este método cumple el requisito de transitividad.

3.3.3 *Método de ajuste*

Este método se denomina también método del error medio. Las personas que intervienen en la prueba controlan la potencia de ruido en el canal de referencia. Se les pide que ajusten el canal de referencia seleccionando el volumen de ruido necesario para que dicho canal y el de prueba sean equivalentes con respecto a la calidad global. En este método, el nivel isopreferencial se obtiene directamente mediante el ajuste del nivel de ruido.

Las pruebas han demostrado que este método satisface también el requisito de transitividad.

3.3.4 *Método de ordenación por rango*

El método consiste en la ordenación por rango de una mezcla de técnicas de tratamiento conocidas y desconocidas, según la calidad, y se le conoce en experimentación estadística por el nombre de modelo de los cuadrados de Youden ($YS = \text{Youden Squares}$). Por técnica de transmisión «conocida» se entiende aquella cuya calidad subjetiva es conocida en términos de una señal analógica con ruido gaussiano. Las técnicas de tratamiento de la transmisión «desconocidas» son todas aquellas cuya calidad tiene que determinarse.

La mezcla de técnicas de transmisión conocidas y desconocidas se compone de tal manera que, mediante una cuidadosa selección previa, la gama de calidad de las técnicas conocidas abarque las de las técnicas desconocidas, y después se obtienen por interpolación los valores de ruido equivalente de las técnicas desconocidas.

El modelo de los cuadrados de Youden permite la ordenación por rango subjetiva de muchos estímulos, por ejemplo, muestras de palabra tratadas por técnicas diferentes, mediante la presentación de sólo una pequeña fracción por vez a las personas que intervienen en la prueba. Esa fracción se denomina bloque. Por ejemplo, se pueden ordenar trece estímulos presentando únicamente bloques de cuatro cada vez, y este modelo se denomina 4×13 YS. Los estímulos utilizados en este caso, designados S_i , son muestras grabadas de palabras tratadas por diversas técnicas en condiciones especificadas.

El método de los cuadrados de Youden garantiza que la división en bloques a fines de ordenación conserve un cierto equilibrio, y en particular que cada estímulo S_i se escuche un número igual de veces en el experimento y que cada par de estímulos posibles aparezca con la misma frecuencia en un número igual de bloques.

El modelo de 13×4 cuadrados de Youden es el que más se utiliza, ya que a la mayoría de las personas les resulta generalmente demasiado difícil comparar más de cuatro muestras a la vez. En este modelo se mezclan seis sistemas «desconocidos» con siete sistemas cuya relación señal analógica/ruido se conoce.

Para cada bloque se pide a la persona que interviene en la prueba que clasifique los cuatro estímulos asignando números entre 1 (el mejor) y 4 (el peor) a las muestras de palabra según sus propias preferencias.

Debe realizarse una prueba de concordancia de coeficientes para asegurar una elevada probabilidad de que exista el orden observado. El análisis consiste en calcular el rango medio de cada sistema. Para la conversión del orden por rango a ruido equivalente se determina una curva de regresión del rango medio en función de la relación señal/ruido, para los tratamientos analógicos, y los rangos medios de los tratamientos desconocidos se transforman en relaciones señal/ruido con ayuda de esta curva.

Puede hallarse un ejemplo del empleo de este método en el Informe 752, anexo II.

4. Pruebas de conversación

4.1 Principios generales

Las pruebas de conversación se refieren en primer lugar a la calidad global de las conexiones telefónicas. Los efectos que surgen en una conversación se tienen presentes de la forma más realista posible.

Las pruebas se realizan mediante pares de personas no adiestradas que conversan por la conexión sometida a prueba. Inmediatamente después de terminada la conversación sin perturbación, se les pregunta su opinión sobre el circuito. El método puede por consiguiente utilizarse en conexiones reales o simuladas.

Aunque los cuestionarios pueden variar, generalmente incluyen una nota de opinión. Se pide a las personas que efectúan la prueba que indiquen la descripción de calidad que coincida con su opinión sobre la calidad del circuito. Además, se les puede preguntar si experimentaron o no dificultades. En caso afirmativo se les puede plantear otras cuestiones para probar el carácter de la dificultad pero sin sugerir los tipos posibles de dificultad. Se les interroga por separado y no se les permite hablar sobre sus opiniones. (Referencia: apéndice 2 al anexo 2 a la Cuestión 4/XII, Libro Verde del CCITT, Tomo V.)

Las conversaciones demandan mucho tiempo ya que se permite a las personas que efectúan las pruebas que las realicen a su propio ritmo. Las conversaciones entre ellos constituyen los principales problemas que plantean los experimentos de este tipo.

El método de análisis de resultados es semejante al empleado en la prueba de las notas de opinión. Se calcula la nota media de opinión y la nota porcentual de cada categoría y de cada grupo de categorías. Además, se especifica el porcentaje de personas que experimentaron dificultades o cualquier otro tipo especial de distorsión enumerado en el cuestionario.

4.2 Métodos prácticos

En el anexo 2 a la Cuestión 4/XII del CCITT, Libro Verde, Tomo V, se describe un método especial para realizar las pruebas de conversación. Normalmente se utiliza una escala de calificación de 5 puntos y se obtiene por ejemplo excelente, bueno, regular, mediocre y malo. La nota media de opinión se calcula como sigue:

$$\text{Nota media de opinión} = (4E + 3G + 2F + P)/N$$

en donde

- E*: Número de conferencias calificadas de excelentes,
- G*: Número de conferencias calificadas de buenas,
- F*: Número de conferencias calificadas de regulares,
- P*: Número de conferencias calificadas de mediocres,
- N*: Número total de conferencias.

(El número de conferencias calificadas de «malas» recibe un cero e influye en la nota media de opinión al estar incluido en *N*.)

5. Montaje experimental

En la fig. 1 se representa una simulación de laboratorio de una conexión característica abonado a abonado en la que se incluyen las degradaciones que pueden experimentarse en una comunicación en el servicio móvil marítimo por satélite. Cuando se realizan las pruebas de escucha únicamente no pueden evaluarse algunas de esas degradaciones y pueden suprimirse en la conexión simulada algunos equipos, como el equipo que simula el retardo de propagación y los supresores de eco.

Por razones prácticas, puede ser conveniente efectuar pruebas de escucha únicamente por enlaces de satélite aisladamente; sin embargo, debe tenerse cuidado al sacar las conclusiones de estas pruebas, ya que, en teoría, en ellas deben tenerse en cuenta factores como la atenuación, la distorsión y el ruido del circuito que pueden influir en la señal telefónica aplicada a los moduladores del sistema por satélite tanto en la estación terrena como en el terminal de barco; también deben considerarse la existencia y características de todo tipo de ruidos en los lugares de escucha. Deben efectuarse pruebas por separado en ambos sentidos de transmisión.

Ha de determinarse el ruido que se produce en la prolongación terrenal y que se inyecta a la conexión simulada y habrá que tener en cuenta que la longitud de las prolongaciones terrenales puede variar considerablemente.

A efectos comparativos, durante las pruebas es necesaria una conexión de referencia, que debe ser similar a la conexión sometida a prueba aunque con el enlace del servicio móvil marítimo sustituido por un enlace del servicio fijo por satélite.

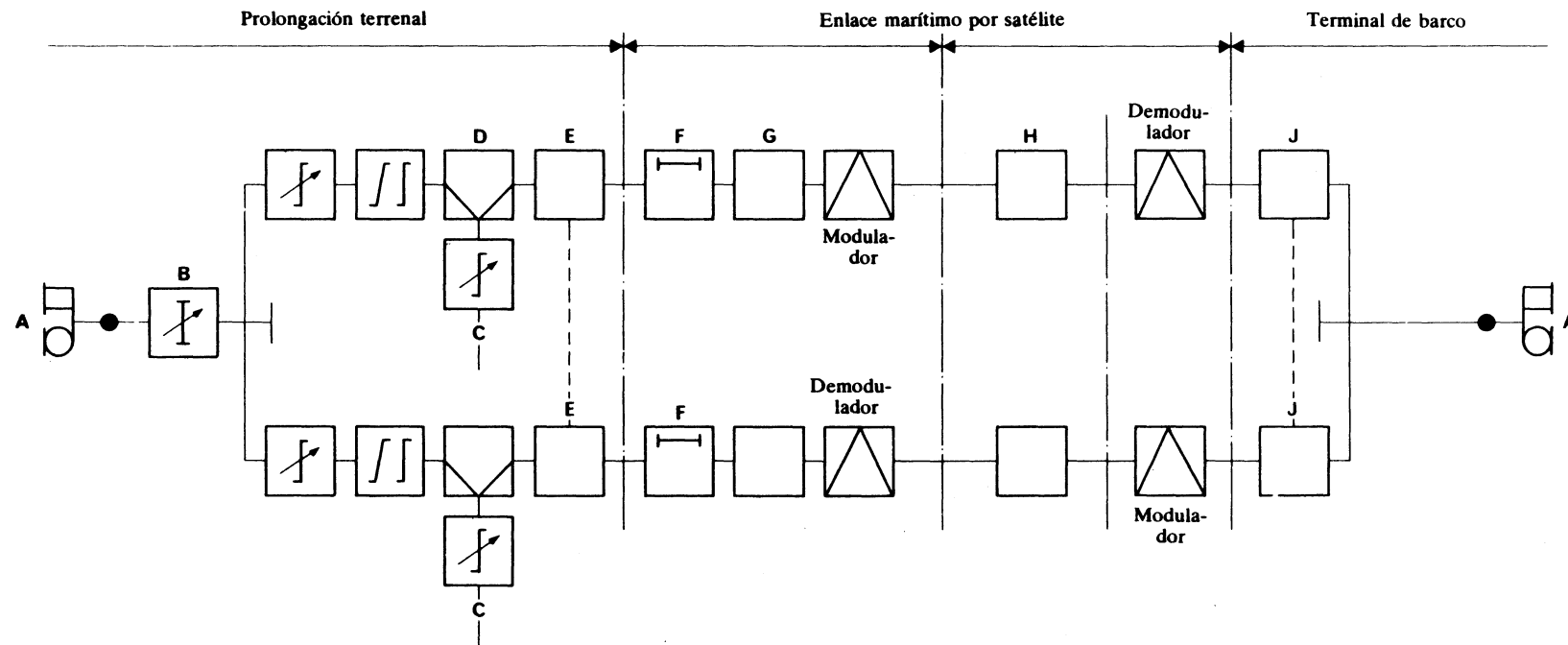


FIGURA 1 — Simulación en laboratorio de una conexión típica de abonado a abonado

- | | | |
|--|----------------------------------|--------------------------|
| A: Ruido ambiente | D: Atenuador del ruido inyectado | G: Activador por la voz |
| B: Línea artificial de longitud variable | E: Semisupresor de eco | H: Simulador de satélite |
| C: Ruido | F: Retardo | J: Semisupresor de eco |

Nota. — En un sistema telefónico a cuatro hilos a bordo de un barco no será necesario el equipo terminal para pasar de cuatro hilos a dos hilos. Véase también la Recomendación 550 en lo que respecta al semisupresor de eco insertado en el extremo de la conexión situado en el barco.

6. Conclusiones

Las pruebas de conversación de abonado a abonado constituyen el único método satisfactorio de evaluar la calidad de una conexión telefónica que incluye un sistema del servicio móvil marítimo por satélite. Dichas pruebas son las únicas que pueden tener en cuenta las degradaciones debidas al retardo de propagación a través del satélite. Desgraciadamente, las pruebas de conversación son difíciles de realizar y duran mucho tiempo; las pruebas de escucha únicamente son más sencillas y pueden emplearse en ciertos aspectos de evaluación, si bien deben tenerse en cuenta sus limitaciones al adoptar decisiones basadas en los resultados de estas pruebas.

En especial, debe reconocerse que las pruebas de nitidez no son muy adecuadas para la evaluación de las conexiones telefónicas ordinarias. No obstante, pueden aplicarse para fines especiales, por ejemplo, determinación del efecto de la conmutación de portadora activada por la voz (truncamiento de las sílabas iniciales) y normas de calidad para las comunicaciones de socorro.

Para determinar el método más adecuado de modulación de las señales vocales y evaluar los niveles de ruido equivalente de un enlace por satélite deben emplearse métodos únicamente de escucha que evalúen la calidad global.

Se formulan las siguientes recomendaciones:

- las pruebas únicamente de escucha pueden utilizarse en las pruebas preliminares para la comparación de métodos de modulación; en teoría deben efectuarse sobre una conexión completa;
- pueden realizarse pruebas de escucha únicamente que evalúen el ruido equivalente de los métodos de modulación a fin de calcular la calidad de transmisión para una diversidad de conexiones completas representativas;
- las pruebas de conversación en laboratorio sobre una conexión simulada deben efectuarse de preferencia bajo los auspicios de un organismo internacional;
- deben realizarse eventualmente pruebas de conversación en servicio.

REFERENCIAS BIBLIOGRÁFICAS

- MILNER, P. [junio de 1973] Advantages of experienced listeners in intelligibility testing. *IEEE Trans. Audio and Electroacoustics*, Vol. AU-21, 161-165.
- MILNER, P. y GOLAB, J. [abril de 1975] Intelligibility of voice transmission through a satellite relay system. *J. Acous. Soc. Amer.*, Vol. 57, Supp. 1, 23.

BIBLIOGRAFÍA

- KENDALL, M. G. [1974] *Rank Correlation Methods*, Capítulos 9 y 10. Charles Griffin and Co., Ltd.
-